Research Note 83-37

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL:

TECHNICAL APPENDIX TO THE ANNUAL REPORT

Newell K. Eaton and Marvin H. Goer
Editors

U.S. Army
Research Institute for the Behavioral and Social Sciences

October 1983

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Research Note 83-37

2. GOVT ACCESSION NO.
AD A137 117

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Improving the Selection, Classification and Utilization of Army Enlisted Personnel - Technical Appendix to the Annual Report

5. TYPE OF REPORT & PERIOD COVERED
1 Oct 82 - 30 Sep 83

6. PERFORMING ORG. REPORT NUMBER
RP-PRD-83-25

7. AUTHOR(s)
Newell K. Eaton and Marvin H. Goer (Editors)

8. CONTRACT OR GRANT NUMBER(s)
MDA 903-82-C-0531

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Human Resources Research Organization
1100 S. Washington Street
Alexandria, VA 22314

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
2Q263731A792

11. CONTROLLING OFFICE NAME AND ADDRESS
U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Ave., Alexandria, VA 22333

12. REPORT DATE
October 1983

13. NUMBER OF PAGES
737

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
The Army Research Institute Technical Point of Contact is: Dr. Newell K. Eaton. His telephone number is (202) 274-8275.

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

Validation	Job Knowledge Tests	Predictors
Predictor Measures	Validity Generalization	Criterion Measures
Construct Validation	Performance Measures	Army-wide Measures
Longitudinal Data Base		

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

This research note describes in detail research performed during the first year of a project to develop a complete personnel system for selecting and classifying all entry level enlisted personnel. Its purpose is to document, in the context of the annual report, a variety of technical papers associated with the project. In general, the first year's activities have been taken up by an intensive period of detailed planning, briefing advisory groups, preparing initial troop requests, and beginning comprehensive predictor and criterion development that will be the basis for later validation work. Research reports associated with the work reported are included.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Research Note 83-37

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL:

TECHNICAL APPENDIX TO THE ANNUAL REPORT

Newell K. Eaton and Marvin H. Goer
Editors

Submitted by:

Newell K. Eaton, Chief
SELECTION AND CLASSIFICATION
TECHNICAL AREA

Approved as technically adequate
and submitted for publication by:

Joyce L. Shields, Director
MANPOWER AND PERSONNEL
RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

October 1983

Army Project Number
2Q263731A792

Manpower and Personnel

FOREWORD

This document describes the research performed during the first year of a project on a path toward achieving the goals of the Army's current, large-scale manpower and personnel research effort for improving the selection, classification, and utilization of Army enlisted personnel. The thrust for the project came from the practical, professional, and legal need to validate the Armed Services Vocational Aptitude Battery (ASVAB--the current US military selection/classification test battery) and other selection variables as predictors of training and performance. The portion of the effort described herein is devoted to the development and validation of Army Selection and Classification Measures, and referred to as "Project A." This work is funded primarily by Army Project Number 2Q263731A792. Another part of the effort is the development of a prototype Computerized Personnel Allocation System, referred to as "Project B." Together, these Army Research Institute research efforts, with their in-house and contract components, comprise a landmark program to develop a state-of-the-art empirically validated personnel selection, classification, and allocation system.

EDGAR M. JOHNSON
Technical Director, ARI and
Chief Psychologist, U. S. Army

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

FIGURES

TABLE OF CONTENTS (Continued)

FIGURES

TABLES

TABLE OF CONTENTS (Continued)

TABLES

TABLE OF CONTENTS (Continued)

ASSOCIATED REPORTS

TABLE OF CONTENTS (Continued)

ASSOCIATED REPORTS

I. INTRODUCTION

Newell K. Eaton (ARI)
Marvin H. Goer (HumRRO)

The purpose of this research note is to document, in the context of the first annual report, a variety of technical aspects of the plans and accomplishments of Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel. Project A, together with the related Enlisted Personnel Allocation System research effort (Project B), are designed to provide a significant increase in Army readiness. These unique, long term, large scale research programs will tie together selection, classification and job allocation of Army enlisted personnel so that personnel decisions are made to optimize soldier performance and utilization of soldier skills and abilities. The research will provide information and procedures required to meet the manpower challenge of the coming decade by assuring that the most qualified people are enlisted, allocated, and retained. The objectives of the research are to develop an integrated personnel management system based on: 1) current and new personnel and performance measures, 2) accurate empirical prediction of future performance, 3) selection/classification, and MOS allocation at enlistment and reenlistment to optimize individual and system performance, and 4) what-if gaming to illustrate the performance impact of possible personnel management decisions.

The thrust of the program came from the practical, professional, and legal need to demonstrate the validity of the Armed Services Vocational Aptitude Battery (ASVAB--the current military selection/classification test battery) and other selection variables used as predictors of training and job performance. Research planners at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) realized the sizable resource commitment required to show ASVAB validity. The resource commitment would be mostly for the development and application of training and job performance measures. It became apparent that with moderate additional effort the predictor space could be greatly enhanced with new tests, and an integrated personnel management system could be developed to more optimally use the predictor and performance information.

The following text provides a short history of Project A, a description of its organization and structure, a more detailed discussion of its goals and objectives, and a report of the activities and accomplishments of the Project through its first year of full scale operation, fiscal year 1983.

Project Background

In response to Army, Congressional, and professional requirements, ARI began in 1980 to develop a major personnel selection, classification, and allocation research program. The basic requirement was to demonstrate the validity of the ASVAB as a predictor of both training and on-the-job performance. In reviewing the design needed to meet that requirement, the concept of a larger project began to emerge. With only a moderate amount of additional resources, new predictors in the perceptual, psychomotor, interest, temperament, and biodata domains could be evaluated as well. And a longitudinal research data base could be developed, linking soldiers' performance on a variety of variables from enlistment, through training, first tour assignments, reenlistment decisions, and for some, to their second tour. Finally, those data could be the basis for a new way to allocate personnel, making near-real-time decisions on the best match between characteristics of an individual enlistee or reenlistee and the requirements of available Army military occupational specialties.

To address the selection and classification portion of the effort, solicitation MDA 903-81-12-R-0158 "Project A: Development and Validation of Army Selection and Classification Measures" was issued Oct. 21, 1981. This milestone document can be viewed as the "official" starting point of this landmark research program which has now completed its first year. The program was intended to bring together the best Army in-house and contract research scientists in a combined effort to meet the Army's requirements for improving their enlisted personnel selection and classification processes and programs. In the solicitation, Army research psychologists mapped out a comprehensive 7-year effort to provide the tools and information necessary for implementation of a state-of-the-art selection and classification system for all enlisted personnel in the U.S. Army.

Changes at ARI

While the contract SOW and RFP process was ongoing, substantial changes were being made within ARI to increase emphasis in the manpower and personnel area. The new manpower and personnel laboratory was created, and Dr. Joyce L. Shields was chosen as director. To accommodate the substantial in-house portion of Project A, the selection and classification technical area was established, with Dr. Newell K. Eaton as chief. A major recruitment effort brought together a staff of experienced research scientists to execute the in-house research and to monitor the contract effort.

Formation of the Consortium

In anticipation of the solicitation (RFP), the presidents of the Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), and Personnel Decisions Research Institute (PDRI) formed a consortium to develop a research proposal to meet the requirements of the forthcoming "Development and Validation of Army Selection and Classification Measures" Request for Proposal (RFP). It was agreed that HumRRO, as prime contractor, would assume responsibilities for overall contract management, technical direction and planning, and for scientific quality assurance.

Proposal and Award

In response to the RFP, the consortium's proposal was submitted in January 1982. In May 1982 the principal scientists and managers of the consortium met with the ARI proposal evaluation team to review the proposal and to respond to technical questions and issues. The consortium was also asked to submit an addendum containing written responses to a number of additional questions raised by ARI. The addendum was submitted in June 1982. In accordance with standard procurement procedures, the consortium was asked in September 1982 to submit a "Best and Final" amendment to the proposal, in which further clarification was provided for the cost estimates and for the proposed project management structure. This "Best and Final" offer was successful, negotiations were conducted, and a contract was awarded to the HumRRO-AIR-PDRI consortium on September 30, 1982. The contract covered a 7-year research program at an estimated overall total cost of $16,390,000.

Project Outline

The overall purpose of Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel is to enhance the Army's ability to accomplish its peacetime and mobilization missions through improved matching of individuals to military occupational specialties (MOS). Specifically, Project A is to:

(1) validate existing selection measures against both existing and project-developed criteria, the latter to include both Army-wide performance measures based on newly developed rating scales and direct measures of MOS-specific task performance;

(2) develop and validate new and/or improved selection and classification measures;

(3) validate proximal criteria, such as performance in training, as predictors of later criteria, such as job performance ratings, so that more informed reassignment and promotion decisions can be made throughout the individual's tour;

(4) determine the relative utility to the Army of different performance levels across MOS; and

(5) estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility for making operational selection and classification decisions.

The project must not be viewed and is not being conducted as a set of separate tasks that make "inputs" to one another and that are to be "integrated" somehow. Such a view misses the essential unity of the effort; Project A is one project and is organized into five major tasks.

Task 1. Validation

Task 1 has two major components. The first component is to maintain the data base and provide the analytic procedures to determine the degree to which performance in Army jobs is predictable from some combination of new or existing measures. The second component is to conduct the appropriate analyses to determine whether the existing set of predictors, new predictors, or some combination of new and existing predictors has utility over and above the present system. These two components must be accomplished using state-of-the-art technology in personnel selection research and data analytic methods.

Task 2. Developing Predictors of Job Performance

To date, a large proportion of the efforts of the armed services in this area have been concentrated on improving the ASVAB, which is now a well-researched and valid measure of general cognitive abilities. However, many critical Army tasks appear to require psychomotor and perceptual skills for their successful performance. Further, neither biodata nor motivational variables are now comprehensively evaluated. It is perhaps in these four noncognitive domains that the greatest potential for adding valid independent dimensions to current classification instruments is to be found. The objectives of Task 2 are to develop a broad array of new and improved selection measures and to administer them to three major validation samples. A critical aspect of this task is the demonstration of the incremental validity added by new predictors.

Task 3. Measurement of School/Training Success

The objective of Task 3 is to derive school and training performance indexes that can be used: (1) as criteria against which to validate the initial predictors, and (2) as predictors of later job performance. Comprehensive job knowledge tests will be developed for the sample of MOS investigated and their content and construct validity will be determined.

Task 4. Assessment of Army-wide Performance

In contrast to performance measures which may be developed for a specific Army MOS, Task 4 will develop measures that can be used across all MOS (i.e., Army-wide). The intent is to develop measures of first- and second-tour job performance against which all Army enlisted personnel may be measured. A major objective for Task 4 is to develop a model of soldier effectiveness that specifies the major dimensions of an individual's contribution to the Army as an organization. Another important objective of Task 4 is to develop measures of utility. It is critical to define, in dollar terms, the benefits likely to accrue from what will probably be more costly selection/classification procedures.

Task 5. Develop MOS-Specific Performance Measures

The focus of Task 5 is the development of reliable and valid measures of specific job task performance for a selected set of MOS. This task may be thought of as consisting of three major components: job analysis, construction of job performance measures, and construct validation of the new measures. While only a subset of MOS will be analyzed during this project, the Army may in the future wish to develop job performance measures for a larger number of MOS. For this reason, the methods are intended to apply to all Army MOS.

The Consortium/ARI Team

The initial project organization is shown in Figure 1. The principal consortium task scientists are shown, with their respective organizations, in the lower row. The principal ARI scientists are shown in the upper row. In the project consortium and ARI scientists undertake both independent and joint research activities. ARI scientists also have the administrative role of contract oversight.

During the course of this first year, the consortium's organization structure has remained stable. However, a number of significant personnel changes did occur. In June 1982, Dr. Ming-mei Wang was added to the AIR Task 1 staff and assumed the leadership of the Analysis Group. In July 1983, Dr. Joe Olmstead, after having completed his supervision of the work entailed in achieving the project's "Research Plan" and "Master Plan," asked to be relieved of his responsibilities in order to pursue other interests. Dr. Robert Sadacca assumed responsibilities as Task 4 Leader, on an acting basis, while replacement options were being evaluated.

Technical and management oversight is the responsibility of Dr. Newell K. Eaton, the contracting officer's technical representative (COR). On the project he is the ARI principal scientist, and has responsibility for technical review and guidance for the consortium scientists and managers,

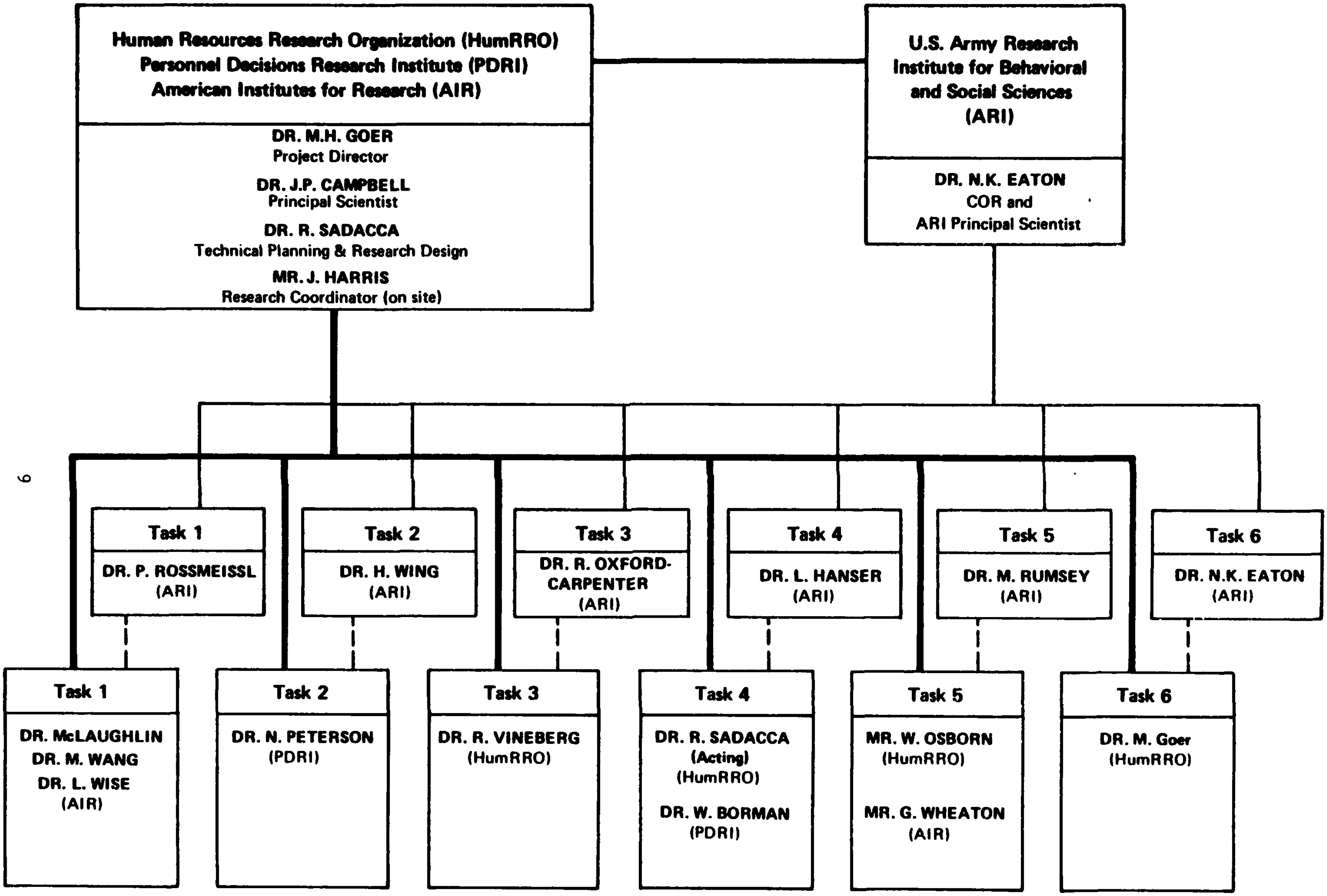


Figure 1. Project Organization

as well as for the ARI research teams. Consortium management is the responsibility of Dr. Marvin H. Goer. He provides management functions to include planning, coordinating, and integrating. Dr. Goer is assisted in his role as Managing Project Director (MPD) by Dr. John P. Campbell, Dr. Robert Sadacca, and Mr. James Harris. Dr. Campbell is the principal scientist responsible for overall scientific quality and for its state-of-the-art procedures. Dr. Sadacca is the assistant for technical planning and research design. In this role, he conceptualizes technical issues and integrates technical plans across tasks. Mr. Harris is the research coordinator on-site at ARI headquarters. As research coordinator, he conducts day-to-day liaison with the COR regarding Project A, Project B interactions, and related research.

A cooperative approach for accomplishing the best possible applied personnel research to meet the Army's initial needs in a collegial and joint consortium/ARI effort has characterized this first year's effort. The level and degree of cooperation and team effort that have been achieved already have been exemplary and have contributed materially to the successsful planning and start-up phases of the project.

The Advisory Group Structure

Because a program and project of this scale and importance would have to maintain close and active coordination with the other military departments, as well as with the Department of Defense, the project planners needed assurance that Project A was consistent with and complementary to the other on-going research programs being conducted in the research units of other armed services. The project also needed a mechanism for assuring that the research program met the highest standards for scientific quality and state-of-the-art technology in personnel selection and classification research. Finally, because it takes some time in a longitudinal research program to arrive at definitive answers to some questions, a method was needed to receive feedback from senior officers on priorities and objectives, as well as to identify current problems where an appropriate research focus would bring operationally useful early results. An effective mechanism was essential because the research program involved a large number of troops. Their commanders would require justification for use of those assets.

With the active assistance of Dr. Joyce L. Shields, Dr. John P. Campbell, and MG H. N. Schwarzkopf, advisory group participants were identified, commitments to participate were obtained, and the groups were established. Figure 2 shows the structure and membership of the Governance Advisory Group (GAG).

The Scientific Advisory Group (SAG) comprises nationally recognized authorities in psychometrics, experimental design, sampling theory, utility analysis, applied research in selection and classification, and in the conduct of psychological research in the Army environment.

The InterService Advisory Group (ISAG) comprises the Laboratory Directors for applied psychological research in the Army, Air Force, and the Navy, and the Director of Accession Policy from the DoD Office of Assistant Secretary of Defense for Manpower and Research Affairs.

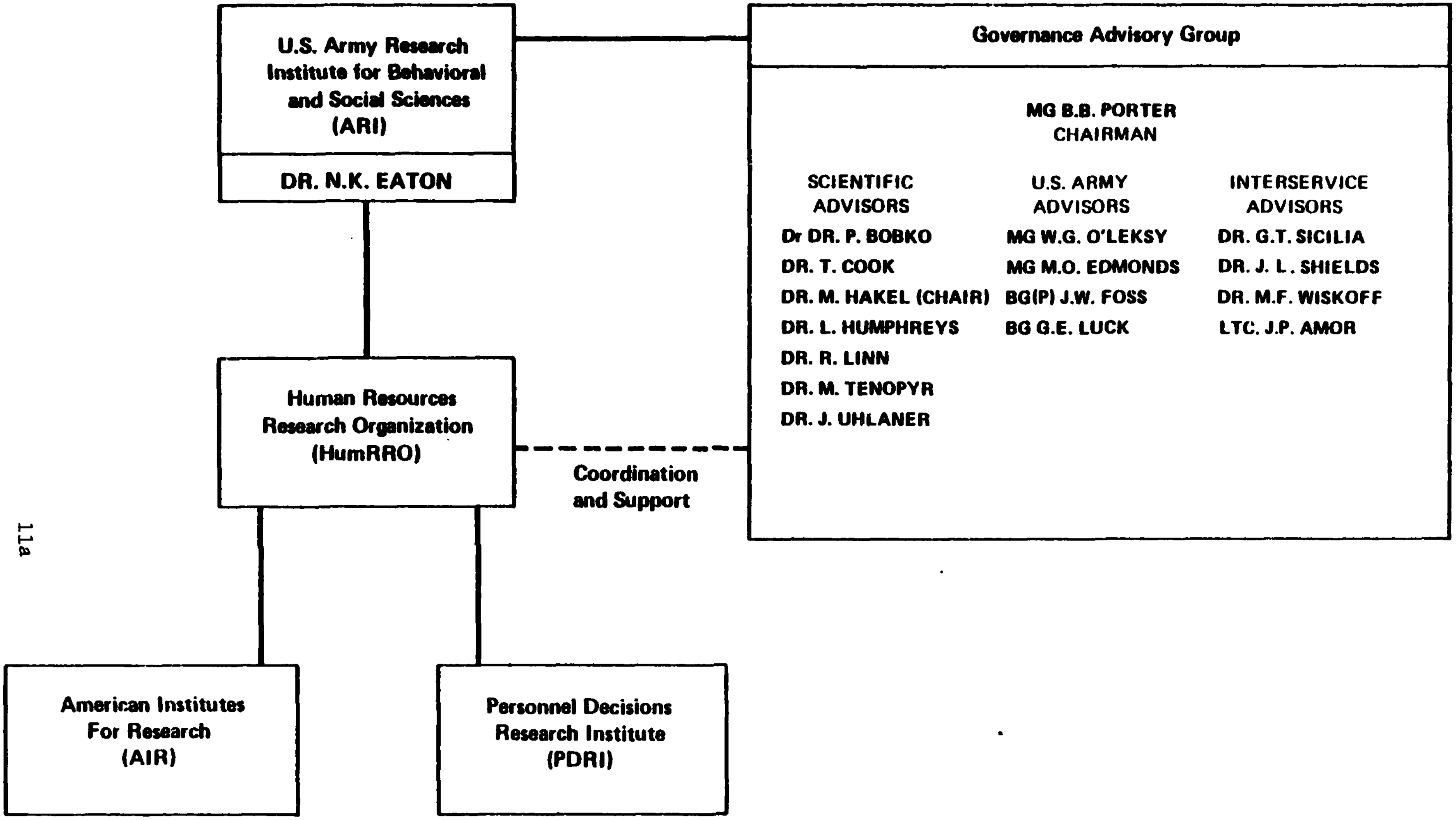


Figure 2. Governance Advisory Group

The U.S. Army Advisory Group includes representatives from the Office of Deputy Chief of Staff for Personnel (DCSPER), Office of Deputy Chief of Staff for Operations (DCSOPS), Training and Doctrine Command (TRADOC), Forces Command, (FORSCOM), and U.S. Army Europe (USAREUR). These senior officers have a significant interest in the project planning and priorities. They also represent the elements which provide the necessary and substantial troop support.

The Research Plan and Integrated Master Plan

The RFP stipulated that during the initial months of the contract a detailed Research Plan and an Integrated Master Plan for the project would be produced. The Research Plan would detail the specific substantive steps to be followed over the life of the contract. The Integrated Master Plan would provide detailed budget allocation, schedules, and product definitions.

Between the time the RFP was developed and the date of project start, a number of events had occurred that required incorporation into the Research Plan. For example, concurrent with the initial work on Project A, ARI had been asked by DoD to analyze and to provide recommendations for possible revisions in the construction of ASVAB 8/9/10, Aptitude Area Composite indexes. Project managers saw that this requirement could be met with the data already partially assembled on the FY81 cohort and that those data assets could be used to meet this priority request.

Having considered all such issues, the project staff turned to an accelerated schedule for the production of a revised Research Plan. A major constraint placed on the research planning was the mandatory requirement to meet both the intended and specific objectives of the research program mapped out in the original solicitation.

During the period January through April of 1983, an intensive period of replanning, documenting, review, modification, and redrafting of these critical documents occupied the consortium and ARI staffs. Drafts of the documents were provided to the SAG and ISAG. Their comments, provided orally during meetings and subsequently written in response to draft documents, were addressed and their suggestions were incorporated in the Research Plan. The culminating review was conducted in April by the U.S. Army Advisory Group, with representatives from the Scientific and Inter-Service Advisory Groups. In that meeting the entire research program, research design, sampling strategy, main cohort and focal MOS recommendation, and troop support implications were reviewed. Changes were incorporated to reduce and to distribute the troop support burden more equitably among the three participating commands (FORSCOM, TRADOC, USAREUR). The review provided valuable guidance and assistance in the practical issues of accomplishing the research activities in their organizations. Most importantly, the research program was endorsed by all three components of the GAG.

In May 1983, ARI issued Research Report 1332 "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel - Project A: Research Plan." In June 1983, the "Project A: Integrated Master Plan"

(HumRRO FR-PRD-83-8) was issued. The revised basic plan for the research program comprises these two mutually supportive documents.

During this entire period of planning, reviewing, and obtaining the endorsement of the Governance Advisory Groups, research activities by Project A consortium and ARI scientists were underway.

In this first year a number of significant research activities were initiated and significant progress was made. The following sections of this report summarize some of the most important efforts through the period ending 30 September, 1983. Associated research reports are included at the end of each section.

General Outcomes

The Project A Research Plan cited above speaks to the specific operational and scientific outcomes that will be produced. The outcomes that will flow from the project are characterized by the following themes:

(1) Project A will generate a broader and more complete sample of the predictor space than has ever been used before in a selection investigation. The taxonomy of predictors that is established will stand as a reference point for many years to come.

(2) Project A will provide the most thorough attempt ever made to develop standardized tests of actual task performance in skilled jobs. The procedure used will stand as a model.

(3) Project A will be by far the most thorough test to date of whether success in training predicts success on the job.

(4) Project A will provide a state-of-the-art model to illustrate how construct validity can be used to study applied problems in selection and performance assessment.

(5) Project A will be the first large selection and classification research effort to incorporate utility in the development of operational decision rules.

(6) Given the broad range of predictors, criteria, and jobs, Project A will be the most comprehensive test ever conducted on questions of different predictability across jobs, criterion measures, and predictor constructs.

Our overall conclusion at this time is that the objectives of Project A can be accomplished. We believe that the Project will make significant contributions to improve Army operational capability and to provide the most satisfactory careers for individual soldiers. Further, we expect that substantial scientific developments will result from this effort. While it will be time consuming and expensive, in our judgment, the benefits of this Project will be well worth the cost.

II. SAMPLE SELECTION

John P. Campbell (HumRRO)

Project A's large, complex requirements address simultaneously a wide range of interrelated research questions pertaining to an entire organizational personnel system. The overall objective in generating the samples has been to maximize the validity and reliability of the information to be gathered, while at the same time minimizing the time and costs involved. In part, costs are a function of the numbers of people in the sample. But costs are also influenced by the relative difficulty involved in locating and assembling the people in a particular sample, by the degree to which the unit's operations are disrupted by the data collection, by the staff costs involved in collecting the data in a particular manner, and by other such considerations. However, cost considerations cannot be used to compromise the validity and statistical reliability of the data to the point where the necessary research and development questions cannot be answered with confidence. We have tried to balance these considerations in as feasible and appropriate a way as the sampling plan was developed and implemented.

The sampling plan itself incorporated two principal considerations. First, a sample of MOS was selected from the universe of possible MOS; then the required sample sizes of enlisted personnel (EP) within each MOS were specified. The MOS are the primary sampling units. This is because Project A is developing a system for a population of jobs (MOS), but only a sample of MOS can be studied. Large and representative samples of enlisted personnel within each MOS are important because stable statistical results must be obtained for each MOS. There is a trade-off in the allocation of project resources between the number of MOS researched and the number of subjects tracked within each MOS: the more MOS are investigated, the fewer subjects per MOS can be tested, and vice versa. Cost versus statistical reliability considerations dictated that 19 MOS could be studied. To samples from all 19, we will administer the new predictors (from Task 2) and collect the school and Army-wide performance data (of Tasks 3 and 4). To nine of these MOS, we will also administer the MOS-specific performance measures developed in Task 5. The nine MOS were chosen to provide maximum coverage of the total array of knowledge, ability, and skill requirements of Army jobs, given certain statistical constraints.

MOS Selection

The selection of the sample of 19 MOS proceeded through a series of stages. An initial sample of MOS was drawn by using the following considerations:

(1) High density MOS that would provide sufficient sample sizes for statistically reliable estimates of new predictor validity and differential validity across racial and gender groups.

(2) Representative coverage of the aptitude areas measured by the ASVAB area composites.

(3) High priority MOS (as rated by the Army in the event of a national emergency).

(4) Representation of the Army's designated Career Management Fields (CMF).

(5) Representation of the jobs most crucial to the Army's mission.

The procedure entailed selecting a variety of CMF within strata of MOS density, given the following initial considerations:

(1) A data table was generated listing for each Army MOS the number of troops acquired in FY81 and the number of those who were female, Black, or Hispanic.[1] The CMF to which an MOS belonged was also listed.

(2) A first pass was made through this table searching for MOS which had at least 1,000 troops overall and a minimum of 300 women, 300 Blacks, and 100 Hispanics. This pass produced 11 MOS in eight CMF. The first eight MOS were identified by selecting the largest from each CMF.

(3) Next, the subgroup criteria were further relaxed by eliminating the requirement for Hispanic representation. This produced four additional MOS, but all were in CMF already present in the initial set of eight. On those grounds, all four were eliminated from further consideration.

(4) Again the criteria were changed, this time by eliminating the requirement for female representation but restoring the minimum requirement for 100 Hispanics. Against these constraints, eight new MOS surfaced representing four new CMF. Four MOS were added to the initial set of eight by retaining the largest in each new CMF.

(5) A final charge in criteria was made in which the total accessions constraint was reduced from 1,000 to 500 and all requirements for minority representation were dropped. An additional 29 MOS in 14 CMF emerged. Seven of these 14 CMF were represented in the set of 12 MOS already selected. Of the remaining seven, one--CMF 98, Intelligence--was dropped because it was classified. That left eight MOS in six CMF. The largest MOS in each of the six remaining CMF was chosen, increasing our sample to 18.

A further indirect indication of the mix of job skills represented in the sample is in the range of ASVAB composites and component subtests pertinent to each MOS. All subtests and all but one (EL) of the nine composites were

[1]FY81 accessions data were available. It was assumed that those data would represent reasonably well the relative distribution over MOS of accessions in FY83 and later.

represented in the 18 MOS initially selected. Consequently, a 19th MOS (27E) was chosen to represent the EL aptitude composite.

The composition of the sample was also examined from the standpoint of mission criticality by comparing it with a list of 42 MOS identified by the Army as high priority for mobilization training.[2] The 42 MOS represent 17 CMF, 13 of which are contained within our set of 19. Of the four not in our sample, two are classified (CMF 96 and 98) and two are small (CMF 23 and 84). The six CMF in our sample not in the mobilization training priority list generally represent jobs for which there are civilian counterparts, a type of job purposely excluded from the mobilization list.

This initial set of 19 MOS represent 19 of the Army's 30 CMF. Of the 11 CMF not represented 2 are classified (CMF 96 and 98), 2 (CMF 33 and 74) have fewer than 500 FY81 accessions, and 7 (CMF 23, 28, 29, 79, 81, 84, and 74) have fewer than 300 FY81 accessions. The initial set includes only 5 percent of Army jobs but 44 percent of the soldiers recruited in FY81. Similarly, of the 15 percent women in the 1981 cohort, 44 percent are represented in the sample; of the 27 percent Blacks, 44 percent are represented in the sample; and, of the 5 percent Hispanic, 43 percent are represented. Although female and minority representation are high absolutely, relatively it remains about the same as in the population. The sample is 15 percent female, 27 percent Black, and 5 percent Hispanic.

Nine of the 19 MOS were tentatively earmarked for the job-specific performance measurement phase of the project. These were selected as a

[2]ODCSOPS (DAMO-ODM), DF, 2 Jul 82, Subject: IRR Training Priorities.

subset with the same general criteria used in identifying the parent list of 19. Since the larger list is composted of five combat and 14 noncombat MOS, it seemed reasonable that these categories were proportionally represented in the subset of nine. To keep travel and field performance measurement costs within bounds, only the largest MOS were selected. So the three large combat MOS--11B (Infantryman), 13B (Cannon Crewman), and 19E/K (Tank Crewman)--were selected first. Of the 14 noncombat MOS, 8 are large and have race and gender subgroups substantially represented. Since five different ASVAB composites are represented among the eight, one MOS was selected for each. Both 64C (Motor Transport Operator) and 94B (Food Service Specialist) share the OF aptitude composite and are roughly the same size, but the former was chosen because it is considered a priority MOS for mobilization. The two clerical (CL) MOS differ neither in size nor in their mobilization priority status, so 71L (Administration Specialist) was chosen over 76Y (Unit Supply Specialist) chiefly because it has more women. Both MOS with the ST composite were selected, since both have priority mobilization status. Thus, the nine MOS designated for hands-on performance measurement development are:

(1) 11B - Infantryman

*(2) 13B - Cannon Crewman

(3) 19E/K - Tank Crewman

(4) 05C - Radio TT Operator

(5) 63B - Vehicle and Generator Mechanic

*(6) 64C - Motor Transport Operator

*(7) 71L - Administration Specialist

(8) 91B - Medical Care Specialist

*(9) 95B - Military Police.

An initial batch of four (see asterisks preceding) was selected and designated as Batch A; the other five, as Batch B. Work has begun on Group A first. Batch B will be taken up in turn.

On the basis of guidance from the Scientific Advisory Group, further refinements of the MOS sample were undertaken. These included a cluster analysis of expert ratings of MOS similarity and a review of the initial sample by the Governance Advisory Group.

Cluster Analysis

To obtain data for empirically clustering MOS on the basis of their task content similarity, a brief job description was generated for each of 111 MOS from the job activities described in AR 611-201. The sample of 111 MOS represents 47 percent of the population of 238 Skill Level 1, Active Army MOS with conventional ASVAB entrance requirements and includes the 84 largest MOS (300 or more new job incumbents yearly) plus an additional 27 selected randomly but proportionately by CMF. Each job description was limited to two sides of a 5-x-7 index card.

Members of the contractor research staff and ARI Army officers--approximately 25 in all--served as expert judges and were given the task of sorting the sample of 111 job descriptions into homogeneous categories based on perceived similarities and differences in job activities as described in AR 611-201. Data from the similarity scaling task were clustered and the initial results used to check the representativeness of the initial sample of 19 MOS, that is, did the initial sample of MOS include representatives from all the major clusters of MOS derived from the similarity scaling? On the basis of these results and guidance received from the Governance Advisory Group, two MOS that had been selected initially were replaced by 51B and 27E, which are in the same CMF and involve the same Aptitude Area Composites as the replaced MOS (62E and 31M).

The sample of MOS resulting from the above procedures is shown in Table 1.

Sampling Enlisted Personnel Within MOS

There are two major considerations relative to sampling individuals within MOS. One concerns the number of people per MOS and the other deals with the schedule or sampling plan for obtaining the data from the enlisted personnel serving as research subjects. The sampling plan, or design, is dictated by the research questions and the kind of information that is needed to answer them. The sample size within MOS is a function of the number of individuals needed for statistical reliability and the amount of sample attrition that must be allowed to obtain such a sample size.

The overall design of Project A is described in detail in the Final Research Plan (June 1983). Briefly, the overall objectives are to develop and validate an experimental battery of new and improved selection measures against a comprehensive array of job performance and training criteria. The validation research must produce sample estimates of the parameters necessary to implement a computerized selection and classification system for all first-tour enlisted MOS. To do this a design that uses two

Table 1

Project A MOS

MOS	Title	CMF	APT Comp	Priority MOS	Total	FY81 Accessions			Trainee Projections		Expected Number Graduates[1]
						Women	Blacks	Hispanic	FY83	FY84	
05C	Radio TT Operator	31	SC	No	3175	585	898	119	2004	2200	1645
63B	Vehicle & Generator Mech	63	MM	No	4653	386	1178	242	5304	4402	4280
64C	Motor Transport Operator	64	OF	Yes	5440	774	1279	141	3706	5000	4484
71L	Admin Specialist	71	CL	No	4484	2744	1967	215	6191	4592	3859
13B	Cannon Crewman	13	FA	Yes	5783	0	2053	367	6092	3553	3572
91B	Medical Care Specialist	91	ST	Yes	3074	924	876	224	3761	unav	3621
19E/K	Tank Crewman	19	CO	Yes	3233	0	604	188	3223	3261	2912
95B	Military Police	95	ST	Yes	6073	704	624	127	5720	5300	4373
11B	Infantryman	11	CO	Yes	7028	0	1128	367	12633	13710	11338
76Y	Unit Supply Specialist	76	CL	No	4565	1179	1998	283	6636	4091	3829
94B	Food Service Specialist	94	OF	No	3859	715	1416	125	5133	5157	4600
12B	Combat Engineer	12	CO	Yes	3707	0	716	147	844	2540	1845
16S	MANPADS Crewman	16	OF	Yes	691	0	206	27	797	1015	815
55B	Ammunition Specialist	55	GM	No	662	171	283	42	620	810	762
76W	Petroleum Supply Spec	92	CL	NO	849	259	559	43	1373	1350	1234
54E	Chemical Operations Spec	54	ST	Yes	557	89	185	41	1012	1247	1068
67N	Utility Helicopter Rpr	67	MM	No	1032	33	68	29	572	465	470
51B	Carpentry/Masonry Spc	51	GM	No	602	6	136	14	120	483	341
27E	Tow/Dragon Rpr	27	EL	No	333	40	76	17	312	308	258
	Total				59800	8609	16001	2758	60052	59511	55306

[1] Weighted average of Trainee Projections (3 months of FY83 and 9 months of FY84) adjusted for expected school attrition (actual FY81 rates).

predictive and one concurrent validation on two major troup cohorts (83/84 accessions and 86/87 accessions) was developed.

A schematic of the data collection plan is shown in Figure 3.

Collection of New Data Within MOS

Five major collections of new data will furnish much of the information to be used to answer the specific questions posed in the research plan.

Data Collection 1

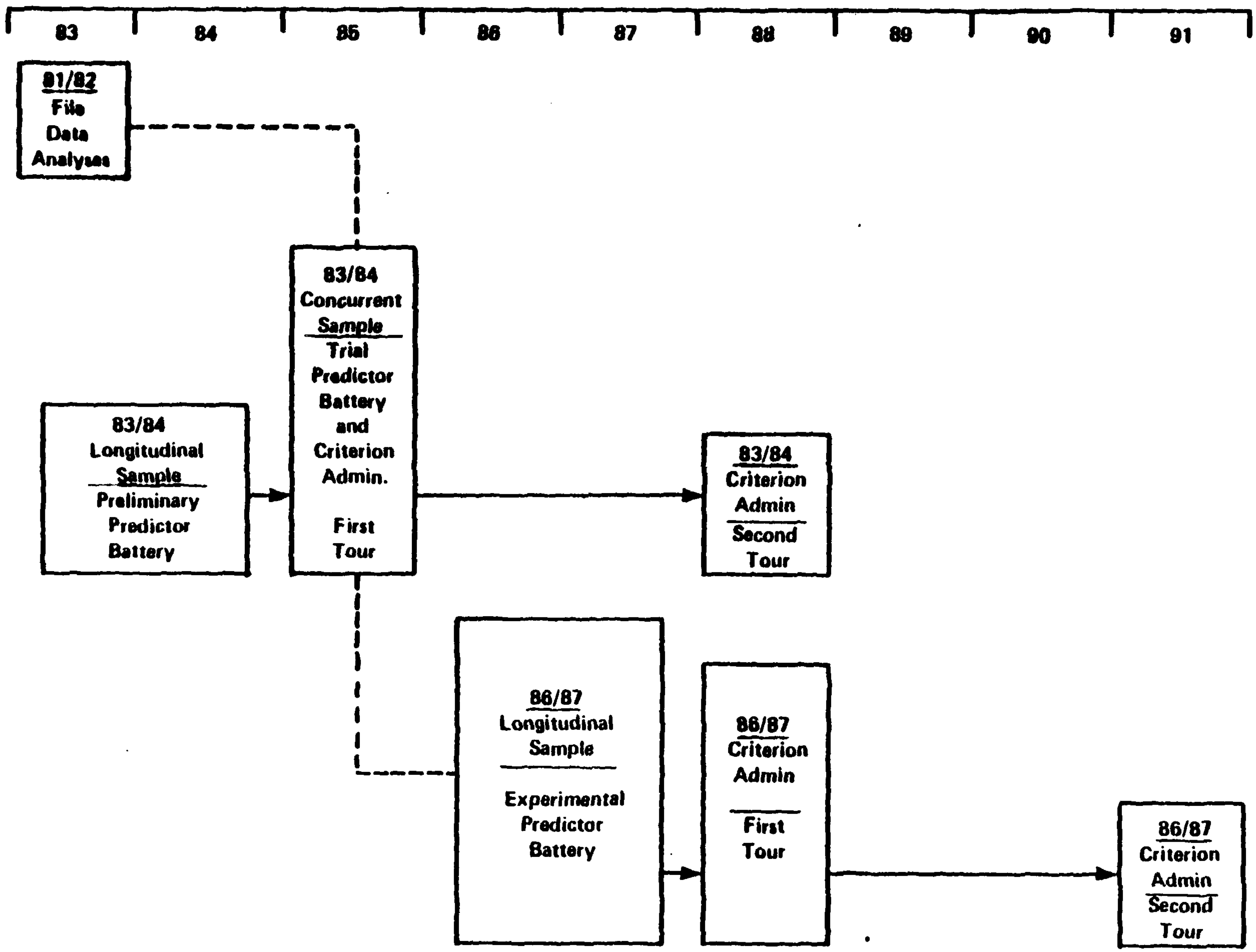
This first major data collection follows a longitudinal design. New recruits will be tested with a preliminary predictor battery, already developed, beginning in October 1983 and continuing until the summer of 1984. The recruits will be sampled from 4 MOS (05C, 19E/K, 63B, 71L). The principal criterion data will be training school achievement measures.

Data Collection 2

The collection of data on new predictors, job knowledge tests, and the Army-wide and MOS-specific performance meaures will be accomplished in a large field administration of these instruments on the FY83/84 cohort-first tour during 6/85-10/85. The target will be to collect data on an average of 500 enlisted personnel (EP) in each of the 19 MOS identified earlier.

Figure 3

The Overall Data Collection Plan



Data Collection 3

A longitudinal prediction sample will be collected from the FY86/87 cohort by testing recruits with the revised predictor battery and obtaining school data beginning in March of 1986 and continuing until February 1987. Recruits will be sampled from the 19 focal MOS. Because this sample will be followed up for purposes of collecting criterion information once during 1988 (first tour) and again during 1991 (second tour), the expected attrition in the sample will be considerable. The expected attrition for a typical MOS is shown in Figure 4. This dictates that it is highly desirable that about 2,200 recruits be tested from each MOS on the average. There will most likely not be that many accessions per year for all MOS. In MOS with fewer accessions, we need to obtain as many of the available recruits as possible.

Data Collection 4

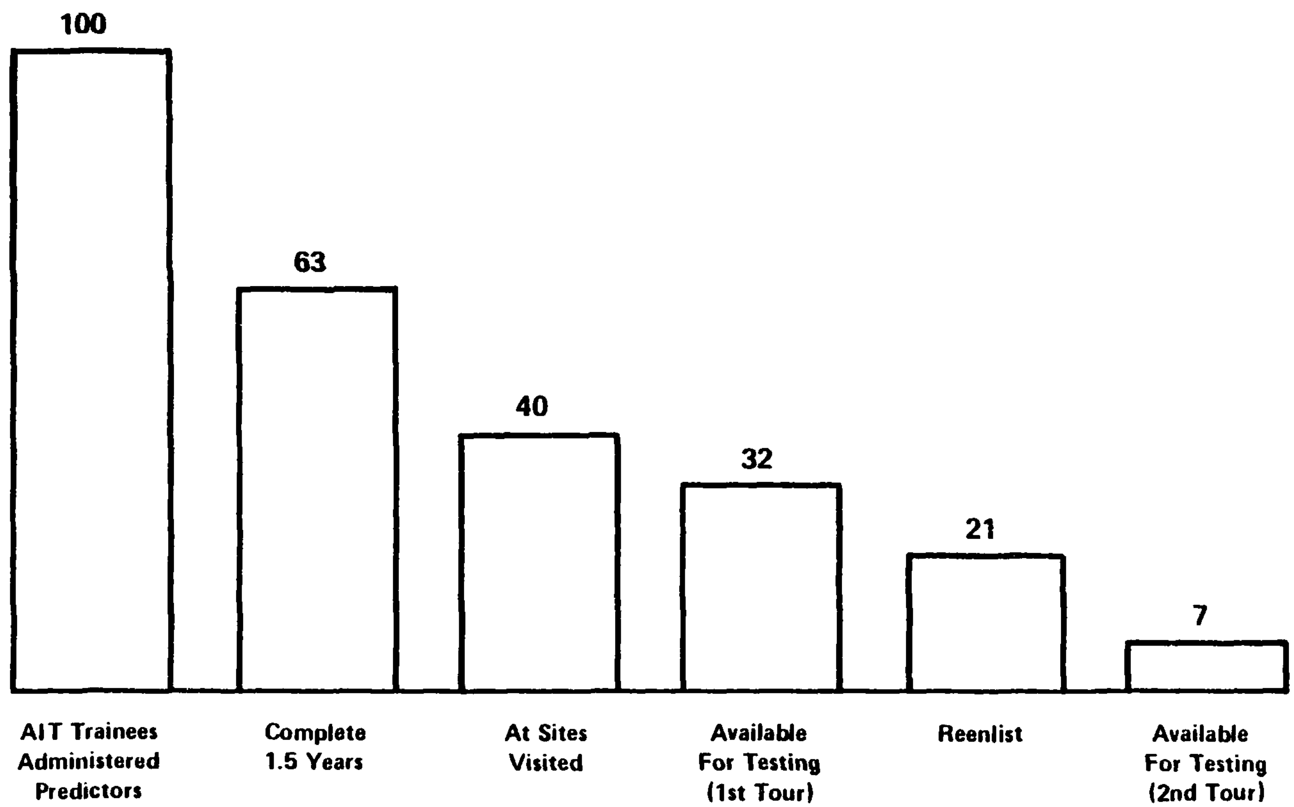
During the period June 1988 through September 1988, Army-wide and MOS-specific performance measures will be collected at 12 to 15 sites from the FY83/84 cohort which will be in its second tour and the FY86/87 cohort which will be in its first tour.

Data Collection 5

From January 1991 to March 1991, Army-wide and MOS-specific criterion data will be obtained from the FY86/87 cohort which will be in its second tour.

Figure 4

Percent of Troops Available for Performance Measurement Research in a Typical MOS



Sample Size as a Function of Attrition Within MOS. There is considerable attrition from the sample as each cohort moves through its tour. The attrition can be summarized by the following points:

(1) A certain percentage of recruits who begin AIT will not finish. Attrition during training is not random, either by MOS or by ability level within MOS.

(2) Of those who finish their AIT, a certain percentage will attrite during the first 1-2 years of their tour.

(3) Since, for purposes of this project, the criterion assessment of people must take place on a relatively small number of installations, not all the sample will be found on those bases (some will be scattered across a much larger number) and a further reduction in the sample will occur.

(4) It is also true that during a given time period, at a given base, not all of the people in the sample will actually be available for testing (e.g., due to leaves, illness, etc.) and additional shrinkage in the sample will occur.

(5) Only a small proportion of the original sample will reenlist and be available for the second tour measures.

(6) Of those who reenlist, only a certain percentage will be on the bases where the testing is taking place at any designated time and will be available for testing.

(7) The attrition rates over the various stages in a soldier's tour, from AIT to reenlistment, are not the same for all MOS. In fact, they vary a great deal, which makes the process of sample selection difficult.

Estimates on attrition and sample shrinkage for the MOS listed in Table 1 are shown in Table 2 and Figure 4. The estimates are based on actual figures for previous or current accessions. As such, they constitute our best estimate of how these decay functions will look in the future. The

Table 2

Estimated Requirements for
Measurement of 83/84 Cohort by MOS

MOS	Title	Projected Input	% Complete 1st Tour	% Re-Up	% at Sites	Concurrent Validity		Predictive Validity		
						No. Tested at Sites (1st Tour Performance Predictors)	No. Tested at Sites (2nd Tour Performance)	No. Tested at AIT (Predictors)	No. Tested at Sites (1st Tour Performance Predictors)	No. Tested at Sites (2nd Tour Performance)
05C	Radio TT Operator	2150	57	16	37			2150	360	100
63B	Vehicle & Generator Mech	4630	63	08	41			4065	650	100
64C	Motor Transport Operator	4370	64	21	37	825	80			
71L	Admin Specialist	4990	63	26	28			4575	650	100
76Y	United Supply Specialist	4725	62	28	38	735	100			
91B	Medical Care Specialist	3760	65	25	33	645	65			
94B	Food Service Specialist	5150	55	20	39	865	100			
95B	Military Police	5405	70	20	19	575				
11B	Infantryman	13,440	58	17	50	860	100			
12B	Combat Engineer	2115	68	17	42	485				
13B	Cannon Crewman	4190	59	18	52	785	100			
16S	MANPADS Crewman	960								
19E/K	Tank Crewman	3250	61	17	49			2720	650	100
55B	Ammunition Specialist	765	61	26	44	165				
76W	Petroleum Supply Spec	1355	64	23	43	295				
54E	Chemical Operations Spec	1210	59	21	42	240				
67N	Utility Helicopter Rpr	490	76	25	43	130				
51B	Carpentry/Masonry Spec	392	70	10	40	90				
27E	Tow/Dragon Rpr	309	66	24	45	75				

initial samples that are required can then be generated by working backward from the sample sizes that are necessary to provide a minimum level of statistical reliability at the crucial data collection points. The specific sample sizes for each MOS for each major data collection were generated in this way.

The required sample sizes within MOS may seem large; however, it is dictated by the following considerations:

(1) The overriding goal is to develop a comprehensive selection and classification system that will be implemented across all nonclassified enlisted MOS that are associated with advanced instructional training. Consequently, the different parts of the system cannot be studied piecemeal. If the system and connections are not studied as a whole, it will not be possible to develop the optimal set of preinduction tests, performance measures, and algorithms that link the parts. We must have a large amount of information on each person and this means that sample sizes must be large to insure statistical reliability.

(2) It is necessary to examine the differences in regressions, correlations, and other statistical indexes between gender groups, racial groups, MOS, etc. As has been frequently demonstrated, testing differences between regression and/or correlation coefficients requires very large sample sizes.

(3) It is necessary, for implementation of the selection and classification system, to draw conclusions about the level of validity for each MOS. Thus, each MOS that is included in the sample of MOS must have a sufficient sample size to make reliable statistical conclusions that can be generalized to the population of MOS.

The FY81/82 Cohort as a Baseline

In addition to collecting data from new samples, the project is making use of existing file data that have been, or can be, accumulated for 1981 and

1982 accessions. The editing and merging of data from the accessions and EMF files for entry into the Longitudinal Research Data Base (LRDB) is now virtually complete and ready for analysis. That work will be described in greater detail in the next chapter of this report.

There are several factors that argue for an extensive analysis of the available file data for the FY81/82 cohort:

(1) These are the best data currently available for evaluating the validity of the current form of the ASVAB 8/9/10.

(2) The file data can be used as a benchmark with which to compare the incremental validity generated by Project A. That is, for the current predictors and the available criteria, subtest validities, composite validities, differential validity across groups, e.g., race, and differential validity across MOS, i.e., validity generalizeability, can be determined. The question is then how much these indexes change when the new experimental battery is tried out with the broader range of criteria.

(3) The FY81/82 cohort is allowing us to try out a number of new analytic techniques so as to determine if they will be useful in later phases of the project. For example, simultaneous estimation techniques could be used to determine how many significantly different prediction equations are needed to predict criterion scores in different MOS.

Summary

During the first year the focal MOS were selected, the sample sizes required from each were specified, and the troop support requests were prepared. In addition, the available computer file data on the FY81/82 cohort were merged from the various sources, were throughly edited, and were made ready for analysis. Although the troop support requirements may seem large, they are made necessary by requirements of the selection and classification system to be developed. A series of smaller efforts over a longer period may indeed be more expensive in the end, and it would not produce the necessary data that Army management could use with confidence.

Associated Report

As indicated in the Introduction, a relevant research report follows.

Grouping Army Occupational Specialties by Judged Similarity[1] *

Rodney L. Rosse and Walter C. Borman
Personnel Decisions Research Institute

Charlotte H. Campbell and William C. Osborn
Human Resources Research Organization

PURPOSE

The main purpose of the present research was to cluster U.S. Army enlisted jobs Military Occupational Specialties (MOS) into homogeneous groups according to rated job content. This is done to provide relevant information in the selection of an MOS sample which is representative of Army MOS. The sample of MOS is, in turn, to be employed in selection and classification research on enlisted army jobs.

BACKGROUND

Motivation for the Study. For personnel selection work, cost prohibits empirically validating predictors against expensively measured criteria of job performance for every single MOS. Instead, predictors may be validated against criteria intensively measured for each of a sample of jobs chosen to represent the occupational domain. Then, prediction equations must be generalized to other jobs by linking profiles of similar job characteristics. This approach underlies the Army's approach to remodeling its personnel selection and classification system (Eaton & Shields, 1982; HumRRO, AIR, PDRI, & ARI, 1983). This research pertains to the problem of choosing a representative sample of MOS.

Practical considerations led to the decision to focus on a sample of 19 MOS (about 8% of the 238 entry-level jobs). In addition to spanning a range of job skills representative of the occupational domain in the Army, size (number of incumbents) and minority/gender representation were factors in selection of MOS since sufficient numbers are needed for validation research.

Job Clustering. It has been stressed that the choice of job content descriptors in effect defines the way that the system may be used (Cornelius, Carron, & Collins, 1979; Dunnette, 1976). Four commonly used types of content descriptors are (1) job-oriented statements (usually task or observable activity statements), (2) worker-oriented statements (more general statements of overt and inferred human behaviors), (3) aptitude requirements, and (4) global descriptions (job titles or general descriptions).

Eventually, prediction equations derived for the sample of 19 MOS are to be generalized to all other specialties through a network linking specialties according to similarity of performance requirements. Accordingly, job-oriented task statements seemed to be most appropriate in this context. The idea would be to obtain judgments about the importance of a relatively large number of tasks related to performance in each MOS for clustering MOS with similar profiles of task importance.

At this point we did not have comprehensive task lists for each Army MOS. Thus, the other three approaches were further scrutinized for our purposes.

Fortunately, the clustering of global job descriptions has yielded results quite similar to those obtained by grouping jobs based on patterns of importance on individual tasks within those jobs (i.e., the job oriented/task statement approach).

Sackett, Cornelius, and Carron (1981) had judges compare foreman jobs in a chemical processing plant through a whole-job paired comparison procedure. They also asked judges to rate, for each job, the importance of each of 237 tasks for performing the job. Results of cluster analysis showed that identical job families would have been

---

1. This research was funded by Army Research Institute Contract No. MDA 903-82-C-0531. All statements expressed in this paper are those of the authors and do not necessarily express the official opinions or policies of the U. S. Army Research Institute or the Department of the Army.

*Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983.

formed using the two methods. Thus, the whole-job strategy should yield results similar to those that would be obtained using the job-oriented approach.

As a workable solution, we chose to use a simple sorting task which requires judges to sort jobs into categories according to similarities in content. The method has been used to generate clusters of Army recruiter tasks (Borman, Toquam, and Rosse, 1977).

METHOD

Deriving the MOS Set for Clustering. Army Regulation 611-201, Enlisted Career Management Fields and Military Occupational Specialties, contain summary descriptions of all Army occupational specialties (approximately 350). Eliminating those having no entry-level job, pertaining to the Reserve force only, being phased out, or having entrance requirements based on something other than ASVAB, produces 238 MOS in 28 Career Management Fields (CMFs are clusters of jobs in the Army's personnel management system).

Availability of time from expert judges dictated further reduction in the domain. Pilot tests indicated that raters could comfortably and reliably sort about 100 MOS in a two-hour period. A two-step procedure brought the MOS list to near that target. First, the 84 large MOS (annual minimum of 300 new job incumbents during FY 83-84) were selected. Then, 27 MOS were added so that every CMF was proportionately represented by at least one-third of its MOS for a total of 111 MOS from all CMFs. All large MOS were included along with several more specialized MOS.

Job Descriptors and the MOS Sorting Protocol. Descriptions of the Skill Level 1 job requirements for the 111 MOS were taken from AR 611-201. The MOS title, one-sentence MOS description, and task requirement paragraph were printed on a 5x9 card for sorting; no MOS or CMF number identifiers were included. The sorting protocol instructed respondents to sort MOS according to similarities in performance requirements. A pretest was conducted, the directions changed slightly in response to pretest results, and the protocol was readied for the main administration.

Data Collection. A sample of 25 raters completed the sorting: 17 experienced psychologists on the contractor's staff, and 8 field-grade Army officers on the administrative staff of the U. S. Army Research Institute.

The sorting task was essentially self-administered. A deck of the 111 randomly-ordered job cards and instructions were presented to each rater. Each was urged to complete the sorting in one session.

Data Analysis: Generating Similarities Matrices. Data acquired from the 25 raters were used to generate 111 by 111 matrices of elements denoting pair-wise similarity: (1) raw similarities, and (2) correlations.

Raw Similarities. Each element in the raw similarity matrix was derived by computing the proportion of raters placing both MOS in the same category during the sorting task. Thus, for example, if 20 of the 25 raters put MOS 1 and 2 in the same category, the raw similarity for that pair would be .80.

Correlational Similarities. The correlational approach to developing similarity indices is akin to a strategy proposed by Rosenberg and Sedlak (1972). In their approach, a matrix of dissimilarities is formed from the elements of the similarity matrix by computing the Euclidean distance of pairs of rows (or columns) in the matrix. The reasoning behind this index is that both direct and indirect similarities are relevant for assessing similarity/dissimilarity between (in this case) MOS. With the correlational method, for a given pair of MOS (say, A and B), indirect similarity is computed by correlating the vectors of the aggregated raw similarities between, respectively, MOS A and each other MOS and MOS B and each of these same other MOS. Correlational indices of similarity were computed in this manner for all pairs of MOS.

Data Analysis: Evaluating Reliabilities of the Similarity Matrices. To provide an index of interrater reliability, raw similarity matrices were first computed for each of the 25 raters. This was accomplished by placing a 1 in a matrix cell when the rater sorted the two MOS into the same category and a 0 when the two appeared in different categories in the rater's solution. Then, the (111 x 110)/2 = 6105 off-diagonal elements of each rater's similarity matrix were correlated with the corresponding elements of each other rater's matrix, giving rise to a 25 by 25 matrix of correlations. The off-diagonal correlations ranged from .15 to .72 with mean of .39 (which was taken to be an estimate of the reliability of a single rater). Using the Spearman-Brown Prophecy formula to adjust the reliability for 25 raters, the resulting agreement index is .94.

Data Analysis: Use of Orthogonal & Oblique Factor Analysis Methods. An orthogonal, principal factor solution of the matrix of correlations was rotated according to Kaiser's Normalized Varimax Criterion. The sole purpose of the orthogonal factor solution was to provide summary information for assignment of MOS to clusters. This was a judgmental process involving inspection of the MOS content as well as the factor pattern matrix. In assigning MOS to clusters, attention was given to maintaining close similarity in the patterns of loadings across factors for the MOS included in each cluster.

To provide a comprehensive and useful summary of the properties of the final cluster solution, the clusters of MOS were then used to define the axes of an oblique, cluster centroid factor analysis, a method given in Overall and Klett (1972, chap. 7).

This method of summarizing the cluster assignments has some desirable properties. One is that the relationship of each MOS to each cluster can be examined. If the cluster assignments are satisfactory, it is expected that each MOS would not only load heavily on the oblique factor to which it is assigned, but also would show a pattern of loadings across all other factors similar to all other MOS assigned to the same cluster.

RESULTS

The Orthogonal Factor Solution[2]. Fifteen factors were extracted from the matrix of intercorrelations among MOS. This solution was selected for the first attempts at assignment of MOS to clusters because solutions with greater numbers of factors appeared to be residual factors.

The rotated, 15-factor solution was used as the basis for cluster assignment decisions. Most of the decisions were made on the basis of similarity (or lack of similiarity) in the patterns of loadings across the 15 factors. Some clusters with only one MOS were assigned because of low communality, implying lack of common properties with any of the other 111 MOS.

Judgment of the adequacy of a cluster solution is, in large part, dependent upon the amount of detail desired in the solution. In this case, it was decided that reasonably fine distinctions might be useful, and thus, 23 clusters of MOS were tentatively identified.

The Oblique Factor Solution: Final Summary of Clustering. Table 1a contains the factor pattern matrix resulting from the cluster centroid factor analysis. For this solution, the 23 clusters of MOS, identified on the basis of the orthogonal factor analysis results (along with judgment), were used to define the oblique axes. Thus, the 23 cluster structure provides what we believe to be meaningful groupings of MOS based on perceived similarity in job content. Table 1b contains the primary factor cosine matrix, i.e., the correlations between oblique factors.

Thus, the orthogonal, 15-factor solution guided attempts to group MOS according to similarities in factor loadings. After the clusters were formed, the oblique factor analysis method (Overall and Klett, 1972) summarized the cluster structure by providing loading indices for each MOS against each cluster.

DISCUSSION

The MOS within each cluster appearing in Table 1 have patterns of loadings that are remarkably similar. This suggests that the 23-cluster solution is a reasonable one, and it was used to help guide selection of MOS for examination in Project A.

---

[2]Space limitations precluded presentation of the factor solution here. Interested readers may request copies from the first authors.

Essentially, the goal is to represent with a sample of 19 MOS the entire population of 111 large Army MOS. The clusters in Table 1 did help in selecting MOS. Referring to the table, the starred MOS are the original 19 selected. Notice that there is some over-representation in that some clusters have two or even three MOS of the 19. Also, there is under-representation in that some of the clusters are not represented by even a single MOS.

Two substitutions of MOS to some extent ameliorated the over and under-representation. The MOS Tow/Dragon Repairer and Carpentry/Masonry Specialist were substituted for Heavy Construction Equipment Operator and Multichannel Communications Equipment Operator. The two new MOS are both in clusters that originally were not represented and the two MOS removed from the list were in clusters that previously contained more than one.

Although this cluster solution is seen as sensible from what is known about the job content of the 111 MOS, one possible weakness in the method is that MOS might have been grouped primarily by MOS name rather than by content. The job descriptions were relatively short, but there still may have been a tendency for judges to "short-cut" the sorting task by forming groups according to MOS names. Another reason for this possibility is that judges were unfamiliar with many of the MOS. Lack of familiarity may also tend to make judges rely on the MOS names when sorting.

The potential sort-by-name problem could be addressed by asking (for each MOS) persons who are very knowledgeable about the MOS to complete a job survey describing the job content (e.g., the importance or time spent on various tasks) and/or perhaps the ability-skill requirements of the MOS. Provided that each expert completed the same survey items, responses for pairs of MOS could be correlated to index their similarity in terms of content and/or ability or skill requirements. The correlation matrix could then be factored using the same methods employed in the present research. This would seem to be a better procedure for generating between-MOS similarity data; however, it would also require many more judges than the 25 used here, and in defense of the sorting protocol, the similarity matrix (from which the correlation matrix was derived) was quite stable, with the reliability of the mean similarities across the judges equal to .94.

REFERENCES

Borman, W. C., Toquam, J. L., & Rosse, R. L. Dimensions of the Army recruiter and guidance counselor job (TR-77-A5M). Alexandria, Virginia: U. S. Army Research Institute for the Behavioral and Social Sciences, March, 1977.

Cornelius, E. T., Carron, T. J., & Collins, M. N. Job analysis models and job classification. Personnel Psychology, 1979, 33, 110-120.

Dunnette, M. D., Aptitudes, abilities, and skills. In M. D. Dunnette (Ed.), Handbook of industrial and organizational. Chicago: Rand McNally, 1976.

Eaton, N. K., & Shields, J. L. U. S. Army soldier selection, classification, and utilization research program. Paper presented to the 90th Annual Convention of the American Psychological Association, August, 1982.

Human Resources Research Organization, American Institute for Research, Personnel Decisions Research Institute, U. S. Army Research Institute. Improving the selection, classification and utilization of Army enlisted personnel. U. S. Army Research Institute Research Report 1332, May, 1983.

Overall, J. E., & Klett, J. C. Applied multivariate analysis. New York: McGraw-Hill, 1982.

Sacket, P. R., Cornelius, E. J., & Carron, T. J. A comparison of global judgment versus task-oriented approaches to job classification. Personnel Psychology, 1981, 34, 791-804.

Table 1a. Primary (Oblique) Factor Pattern Matrix:
Factors Defined by MOS's in 23 Group Interpretation of 15 Factor Orthogonal Solution.

MOS NR.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
004	99	75	-38	-13	38	-25	-29	02	41	-24	-30	-33	-14	-21	-17	-20	-02	-32	-13	59	05	-08	-28
011	96	72	-38	-14	36	-26	-30	06	46	-25	-31	-33	-10	-21	-17	-20	-03	-33	-14	57	03	-07	-28
017	88	87	-37	-15	31	-25	-27	-08	33	-18	-28	-22	-16	-16	-11	-21	-05	-31	-12	54	08	-12	-12
016	93	62	-38	-12	40	-26	-29	06	45	-23	-26	-30	-14	-17	-15	-17	-03	-32	-10	63	05	-05	-27
027	96	87	-38	-10	37	-26	-30	-02	36	-26	-33	-35	-12	-22	-18	-22	03	-35	-17	52	02	-09	-30
042	99	88	-39	-11	43	-25	-29	-02	37	-25	-31	-33	-15	-21	-17	-20	-00	-32	-11	60	07	-09	-29
045	97	72	-38	-10	38	-26	-30	08	49	-26	-32	-34	-14	-22	-18	-21	02	-34	-17	54	01	-08	-29
051	97	60	-37	-15	34	-26	-30	04	42	-26	-32	-34	-01	-22	-18	-21	-04	-34	-14	57	02	-06	-29
072	99	74	-38	-12	40	-25	-29	-00	38	-25	-31	-33	-11	-21	-17	-20	-01	-33	-13	59	05	-07	-29
075	86	38	-34	-25	20	-28	-30	11	45	-12	-19	-10	-07	-09	-08	-20	-15	-34	-18	48	-14	20	-15
082	94	82	-41	-08	43	-26	-30	-04	33	-26	-33	-36	-16	-22	-18	-21	15	-34	-13	56	09	-11	-30
084	87	69	-39	-08	56	-27	-31	-06	28	-21	-24	-28	-17	-22	-17	-20	08	-33	-06	53	18	-09	-26
090	99	78	-40	-07	42	-27	-31	-00	38	-27	-33	-36	-12	-23	-19	-21	04	-34	-12	58	09	-09	-31
091	92	77	-41	-07	44	-27	-31	-03	31	-27	-34	-36	-12	-23	-19	-22	18	-35	-13	56	10	-10	-32
035	47	85	-36	32	30	-13	-22	-19	09	-27	-32	-30	-21	-24	-20	-21	31	-28	-12	18	18	-23	-32
044	82	97	-38	-07	34	-22	-26	-07	28	-25	-32	-34	-16	-22	-17	-28	02	-31	-12	46	07	-15	-28
053	76	97	-38	07	43	-23	-28	-15	18	-25	-32	-36	-18	-22	-18	-20	12	-31	-06	48	20	-20	-31
061	66	96	-36	04	25	-18	-25	-06	28	-25	-31	-36	-17	-22	-18	-20	30	-30	-17	34	01	-15	-29
064	82	98	-38	-02	38	-24	-29	-08	28	-25	-33	-35	-17	-22	-18	-21	08	-33	-11	46	11	-16	-30
068	87	97	-40	-10	36	-21	-27	-06	32	-25	-33	-35	-16	-22	-18	-21	01	-32	-13	51	05	-14	-29
083	72	94	-38	11	38	-21	-28	-08	28	-28	-34	-39	-18	-24	-20	-22	16	-33	-15	41	08	-16	-32
105	78	96	-38	08	44	-22	-26	-15	19	-25	-32	-36	-19	-23	-18	-20	11	-30	-06	42	20	-20	-31
107	88	98	-40	04	47	-23	-25	-14	21	-25	-32	-36	-19	-23	-18	-20	11	-32	-06	46	18	-18	-31
001*	-20	-20	51	-07	-28	-08	20	-06	-12	-06	17	15	00	14	-03	01	-13	16	-06	-20	-15	-06	12
005*	-17	01	-08	98	60	-10	-10	-22	-25	-20	-15	-33	-18	-19	-16	-09	66	-11	15	-22	60	-23	-26
007	-18	01	-08	98	97	-08	-08	-23	-25	-20	-15	-33	-18	-19	-16	-10	65	-09	13	-23	58	-23	-25
022	-03	11	-22	97	79	-20	-22	-24	-25	-21	-21	-34	-20	-20	-17	-09	68	-18	28	-06	78	-24	-31
024	-16	02	-08	99	60	-14	-12	-22	-25	-19	-14	-30	-18	-18	-15	-08	65	-12	17	-21	62	-23	-26
030	-04	10	-21	97	79	-19	-21	-24	-25	-21	-20	-33	-19	-20	-17	-09	67	-18	27	-06	78	-23	-30
038	-16	02	-09	98	61	-14	-13	-22	-25	-19	-14	-30	-18	-18	-15	-08	65	-12	18	-20	63	-22	-26
041	-16	03	-09	99	61	-14	-13	-22	-25	-19	-14	-30	-18	-18	-16	-08	66	-12	18	-20	63	-23	-26
056	-04	00	-23	96	81	-20	-22	-25	-26	-21	-21	-34	-20	-21	-17	-02	67	-14	36	02	83	-23	-29
062	-17	01	-03	99	60	-15	-13	-22	-24	-18	-12	-29	-18	-16	-15	-07	63	-12	17	-21	61	-22	-26
088	-04	10	-21	97	79	-19	-21	-24	-25	-21	-20	-33	-19	-20	-17	-09	67	-18	27	-06	78	-23	-30
109	-16	04	-07	98	58	-13	-10	-22	-24	-20	-15	-31	-18	-19	-16	-10	66	-10	14	-22	59	-23	-26
055	34	35	-36	66	94	-26	-30	-23	-11	-26	-27	-39	-23	-25	-21	-08	90	-23	38	31	81	-20	-36
100	43	41	-40	63	96	-27	-32	-17	-00	-28	-32	-42	-23	-26	-21	-16	53	-28	30	37	78	-23	-37
010	-25	-19	-08	-16	-27	97	66	-19	-22	-16	-17	-24	-14	-15	-12	-10	-18	48	-11	-17	-22	-23	18
029	-25	-19	-08	-16	-27	97	66	-19	-22	-16	-17	-24	-14	-15	-12	-10	-18	48	-11	-17	-22	-23	18
034*	-25	-20	-06	-17	-28	98	69	-19	-22	-17	-17	-24	-14	-15	-12	-10	-19	50	-11	-18	-22	-24	20
037*	-29	-24	11	-14	-30	94	91	-20	-23	-17	-11	-25	-14	-17	-13	-11	-20	69	-11	-21	-23	-25	29
046	-28	-23	10	-09	-22	90	80	-20	-23	-16	-11	-25	-15	-17	-13	-12	-17	60	-08	-24	-16	-25	23
066	-25	-19	-08	-17	-27	97	64	-19	-22	-16	-17	-23	-13	-15	-12	-09	-18	48	-11	-17	-22	-23	18
069	-28	-23	12	-10	-23	91	85	-21	-24	-17	-11	-25	-15	-17	-14	-12	-17	62	-09	-24	-16	-26	25
070	-25	-20	-07	-17	-28	98	68	-19	-22	-17	-17	-24	-14	-15	-12	-10	-19	50	-11	-18	-22	-24	20
080	-25	-19	-08	-17	-27	98	66	-19	-22	-16	-17	-24	-14	-15	-12	-10	-19	48	-11	-18	-22	-23	18
081	-26	-20	-04	-17	-28	99	74	-20	-23	-17	-16	-24	-14	-16	-13	-10	-19	53	-12	-19	-23	-24	22
086*	-28	-24	12	-13	-27	91	82	-19	-22	-16	-18	-24	-13	-16	-13	-12	-19	62	-12	-25	-21	-24	25
017	-30	-27	21	-15	-31	64	84	-17	-22	-12	08	-10	-02	-15	-13	-08	-21	75	-04	-14	-22	-02	33
023	-32	-28	27	-16	-34	79	96	-20	-24	-09	-04	-13	-12	-12	-08	-10	-22	77	-09	-21	-26	-21	46
076*	-27	-24	46	-13	-28	72	97	-18	-21	-14	-06	-20	-11	-15	-12	-12	-18	75	-11	-23	-21	-23	40
097	-29	-27	54	-14	-28	68	95	-18	-22	-15	-06	-21	02	-16	-13	-13	-20	74	-12	-26	-23	-23	37
108	-29	-25	37	-14	-29	78	98	-19	-22	-15	-07	-22	-12	-16	-13	-10	-20	78	-09	-20	-22	-24	38
002	-02	-11	-08	-17	-19	-14	-12	97	88	-15	-12	-12	-00	-12	-08	07	-18	-11	-13	08	-22	09	02
009	08	-06	-18	-22	-14	-19	-20	95	90	-17	-17	-16	02	-13	-07	10	-20	-13	-06	33	-20	10	-03
014	-03	-13	-18	-22	-21	-19	-19	98	82	-13	-15	-12	09	-06	-06	10	-21	-14	-09	11	-23	14	-03
028	-04	-15	-15	-25	-25	-21	-20	96	81	-05	-07	02	06	-02	01	06	-23	-15	-09	09	-26	19	07
036*	-03	-12	-15	-20	-19	-17	-17	99	86	-14	-14	-14	05	-11	-06	10	-19	-12	-09	12	-21	11	-02
043	01	-12	08	-27	-25	-23	-21	64	54	-08	01	06	20	02	-08	05	-24	-21	-11	10	-26	30	-02
052	03	-09	-15	-21	-20	-18	-18	98	88	-16	-15	-16	09	-12	-09	07	-19	-18	-14	09	-23	12	-04
071*	06	-08	-11	-21	-17	-20	-20	98	89	-15	-13	-15	04	-09	-08	08	-19	-19	-14	11	-23	11	-05
088	51	36	-30	-19	04	-26	-28	76	94	-25	-27	-29	-07	-20	-16	-06	-16	-29	-18	44	-18	03	-17
103	38	22	-16	-24	-06	-25	-26	76	84	-19	-13	-23	02	-12	-13	-07	-19	-28	-18	25	-22	06	-15
110	18	06	-13	-23	-13	-10	-07	78	82	-15	-15	-08	-10	-11	-06	-02	-23	-07	-11	48	-25	04	08
003	-21	-23	-12	-19	-26	-16	-13	-12	-20	99	78	66	-09	24	11	12	-18	-13	-07	-26	-20	05	25
032	-23	-26	-07	-19	-24	-14	-07	-10	-19	71	95	52	-10	04	-02	45	-19	-02	01	-14	-16	16	23
050	-25	-25	-09	-17	-24	-18	-17	-15	-24	96	66	66	-11	40	15	06	-19	-18	-06	-30	-17	12	17
054	-26	-27	-09	-15	-25	-19	-17	-15	-24	96	65	65	-11	48	15	07	-15	-19	-09	-32	-17	12	16
077*	-23	-25	-11	-26	-28	-17	-15	-13	-21	99	72	70	-10	35	14	09	-20	-16	-09	-28	-22	11	23
079*	-22	-25	-10	-20	-28	-16	-12	-13	-21	99	70	66	-09	29	12	10	-19	-11	-08	-27	-21	06	30
099	-23	-26	-04	-21	-29	-18	-16	-13	-20	95	60	64	-10	39	30	04	-21	-17	-10	-30	-24	04	23
111*	-24	-26	-10	-21	-29	-13	-08	-14	-22	99	73	64	-11	29	11	10	-20	-09	-09	-28	-22	09	30
040	-25	[illegible]	27	-13	-25	-12	-04	-10	-17	63	79	69	-10	12	03	18	-18	03	03	-19	-13	11	20
065*	-23	-26	20	-22	-30	-17	-14	-07	-16	47	60	69	-09	37	26	08	-21	-13	-05	-24	-23	21	31

* The starred MOS are those selected before consulting the clustering results.
Decimal points omitted throughout to conserve space.

Listing of MOSs, CNFs, and Titles by Cluster:

Cluster	MOS	MOS NR.	Name
	36L	004	Electronic Switching Systems Repairer
	32F	011	Fixed Ciphony Repairer
	35F	017	Nuclear Weapons Electronics Specialist
	31J	016	Teletypewriter Repairer
	26K	027	Aerial Electronic Warning/Defense Equipment Repairer
	35B	042	Electronic Instrument Repairer
A	31E	045	Field Radio Repairer
	33S	051	EW/Intercept Systems Repairer
	25J	072	Operations Control Repairer
	26T	075	Radio/Television Systems Specialist
	35M	082	AVIONIC Navigation & Flight Control Equipment Repairer
	34B	084	Punch Card Machine Repairer
	26Y	090	Satellite Communications Ground Station Equipment Repairer
	35R	091	AVIONIC Special Equipment Repairer
	24C	035	Improved HAWK Firing Control Mechanic
	21L	044	PERSHING Electronics Repairer
	27N	053	SHILLELAGH Repairer
B	24U	061	HERCULES Electronics Mechanic
	27G	064	CHAPARRAL/REDEYE Repairer
	24J	068	Improved HAWK Pulse Radar Repairer
	24Q	083	NIKE HERCULES Fire Control Mechanic
	27E	105	TOW/DRAGON Repairer
	23R	107	NIKE Track Radar Repairer
C	95B	001	Military Police
	63B	005	Light Wheel Vehicle/Power Generation Mechanic
	45D	007	Self-Propelled Field Artillery System Mechanic
	45K	022	Tank Turret Repairer
	63N	024	M60A1/A3 Tank System Mechanic
D	63H	030	Track Vehicle Repairer
	63S	038	Heavy Wheel Vehicle Mechanic
	45N	041	M60A1/A3 Tank Turret Mechanic
	62B	056	Construction Equipment Repairer
	63Y	062	Track Vehicle Mechanic
	63W	088	Wheel Vehicle Repairer
	63T	109	ITV, IFV, CFV System Mechanic
E	52C	055	Utilities Equipment Repairer
	52D	100	Power Generation Equipment Repairer
	15D	010	LANCE Missile Crew Member
	15E	029	PERSHING Missile Crew Member
	16S	034	MANPADS Crewman
	13B	037	Cannon Crewman
F	19E	046	M48-M60A1/A3 Armor Crewman
	16T	066	PATRIOT Missile Crewmember
	11H	069	Heavy Antiarmor Weapons Crewman
	16P	070	ADA Short Range Missile Crewman
	16E	080	HAWK Fire Control Crewman
	16R	081	ADA Short Range Gunnery Crewman
	19K	086	M1 Armor Crewman
	82C	017	Field Artillery Surveyor
	13F	023	Fire Support Specialist
G	11B	076	Infantryman
	19D	097	Cavalry Scout
	11C	108	Indirect Fire Infantryman
	05B	002	Radio Operator
	36C	009	Wire Systems Installer/Operator
	72E	014	Telecommunications Center Operator
H	72G	028	Data Communications Switching Center Specialist
	05C	036	Radio Teletype Operator
	93J	043	ATC Radar Controller
	26Q	052	Tactical Satellite/Microwave Systems Operator
	31M	071	Multichannel Communications Equipment Operator
	31V	088	Tactical Communications Systems Operator/Mechanic
I	32D	103	Station Technical Controller
	36K	110	Tactical Wire Operations Specialist
	76V	003	Material Storage & Handling Specialist
	57H	032	Terminal Operations Coordinator
	76C	050	Equipment Records & Parts Specialist
J	76P	054	Material Control & Accounting Specialist
	76Y	077	Unit Supply Specialist
	76W	079	Petroleum Supply Specialist
	76J	099	Medical Supply Specialist
	55B	111	Ammunition Specialist
K	43E	040	Parachute Rigger
L	94B	065	Food Service Specialist

Table 1a. Primary (Oblique) Factor Pattern Matrix:
(Page 2)

MOS NR.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
026	-14	-23	18	-29	-34	-18	-01	25	12	-03	-06	03	81	-06	01	-05	-27	-02	-08	-08	-38	12	17
057	-11	-16	12	-16	-21	-13	-08	03	-09	-12	-14	-16	99	-09	-08	-10	-14	-15	-14	-20	-18	01	-12
059	-11	-16	11	-16	-21	-13	-08	04	-09	-12	-15	-16	99	-10	-08	-10	-14	-15	-14	-20	-18	01	-12
067	-17	-21	28	-17	-24	-14	-08	-03	-15	-08	-09	-10	95	02	-02	-06	-16	-14	-12	-23	-20	02	-08
078	-11	-16	11	-16	-21	-13	-08	04	-09	-12	-15	-16	1.0	-09	-08	-10	-14	-15	-14	-20	-18	01	-12
089	-11	-16	11	-16	-21	-13	-08	04	-09	-12	-15	-16	1.0	-09	-08	-10	-14	-15	-14	-20	-18	01	-12
093	-12	-17	10	-17	-22	-14	-09	13	-01	-13	-16	-17	99	-10	-08	-10	-16	-16	-15	-20	-20	01	-12
047	-21	-23	34	-19	-27	-18	-15	-07	-16	27	13	53	-07	98	18	-06	-18	-18	-10	-25	-22	29	-00
049	-21	-24	33	-19	-27	-17	-16	-08	-17	32	20	50	-07	98	13	-06	-18	-19	-10	-24	-22	33	-05
058	-22	-24	20	-20	-27	-16	-16	-07	-17	36	12	55	-08	99	20	-04	-19	-15	-06	-22	-21	23	02
063	-21	-23	25	-19	-27	-16	-16	-09	-17	35	14	51	-08	1.0	17	-06	-18	-18	-10	-25	-22	26	00
092	-21	-23	25	-19	-27	-16	-16	-09	-17	35	14	51	-08	1.0	17	-06	-18	-18	-10	-25	-22	26	00
102*	-22	-24	25	-20	-27	-16	-15	-09	-17	36	15	61	-08	99	19	-07	-19	-18	-10	-26	-22	23	04
006	-17	-19	-11	-17	-22	-13	-12	-06	-12	18	10	38	-06	09	98	-03	-16	-07	-01	-14	-17	-03	11
033	-17	-19	00	-16	-22	-13	-13	-09	-14	13	03	33	-06	23	99	-05	-15	-12	-06	-19	-18	-03	04
060	-18	-20	-08	-17	-23	-14	-13	-06	-13	14	01	37	-06	21	99	-04	-16	-09	-02	-15	-18	04	05
101*	-16	-18	-07	-15	-21	-12	-11	-08	-13	12	01	40	-06	17	99	-06	-14	-11	-06	-18	-17	-04	08
074*	-21	-20	-07	-09	-11	-14	-15	09	-06	19	32	26	-10	-04	-02	96	-14	09	25	12	03	02	08
085	-22	-22	01	-06	-13	-15	-15	06	-08	25	36	19	-10	-02	-05	98	-13	07	15	08	01	00	10
094	-21	-21	09	-04	-09	-07	-08	07	-06	04	10	-02	-08	-07	-07	97	-07	19	29	23	07	-07	08
106*	-28	-20	02	-13	-15	-06	-06	11	-03	04	09	01	-07	-09	-04	98	-15	24	32	29	03	-04	11
018	-01	09	-24	66	53	-19	-21	-21	-23	-19	-22	-30	-17	-18	-15	-12	1.0	-22	02	-18	50	-21	-28
020*	-01	09	-24	66	53	-19	-21	-21	-23	-19	-22	-30	-17	-18	-15	-12	1.0	-22	02	-18	50	-21	-28
021	-01	09	-24	66	53	-19	-21	-21	-23	-19	-22	-30	-17	-18	-15	-12	1.0	-22	02	-18	50	-21	-28
031	11	24	-28	67	59	-21	-24	-23	-19	-23	-26	-34	-19	-21	-17	-15	97	-27	-01	-13	50	-23	-31
398	-02	09	-24	60	56	-19	-21	-22	-22	-19	-22	-30	-17	-18	-15	-12	1.0	-22	03	-17	54	-21	-28
104*	-30	-08	28	-12	-23	40	60	-14	-21	-12	03	-16	-12	-15	-08	13	-20	86	44	23	03	-19	40
048	-06	-02	-19	30	57	-16	-16	-13	-17	-14	-02	-20	-15	-14	-09	14	15	29	92	56	91	-07	-16
096	-19	-19	-01	-01	08	-04	-01	-08	-16	01	18	07	-09	-03	03	34	-12	64	87	67	49	-07	19
087	49	36	-32	-12	29	-17	-18	11	37	-23	-19	-28	-16	-20	-14	15	-14	22	56	82	27	-01	-09
095	04	12	-27	64	78	-20	-23	-23	-23	-20	-15	-31	-20	-21	-16	03	40	04	73	30	93	-21	-27
013	-02	-16	-15	-28	-27	-23	-18	33	22	06	-02	22	10	18	07	-01	-28	-15	-06	03	-27	73	02
015	06	-03	-21	-08	-02	-24	-23	17	14	05	12	15	-12	08	-09	04	-07	-24	-03	14	-06	76	-09
025	-15	-19	-07	-16	-18	-12	-05	-04	-12	13	20	25	06	15	-02	-01	-15	-10	-06	-10	-16	79	-05
073	-09	-17	11	-18	-21	-16	-14	00	-09	06	11	27	05	40	-00	-09	-17	-18	-09	-10	-19	71	-10
019*	-24	-20	24	-26	-35	11	29	10	01	23	24	44	-06	01	07	10	-26	23	-06	-17	-28	-01	87
030	-27	-29	17	-25	-34	29	45	-12	-17	21	22	37	-08	-01	05	07	-26	61	09	-02	-23	-12	92

* The starred MOS are those selected before consulting the clustering results.
Decimal points omitted throughout to conserve space.

Listing of MOSs, CNFs, and Titles by Cluster:

Cluster	MOS	MOS NR.	Name
	17M	026	Unattended Ground Sensor Specialist
	98C	057	Electronic Warfare/Signal Intelligence Analyst
	05H	059	Electronic Warfare/Signal Intelligence Morse Interceptor
M	96B	067	Intelligence Analyst
	98G	078	Electronic Warfare/Signal Intelligence Voice Interceptor
	98J	089	Electronic Warfare/Signal Intell. Noncommunications Interceptor
	05K	093	Electronic Warfare/Signal Intelligence Non-Morse Interceptor
	71M	047	Chapel Activities Specialist
	71D	049	Legal Clerk
N	75B	058	Personnel Administration Specialist
	75E	063	Personnel Actions Specialist
	75D	092	Personnel Records Specialists
	71L	102	Administrative Specialist
	92B	006	Medical Laboratory Specialist
O	91E	033	Dental Specialist
	91T	060	Animal Care Specialist
	91B	101	Medical Specialist
	64C	074	Motor Transport Operator
P	62F	085	Lifting & Loading Equipment Operator
	62J	094	General Construction Equipment Operator
	62E	106	Heavy Construction Equipment Operator
	67V	018	Observation/Scout Helicopter Repairer
	67N	020	Utility Helicopter Repairer
Q	67Y	021	Attack Helicopter Repairer
	68M	031	Aircraft Weapon Systems Repairer
	67U	398	Medium Helicopter Repairer
R	12B	104	Combat Engineer
S	44B	048	Metal Worker
	51B	096	Carpentry & Masonry Specialist
T	51R	087	Electrician
U	61F	095	Marine Hull Repairman
	84C	013	Motion Picture Specialist
V	83F	015	Photolithographer
	81C	025	Cartographer
	71R	073	Broadcast Journalist
W	54E	019	Chemical Operations Specialist
	12E	030	Atomic Demolition Munitions Specialist

Table 1b. Primary Factor Cosine Matrix:
Twenty Three Oblique Factors.

Factor	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
A	1.0*	78	-40	-12	41	-27	-31	01	41	-25	-32	-33	-13	-21	-17	-22	01	-35	-14	59	05	-07	-28
B	78	1.0	-40	05	40	-22	-28	-11	25	-27	-34	-38	-19	-24	-19	-22	12	-33	-11	43	13	-18	-32
C	-40	-40	1.0	-13	-40	-00	39	-12	-23	-09	34	29	15	27	-07	02	-25	32	-11	-39	-29	-11	23
D	-12	05	-13	1.0	68	-16	-15	-24	-26	-20	-17	-32	-19	-19	-16	-08	67	-14	21	-15	69	-23	-28
E	41	40	-40	68	1.0	-28	-32	-21	-06	-28	-31	-43	-24	-27	-22	-12	55	-27	36	36	84	-23	-38
F	-27	-22	-00	-16	-28	1.0	77	-20	-24	-17	-15	-25	-14	-17	-13	-11	-19	57	-11	-21	-22	-25	23
G	-31	-28	39	-15	-32	77	1.0	-20	-24	-14	-05	-20	-07	-16	-12	-11	-22	81	-10	-22	-24	-20	41
H	01	-11	-12	-24	-21	-20	-20	1.0	88	-14	-12	-10	07	-08	-07	08	-22	-16	-12	14	-25	15	-01
I	41	25	-23	-26	-06	-24	-24	88	1.0	-23	-21	-23	-06	-17	-13	-06	-22	-25	-18	45	-25	05	-09
J	-25	-27	-09	-20	-28	-17	-14	-14	-23	1.0	79	68	-11	34	14	14	-20	-14	-07	-29	-21	10	25
K	-32	-34	34	-17	-31	-15	-05	-12	-21	79	1.0	87	-13	15	04	23	-23	04	04	-24	-17	14	25
L	-33	-38	29	-32	-43	-25	-20	-10	-23	68	87	1.0	-13	54	37	11	-31	-19	-07	-34	-34	30	45
M	-13	-19	15	-19	-24	-14	-07	07	-06	-11	-13	-13	1.0	-08	-06	-09	-17	-14	-14	-20	-21	03	-07
N	-21	-24	27	-19	-27	-17	-16	-08	-17	34	15	54	-08	1.0	17	-06	-19	-18	-10	-25	-22	27	00
O	-17	-19	-07	-16	-22	-13	-12	-07	-13	14	04	37	-06	17	1.0	-05	-16	-10	-04	-17	-18	-01	07
P	-22	-22	02	-08	-12	-11	-11	08	-06	14	23	11	-09	-06	-05	1.0	-13	15	27	18	04	-02	10
Q	01	12	-25	67	55	-19	-22	-22	-22	-20	-23	-31	-17	-19	-16	-13	1.0	-23	01	-17	51	-21	-29
R	-35	-33	32	-14	-27	57	81	-16	-25	-14	04	-19	-14	-18	-10	15	-23	1.0	52	27	04	-22	47
S	-14	-11	-11	21	36	-11	-10	-12	-18	-07	04	-07	-14	-10	-04	27	01	52	1.0	69	78	-08	02
T	59	43	-39	-15	36	-21	-22	14	45	-29	-24	-34	-20	-25	-17	18	-17	27	69	1.0	33	-01	-11
U	05	13	-29	69	84	-22	-24	-25	-25	-21	-17	-34	-21	-22	-18	04	51	04	78	33	1.0	-23	-29
V	-07	-18	-11	-23	-23	-25	-20	15	05	10	14	30	03	27	-01	-02	-21	-22	-08	-01	-23	1.0	-07
W	-28	-32	23	-28	-38	23	41	-01	-09	25	25	45	-07	00	07	10	-29	47	02	-11	-29	-07	1.0

*Except for diagonal elements, decimal points are omitted throughout to conserve space.

III. GENERATION OF THE LONGITUDINAL RESEARCH DATA BASE (LRDB) AND FY81/82 DATA FILE

Lauress L. Wise (AIR)
Winnie Y. Young (AIR)
Paul G. Rossmeissl (ARI)

Project A will generate a large amount of interrelated data that must be assembled into an integrated data base that can be accessed easily by the research teams for various analytical purposes. Therefore, one of the major tasks in Project A is to establish and maintain the longitudinal research data base (LRDB). This data base will link together data on diverse measures gathered in the various tasks of Project A and, in addition, incorporate existing data that are routinely collected by the Army. Such a comprehensive LRDB will enable Project A to conduct a full analysis of how information gathered at each stage of the enlistee's progress through his/her Army career can add to the accuracy of predicting later performances.

The LRDB will not only facilitate efficient validation analyses that concern Project A, but it will also enable Project B to test and revise the prototype selection/allocation system. In addition, the usefulness of the LRDB extends beyond Projects A and B since it can also be used to support other research by the ARI staff (e.g., to address specific policy issues that might arise).

Overview of LRDB Contents

In accordance with the Project A Research Plan, three major sets of data will be assembled within the LRDB. The first set will consist of already existing data on FY81/82 accessions. These data will include accession information (demographic/biographical data, test scores, and enlistment options), training success measures, measures of progress or attrition taken from the Enlisted Masterfile (EMF), and specific information on Skills Qualification Test (SQT) scores. This first set of data will be employed to validate the current version of the Armed Services Vocational Aptitude Battery (ASVAB) insofar as that can be done with available criteria. (This cohort was the first to receive forms 8, 9, and 10 of the ASVAB.) It will also be used to investigate the major methodological and conceptual issues that must be resolved before the optimal estimates of the classification algorithm parameters can be made using the validation data from the 83/84 and 86/87 cohorts. (See Task 1 of the Project A Research Plan.)

The second and third major sets of data to be assembled into the LRDB will involve the new data collection efforts described in the research plan for the 83/84 and 86/87 cohorts.

General Objectives

The primary role of the LRDB is to support efficient data analyses as required by the research teams of both Projects A and B. The data collected throughout the research process of Project A and the data to be acquired from existing Army files must be organized and stored in such a way that they are simple and economical to access. Accordingly, the generation of the LRDB must meet the following objectives:

(1) It must develop systematic and efficient procedures for entering and editing the data.

(2) It must establish linkages of data from various sources and resolve all data inconsistencies.

(3) It must develop and maintain complete documentation of the data organization and contents.

(4) It must store both the data and the documentation cost-effectively and must provide fast and easy access to both simultaneously.

(5) It must insure the security and integrity of the data.

Summary of First Year Activities

A significant portion of the first year's LRDB activities involved planning the data base contents and procedures for the duration of the project. The main result of this activity was the draft and final LRDB plan. Other planning accomplishments included the installation and testing of the RAPID data storage and retrieval system, the development of workfile generation and data set documentation programs, the identification and implementation of data file integrity and security procedures, and the evolution of data editing procedures.

Creation of the FY81/82 Cohort Data Base

Most of the substantive LRDB results during the first year were related to the creation of the FY81/82 cohort data base for use in the preliminary validation of the current ASVAB and the evaluation of new aptitude area (AA) composites. The FY81/82 cohort consists of 885,238 different individuals who applied for regular Army enlistment one or more times during FY81 and FY82. The cohort includes a total of 268,297 regular Army accessions for whom subsequent progress and performance data have been assembled. Table 3 summarizes the types of data records that were assembled for the cohort and the number of applicants for which each type of record was found. A brief description of the different record types follows, along with a summary of the steps taken to enhance the accuracy and usability of the data received from each source. A more detailed description of the data elements is given in Appendix A and a more detailed description of the editing steps for one of the data sources (the training data) is included as Appendix B.

Applicant/Accession Files. ARI receives a monthly data file from MEPCOM, which contains information on all DoD applicants and accessions. Another contractor has already selected out the regular Army applicants and created a merged file for each fiscal year. In the course of editing these data we

(1) "found" over 20,000 applicant records that had been inadvertently lost durng the merging,

(2) identified over 5,000 cases with erroneous SSN codes (cases with identical names and birth dates and only one to three SSN digits mispunched or transposed),

(3) corrected over 2,000 cases where the ASVAB raw subtest scores had been misentered.

(4) unraveled the various uses of the different date fields, corrected errors, and created a set of date variables that are used consistently for all types of cases. (The date of the enlistment contract had been initially stored in the entry date field and then moved into the DEP date field at time of actual accession; we created a variable which was always the enlistment contract date and reserved the entry date field for only the actual accession date, for example, and made similar changes in instances where prior service cases had been treated differently from nonprior service cases.),

Table 3

Number of Cases in the FY81/82 Cohort Data Base by Data Source

DATA SOURCE	FY81	FY82	TOTAL
Applicant/Accession Cases			
Applicants not enlisted	273,175	264,839	538,103
Enlisted but not yet shipped	2,796	42,970	45,766
Discharged without being shipped	8,708	9,610	18,318
Accessions	136,928	128,794	265,722
Total Applicant Cases	421,607	446,212	867,819
MAP Cases			
Cases with MAP and accession data	4,618	3,794	8,412
Cases without applicant data			1,914
Total MAP Cases			10,326
Training Cases			
Cases with applicant data	49,728	6,077	55,805
Cases without applicant data			13,134
Total Training Records			68,939
EMF Cases			
Cases with applicant data	105,519	88,193	193,712
Cases without applicant data			2,575
Total EMF Cases			196,287
SQT Cases			
Cases with applicant data	47,746	12,167	59,913
Cases without applicant data			3,793
Total SQT Cases			63,706

(5) corrected inconsistencies in the recording of entry status,

(6) corrected, wherever possible, ubiquitous errors in the entry of the MOS fields, and in other key fields such as the ASVAB form code,

(7) resolved inconsistencies in sex, race, and enlistment program information between values on the accession files and values on the Enlisted Masterfile, and

(8) began work on the documentation of these data that includes both codebooks giving the frequency of each value of each variable and a more detailed explication of the meaning and use of each variable.

Some work remains to be done on the applicant/accession data including the editing of some variables not critical to the initial validation analyses (e.g., medical block data) and the resolution of additional interfile inconsistencies.

Military Applicant Profile (MAP) Data. Applicants who are not high school graduates are required to complete a special biographical qustionnaire from which an overall "suitability" score is derived. The questionnaire item responses are coded on scan sheets, but have been scored by hand. ARI has accumulated the scan sheets for most or all of the FY81/82 applicants who completed the MAP. During the first year, we oversaw the scanning of these sheets and the loading of the resultant data at NIH. Data on 10,326 different applicants resulted from this activity. These data are now being checked, new scores are being generated, and attempts are being made to resolve SSN errors that result in mismatches to the overall applicant files. When this editing is completed, the data file documentation will be prepared and the data will be merged with the main applicant files.

Training Records Data. Training records were processed for 68,939 recruits who went through training during calendar year 1981. (Actually many more records were processed as some recruits went through more than one course.) As a first step, we reprocessed about 20,000 records for which significant information had not been entered initially. As with the applicant/accession data, the editing consisted of resolving erroneous or inconsistent values in every data field. A field-by-field description of the steps in editing the training data is attached as Appendix A to this report. The editing also included significant efforts to differentiate duplicate training records from cases where the same soldier actually participated in more than one course. We also identified and corrected over 1,100 SSN errors that had led to a failure to match the training record to a corresponding applicant/accession record.

Enlisted Masterfile Data. Information from the FY82 year-end EMF was captured and entered into the LRDB. This information has been used to check key fields (e.g., race, sex, and SSN) in the other data files and to check against hardcopy records in Task 4's efforts to identify sources of information on general Army performance. These data will be used to assess soldier progress in the Army. Since these data are not involved in the initial validation analyses, the cleaning and documentation of this file has been given a lower priority and is only just now getting underway. Beginning with the last quarter in FY83, quarterly progress information will be extracted from the EMF for both the FY81/82 cohort and also for the FY83 study cohort.

Skills Qualification Test Data. Data were received on SQT testing from FY79 through about January 1983. These files were found to contain SQT records for 59,913 of the individuals in the applicant file (some of whom were prior service cases and had taken an SQT prior to reentry into the Army in FY81/82). In general, only minor editing was required on the test date, MOS, and skill level fields. Some delays have been experienced in obtaining more recent SQT information (through at least June 1983), but a new tape is expected momentarily.

Data Base Security

Whenever a large amount of data on individuals is maintained and stored, it is necessary to protect that data from compromise. The security of the Project A and B data base is particularly important for a number of reasons. Some of the data collected on individual soldiers, such as promotions, paygrade, or disciplinary actions, will be private in nature, and the privacy of that information must be protected. Since many researchers will be accessing the data base for a variety of uses, the integrity of the data must be maintained in a manner that insures the data remain accurate and consistent across uses. Finally, it is necessary to secure the data base so that the Army maintains complete ownership of the data, to insure that the data within the data base are used only for authorized project A and B research.

The security of the data base is protected in three ways. Soldier social security numbers (SSN) will be routinely encrypted to insure the privacy of each soldier's records. Access to the data base will be carefully controlled both to further protect soldier privacy and to insure proper use of the data. Finally, a log will be maintained for the system that will note each attempted access of the data base and whether or not the access was authorized. Each of these procedures is outlined in turn below.

The key procedure in guaranteeing the privacy of individual soldier data is the coding or encrypting of each soldier's identifier. This encryption is accomplished by scrambling each soldier's SSN in an unpredictable manner. The specific algorithm that does the encrypting (and if needed, decrypting) is known only to the data base administrators, with a printed copy of the algorithm being securely maintained. All of the data files of the data base that are routinely accessed and any project workfiles that are generated from the larger data base files will use only the encrypted SSN as an identifier.

Data integrity and accuracy are maintained by controlling the access to the large files or relations within the data base. This procedure also helps contribute to the privacy protection of soldier records. The system uses the RACF procedure at NIH to restrict the access of selected files to authorized users. Under RACF different degrees of access can be structured. By specifying a "universal access" of "NONE", access can be restricted to only those users granted specific exceptions. Users would then have to provide an eight character RACF password (different for each user) in order to read the datafiles for which they are authorized. Using the provisions of RACF, a series of access "levels" has been installed to provide timely access to relevant data needed by project researchers and yet protect the data from compromise.

At the highest level of access are the data base administrators who will have access to all of the files and relations within the data base, and are the only personnel who will be able to enter data into the data base or to modify data already stored in the data base. Thus, the data base administrators have responsibility for all data entry and editing. It is also the duty of the data base administrators to create work files based upon data base files and relations as they are needed by other project researchers.

Project personnel with the next highest data base access authority will be able to read data from all of the files within the data base, with the exception of the Link File which contains all of the basic identifying information for each soldier. This exception is being made to help maintain soldier privacy. Only a few of the project staff will be at this access level and their primary responsibility with regard to the data base will be to back up the data base administrators.

Most project researchers will have some level three data base access. Researchers at this level will have direct access to those files that are generated by the particular issues they are investigating. They will also have direct access to files created by other research that should have a direct influence upon their research. For example, researchers investigating the development of new preinduction predictors of performance will have direct access to the task analysis data that is being collected by those investigating criterion development so that the new predictors that are developed will address areas of the criterion space not currently covered by ASVAB.

than tape by providing information as to how frequently data are requested from any given file.

Any set of procedures designed to store data electronically needs to balance the ease with which data can be accessed against the security of the data base. The procedures described here tend to favor the security aspect of this balance. The number of data files that most project scientists will be able to access directly will be small in comparison to the total amount of stored data. Furthermore, only the data base administrators will have access to the true soldier identifying information and be able to add or modify data. However, these limitations should not prove to be too restrictive since prompt creation and efficient use of work files should provide each scientist with the data that he or she needs to perform the required research.

Other LRDB Activities

While the main work of the past year has been to develop the overall plans and to establish the FY81/82 cohort database, there were a number of additional accomplishments. One effort was to assist other members of the research staff in creating a computerized data bank of job task descriptions for each of the focal MOS.

Another activity has involved working with Task 4 staff to create samples of administrative records from which manual records could be checked and to enter the information obtained from these checks for analysis.

A final activity has been the creation and documentation of work files for users at ARI and for outside users for whom ARI has been requested to provide data.

Next Steps

Creation of a data base of this magnitude is a massive undertaking. Much progress has been made but it is not yet complete. During the next contract period the accumulation, editing, and documentation of the FY81/82 data files will be completed. The result will be the largest single data base ever created for purposes of personnel selection and classification research.

Also during this period: (a) additional special files will be created for specific research purposes, (b) data from the administration of the preliminary battery to the 83/84 longitudinal sample will be entered, (c) data from the pilot testing of training school measures and MOS-specific measures will be entered in the data base, and (d) assistance will be provided to each of the other tasks and to ARI staff as they begin their initial analyses of Project A data.

Associated Report

As noted in the Introduction, a relevant research report follows.

DEVELOPMENT AND VALIDATION OF *
ARMY SELECTION AND
CLASSIFICATION MEASURES

PROJECT A:
LONGITUDINAL RESEARCH DATABASE PLAN

JUNE 1983

Lauress L. Wise

Ming-mei Wang

American Institutes for Research

Paul Rossmeissl

U.S. Army Research Institute
for the Behavioral and Social Sciences

* To be published as ARI Research Report 1356

1. INTRODUCTION

1.1. Nature and Purpose of the Overall Project

The Army Research Institute (ARI) is currently funding two large-scale research projects in order to develop a new selection and classification system that will improve the efficiency of personnel utilization in the Army. The purposes of the first project, Project A: Development of Improved Army Selection and Classification Systems, are:

(1) to validate current predictors (primarily the ASVAB composites, supplemented by other available predictors such as high school graduation, Military Applicant Profile (MAP) data, and physical capacities);

(2) to develop new or improved predictors and performance measures; and

(3) to conduct a longitudinal validation of current and newly developed classification measures for prediction of the enlistee's performance from training through the second tour of duty.

The second project, Project B: Development of a Enlisted Personnel Allocation System (EPAS), is to develop a state-of-the-art system, for implementation at the Military Entrance Processing Station (MEPS), to facilitate the initial enlistment decisions. The success of the EPAS depends on a set of cost-effective classification measures that can accurately predict the recruit's future performance throughout his/her Army career. The major objective of Project A is to ensure that such a set of predictors is available to drive the new allocation system. Thus, although the two projects are being conducted separately, they have a common goal to improve the Army's personnel decision mechanism and thereby increase the overall effectiveness of the Army.

The selection of a set of cost-effective classification measures for use in the EPAS requires first a careful evaluation of the relationships of the current predictors to performance on Army jobs. The Army presently has a systematic testing and validation program to support its current selection practice. Specifically, the Armed Forces Vocational Aptitude Battery (ASVAB) is administered to all applicants, and aptitude area composites are developed for use in the initial selection and assignment to training courses. The ASVAB composites are traditionally validated in terms of the extent to which they predict the enlistee's success in training (essentially measured by course grades). Although these composites generally are quite effective in predicting how well the enlistees will perform during training, their validities for predicting other important areas of Army performances -- general soldiering and job-specific performances -- have not been extensively investigated because valid, sound, and economical measures of these additional aspects of performance are not currently available.

In addition to valid predictions of performance, the new EPAS will require information on the relative utility to the Army of different levels of performance in different MOS. The collection and analysis of such data is another major objective of Project A.

While the greatest concern is with initial selection and classification decisions, Project A will also address subsequent personnel allocation decisions. Two major decision points will be investigated. The first point is the decision at the end of training whether to pass a recruit out of training, or to recycle him or her into additional training for the same or some other MOS, or to drop him or her from the Army altogther. At this point, both pre-induction and training performance measures will be used to predict subsequent performance in the MOS. The second major decision point is at the completion of each soldier's first tour. At this point the Army must decide whether to encourage or to bar the soldier's reenlistment for a second tour. Here first-tour performance measures must be used, along with the training and pre-induction measures, to predict second-tour performance. Thus, valid training and first-tour performance measures are needed both as criteria for the validation of earlier prediction measures and as predictors of subsequent criteria.

In order to accomplish these objectives, the project is organized into five major tasks. These are:

Task 1: Validation

Task 2: Prediction of Job Performance

Task 3: Measurement of School/Training Success

Task 4: Assessment of Army-wide Performance

Task 5: Development of MOS-Specific Performance Measures

In the course of developing these new or improved measures, there will be pilot field tests in order to assess the psychometric characteristics of the measuring instruments and the ease of their administration. Based on these pilot tests, the instruments may be revised and then employed in a large-scale field test to collect data for two purposes. The first purpose will be to evaluate formally the effectiveness of the experimental set of pre-induction predictors for predicting success in the Army. The second purpose will be to examine the practical value of using early performance measures as additional predictors of later performances. The results of these evaluations will be employed to guide further refinements of the experimental measures. Through this iterative development and refinement process, a final set of predictor and performance measures will be selected and administered to a cohort of enlisted personnel. The data collected in these administrations will be analyzed for the validation of the classification battery to be employed in the Army's new selection and classification system.

In conjunction with this complete, longitudinal validation of the classification measures, prediction models of enlistees' performance in the Army will be obtained to generate numerical inputs into the EPAS for determining optimal person-job matches. The development of a dynamic allocation procedure will be most useful to the Army if it uses information accumulated through the early period of the enlistee's career to modify post-enlistment decisions at various choice points (post-training reassignment, promotion, and reenlistment). The capability of the EPAS can be enhanced by incorporating such a dynamic decision process. EPAS would then be used to assist in personnel decisions beyond the initial selection and training assignment made at the MEPS level.

1.2. Role of the Longitudinal Research Database (LRDB)

Clearly, the research process of Project A will generate a large amount of interrelated data that must be assembled into an integrated database that can be accessed easily by the research teams for various analytical purposes. Therefore, one of the major tasks in Project A is to establish and maintain the longitudinal research database (LRDB). This database will link together data on diverse measures gathered in the various tasks of Project A and, in addition, incorporate existing data that are routinely collected by the Army. Such a comprehensive LRDB will enable Project A to conduct a full analysis of how information gathered at each stage of the enlistee's progress through his/her Army career can add to the accuracy of predicting later perfomances.

The richness of the LRDB to be created for the project will not only facilitate efficient validation analyses that concern Project A, but will also enable Project B to test and revise the prototype selection/allocation system. Specifically, Project B will employ data on the training time, subsequent performance, and the utility of subsequent performance levels to develop the classification model and estimate required parameters.

Indeed, the function of the LRDB extends beyond Projects A and B. The database can also support other research work to be conducted by the ARI staff to address specific policy issues that may arise.

1.3. Overview of LRDB Contents

In accordance with the Project A Research Plan, three major sets of data will be assembled within the LRDB. The first set will consist of already existing data on FY81/82 accessions. These data will include accession information (demographic/biographical data, test scores, and enlistment options), training success measures, measures of progress or attrition taken from the Enlisted Masterfile (EMF), and specific information on Skills Qualification Test (SQT) scores. This first set of data will be employed to validate the current version of the Armed Services Vocational Aptitude Battery (ASVAB) insofar as that can be done with available criteria. (This cohort was the first to receive Forms 8, 9, and 10 of the ASVAB.)

Recommendations will be made for revisions in the ASVAB Area Composite scores to be used in classification decisions until EPAS becomes fully functional. These analyses of the existing battery will also produce recommendations for needed additions to the predictor battery. Furthermore, the investigation of methodological and conceptual issues that have plagued personnel decision research will begin with the data on this cohort so that practical solutions may be devised for the validation analyses on two subsequent cohorts. (See Task 1 of the Project A Research Plan.)

The second and third sets of data to be assembled into the LRDB will involve substantial new data collection efforts, in addition to the routinely collected data described above. The second set of data will consist of longitudinal information on FY83/84 inductees. This information will be acquired in three data collection phases:

(1) Beginning in the summer of 1983 and continuing until summer of 1984, samples of recruits will be administered a preliminary predictor battery (consisting of available tests that are not currently employed by the Army but that have potential value for predicting performance on

Army jobs). These data will be analyzed to determine the incremental validity of the new tests (tests such as vocational interest and motivation scales) over the existing predictors (in essence, the ASVAB scores). The results of this evaluation will help guide the development of new preinduction predictors. (They will be developed to be parallel to the preliminary predictors that are found to be effective for predicting performance.)

(2) Later in their first tour (June 1985 to October 1985), data on a revised set of predictors, job knowledge tests, and Army-wide and MOS-specific performance measures under development by Project A will be obtained from this sample. These data, together with the existing data on current predictors and school performance indicators, will be employed to conduct a concurrent validation of the initial predictors using both school measures and subsequent performance as criteria. The school measures will also be validated as predictors of subsequent performance. The findings from this concurrent validation analysis will provide the basis for revision and improvement of new instruments and

for choosing the most cost-effective of them for inclusion in the final set of classification measures.

(3) For members of this sample who reenlist, Army-wide and MOS-specific performance measures will be collected during their second tour (June 1988 to September 1988) and merged with existing EMF measures. These data will be used to validate the pre-induction selection measures and early performance in the Army as predictors of second-tour performance.

Once the new measures are refined on the basis of the analyses of the FY83/84 cohort data, they will be administered to a new cohort (FY86/87 inductees) to allow a complete, longitudinal validation of the final classification battery. The data for this final validation will also be collected in three phases. Briefly, the three data collection points are:

(1) From March 1986 until February 1987, samples of recruits from the 19-focal MOS will be tested with the revised predictor battery, and their school data will be obtained.

(2) During their first term of enlistment (June 1988 to September 1988), the sample will be followed up and the Army-wide and MOS-specific performance measures will be obtained. These data will not only support the predictive validation of the classification measures but will also permit analysis of criterion equivalence between training and first-tour performance and between Army-wide and job-specific performance.

(3) Similarly, during their second tour (January 1991 to March 1991), Army-wide and MOS-specific performance measures will again be obtained from this sample. These data will be used for the longitudinal validation of the predictor measures and the investigation of criterion equivalence between first-tour and second-tour performance measures.

1.4. Specific Objectives of the LRDB Plan

As pointed out in the preceding section, the primary role of the LRDB is to support efficient data analyses as required by the research teams of both Projects A and B. To fulfill this role, the LRDB must be created and maintained in coordination with the data collection activities and the research process. The data collected throughout the research process of Project A and the data to be acquired from existing Army files must be organized and stored in such a way that they are simple and economical to access. Accordingly, the LRDB plan must meet the following objectives:

(1) To develop systematic and efficient procedures for entering and editing the data.

(2) To establish linkages of data from various sources and resolve all data inconsistencies.

(3) To develop and maintain complete documentation of the data organization and contents.

(4) To store both the data and the documentation cost-effectively and to provide fast and easy access to both simultaneously.

(5) To insure the security and integrity of the data.

2. CONTENTS OF THE LRDB

The adequacy of the LRDB depends heavily on the specific variables that are included. It is not sufficient, for example, to specify that "training measures" will be included. The planning and conduct of the validation analyses require knowledge of the specific predictor and criterion measures that will be available and of the data elements that provide qualifying information. Unfortunately, tne degree to which specific variables can be listed varies widely for the three main cohorts. For the FY81/82 cohort, most of the specific variables that will be available are now known. For the FY83/84 and FY86/87 cohorts, a great deal of work will go into defining and collecting new items of information. It is not now possible to give a complete list of specifics at this time.

What follows is a listing of specific data elements to be included in the database. This list should be helpful to all project staff in planning both analyses and future data collections.

Variable names. Before proceeding, however, a word is in order on conventions and standards regarding variable names. There is a wide range of possible naming conventions, ranging from a strict numbering (e.g., VAR129) to acronym conventions (e.g., HTIN=height in inches) to single word descriptors (e.g., HEIGHT). The naming convention to be used in this project will combine (a) a two character prefix indicating data source with (b) a descriptive label of up to six characters. (Note that eight characters is the maximum variable name length in most statistical packages.)

The two characters indicating data source will consist of an initial character, designating the type of data, followed by a sequencing character (1 through 9 and then A through Z). The type of data codes are:

A - existing applicant/accession data (including existing predictors)

B - new predictor-battery data

E - Enlisted Masterfile data (including existing Army-wide performance measures)

G - new general (Army-wide) performance data

P - existing MOS-specific performance data (e.g., SQT)

M - new MOS-specific performance data

T - existing school/training data

S - new school/training data

Additional codes may be defined f›r derived variables that combine data from different sources.

In cases where data are extracted from existing datafiles, the established variable name will be used as the descriptive portion of the variable name in our system (in characters 3 through 8). EMF variable names, for example, are limited to five characters and thus fit nicely into the system. In other cases, it may be necessary to shorten the name to six characters. Where appropriate, more obvious mnemonics may replace the variable name in the original file.

2.1. Data Elements for the 81/82 Cohort

2.1.1. 81/82 APPLICATION AND ACCESSION DATA

The Army collects a great deal of information on each soldier at the time that he/she applies to and is accepted into the Army. Some of this information is retained in each soldier's permanent computerized records (the Enlisted Masterfile), but much is not. Some information, such as responses to individual ASVAB items, is not retained in machine-readable form at all.

For the most part, analyses of the FY81-82 cohort will be limited to data that are already in machine-readable form. One exception will be information used in the

Military Applicant Profile (MAP) score. The MAP is a battery of behavioral indicators now administered to all applicants who have not completed high school. It has been discovered that the overall profile score has not been included in the computerized accession file. The answer sheets, including responses to the approximately 60 (noncognitive) items that comprise the profile, have been retained at ARI and are available for entry. We plan to enter this information for a sample of about 2,500 applicants to allow for analyses of the current MAP items.

The following items of information will be taken from the existing accessions file and retained in the database:

2.1.1.1. Basic Identifying Information

The data will be used to link the accession data to the EMF data. The linking variables will NOT be stored on the main data files: only a scrambled identifier will be retained on the main data files for linking new data. The data needed for linking and checking the validity of the linkage are:

A1SSN	SOC.SEC.NO.
A1NAME4	4 CHAR NAME ABBREVIATION
A1NAME	FULL NAME
A1DOB	DAY OF BIRTH

2.1.1.2. Individual Background Data

These data will be used to identify differences in backgrounds that may be predictive of performance differences. In addition, certain variables may be valuable as moderators, predicting differences in the relationships between predictor and criterion variables. Other variables such as ZIP codes, will permit links to other (e.g., Census) data files containing information on lccal geographic or economic conditions.

A1HOMADD	STATE/COUNTY CODE OF HOME ADDRESS
A1HOMZIP	HOME ZIP CODE
A1PRSADD	STATE/COUNTY CODE OF PRESENT ADDRESS
A1PRSZIP	ZIP CODE OF PRESENT ADDRESS
A1MARST	MARITAL STATUS
A1NRDEP	NUMBER OF DEPENDENTS
A1YOB	YEAR OF BIRTH
A1MOB	MONTH OF BIRTH
A1CITIZ	CITIZENSHIP
A1SEX	SEX
A1RACE	POPULATION GROUP
A1ETHNIC	ETHNIC GROUP
A1RELIG	RELIGIOUS PP ERENCE

A1HGT	HEIGHT
A1WGT	WEIGHT
A1PULHES	PHYSICAL PROFILE
A1EDYRS	YEARS OF EDUCATION
A1EDCERT	EDUCATION CERTIFICATION

2.1.1.3. Enlistment Information

These data describe the timing and conditions of enlistment. This information is of primary importance in the development of forecasting models by Project B. In addition, some of these variables (e.g., entry date, primary MOS, pay grade) provide the starting points against which the relationship between the test scores and progress will be charted. Other variables (e.g., moral waivers, additional skill indicators) will be useful as additional predictors.

A1ENTDTE	ENTRY DATE
A1PADDTE	PROJECTED ACT. DUTY DATE
A1AITGRD	AIT GRADUATION DATE
A1AAAS	ACCESSION TO ACTIVE ARMY STRENGTH
A1UPSTAT	STATUS CODE
A1ACTDTE	DATE OF ACTION
A1SERV	BRANCH OF SERVICE

A1PRMMOS	PRIMARY MOS
A1TRNMOS	TRAINING MOS
A1ENLOP	ENLISTMENT OPTION GUARANTEED
A1ENOPT1	ENLISTED OPTION
A1DESGOP	DESIGNATED OPTION
A1ENTPRG	PROGRAM FOR WHICH ENLISTED
A1ENLTRM	TERM OF ENLISTMENT
A1ENTST	ENTRY STATUS
A1BONLVL	ENLISTMENT BONUS LEVEL
A1ABGRD	ABBREVIATED GRADE CODE
A1PAYGRD	PAY GRADE
A1GRDDTE	DATE OF GRADE
A1ASI	ADDITIONAL SKILL INDICATOR
A1WAIVER	WAIVER TYPE
A1MORWVR	REASON FOR MORAL WAIVER
A1WAPLVL	WAIVER APPROVAL LEVEL
A1AFEES	AFEES IDENTIFICATION

2.1.1.4. Prior Military Experience

A1YTHPRG	YOUTH PROGRAM
A1CONDBY	YOUTH PROG.CONDUCTED BY
A1YRSCMP	NO.OF YEARS COMPLETED IN YOUTH PROG
A1PRISRV	PRIOR SERVICE
A1SRVBRK	BREAK IN PRIOR SERVICE

2.1.1.5. Delayed Entry Program (DEP) Information

A1DEPDTE	DEP DATE
A1ENTDT2	DATE OF CONTRACT/ENTRY
A1DISDTE	ENTRY OR DISCHARGE DATE
A1ENTRST	ENTRY STATUS
A1DEPPG	PROGRAM FOR WHICH ENLISTED
A1DESDEP	DESIGNATED OPTION
A1OPTDEP	ENLISTMENT OPTION
A1TNGMOS	TRAINING MOS
A1NODEPR	DEP NON-ENLISTMENT REASON

2.1.1.6. Hometown Recruiter Aide Program (HRAP) Information

A1HRAP	HOMETOWN RECRUITER AIDE
A1HRAPLC	HRAP LOCATION

2.1.1.7. Testing Information

A1CYCL	DATE OF CYCLE NUMBER
A1MCAT	MENTAL CATEGORY
A1TSITE	TEST SITE
A1TSESS	TEST SESSION
A1ASVBFM	ASVAB FORM
A1AFQTPC	AFQT PERCENTILE
A1ASVBxx	ALL ASVAB SUBTEST SCORES

Current ASVAB Area Composite Scores

AlASCVGT	GENERAL TECHNICAL
AlASCVGM	GENERAL MAINTENANCE
AlASCVEL	ELECTRONICS
AlASCVCL	CLERICAL
AlASCVMM	MECHANICAL MAINTENANCE
AlASCVSC	SURVEILLANCE/COMMUNICATIONS
AlASCVCO	COMBAT
AlASCVFA	FIELD ARTILLERY
AlASCVOF	OPERATIONS/FOOD
AlASCVST	SKILLED TECHNICAL
AlASCVWS	AFWST(WOMEN ONLY)

Previous ASVAB Subtest and Composite Scores

AlPASVFM	PREVIOUS ASVAB TEST FORM
AlPAFQTS	PREVIOUS ASVAB TEST-AFQT
AlPASVxx	PREVIOUS ASVAB SUBTEST SCORES
AlPASCXX	PREVIOUS ASVAB CONPOSITE SCORES

2.1.2. TRAINING DATA

ARI has expended considerable effort to collect training information on 1981 and some 1982 accessions. These data indicate the timing and duration of training, the course(s) taken, the overall outcome, and some measure of performance in the course. It is important to note that the nature of these performance measures varies widely by school and sometimes by course or class within school.

2.1.2.1. Basic Identifying Information

These data will be used for identification purposes only and will NOT be stored on the main data files; only a scrambled identifier will be retained in the main data files for linking in new information.

T1NAME5	5 CHAR ABBREVIATION FOR NAME
T1SSN	SOCIAL SECURITY NUMBER

2.1.2.2. School Identification Information

These data will be used to identify the school, the class, and the course for which the scores on each file have been collected.

T1SCHOOL	SCHOOL/ATC CODE
T1COURSE	NAME OF COURSE
T1CLASS	CLASS ID NUMBER WITHIN COURSE
T1MOSAWD	MOS AWARDED UPON COMPLETION
T1SKLLVL	MOS SKILL LEVEL AWARDED

2.1.2.3. Students' Progress Through the Training Program

Essential to this project are the data which describe each student's progress through training.

T1ENRDTE	ENROLLMENT DATE
T1GRDDTE	DATE OF RECYCLE, TRANSFER, OR GRADUATION
T1ATTRIT	TYPE OF ATTRITION
T1DISP	DISPOSITION (PASS, RECYCLE, TRANSFER OR DROP)
T1SCORE1	STUDENT'S COURSE GRADE OR TEST SCORE
T1STYPE1	TYPE OF SCORE
T1SCORE2	SECONDARY PERFORMANCE MEASURE (FOR SOME MOS)
T1STYPE2	TYPE OF ADDITIONAL SCORE
T1SELECT	WAS SPECIFIC MOS GUARANTEED FOR BASIC INFANTRYMEN
T1MORSE	MORSE CODE TAKEN FOR 05B AND 05C

2.1.3. DATA FROM THE ENLISTED MASTERFILE (EMF)

The Army Enlisted Masterfile (EMF) contains a significant amount of information that is essential to Project A. In particular, information on each soldier's progress through his or her Army career is captured by the EMF. The EMF also contains important information on the individual background and enlistment conditions of each soldier that are important checks against similar information obtained from the Accession files.

While some of the analyses will focus on a specific MOS, others will require information on a broadly representative cohort of soldiers. In particular, in analyzing the generalizability of results from samples of MOS, it is essential that such representative cohorts be analyzed. The EMF provides one source of information on the progress of all recruits, against which the results for specific samples can be compared.

The following list indicates the EMF data elements that will be needed by this project in order to avoid large and redundant data collection costs. The variables are grouped into seven types of information, and the use of each type of information in the planned analyses is indicated. Basic and background information will be retained only once in the system. Other information, on progress and problems, will be obtained at regular intervals and accumulated into the database.

2.1.3.1. Basic Identifying Information

Again, these data will be used to link new EMF information to data previously collected on the samples of soliders in the project. The linking variables will NOT be stored on the main data files. Only a scrambled identifier will be retained on the data files for linking in new information. The EMF variables needed for linking and for checking the validity of the linkage are:

E1SSN	SOCIAL SECURITY NUMBER
E1SSNPR	PREVIOUS (INCORRECT) SSN
E1SNCTL	SSN CORRECTION DATE

2.1.3.2. Individual Background Data

These data will be used to identify differences in backgrounds that may be predictive of performance differences. In addition, certain key variables will be used to check the "cultural fairness" of any proposed selection and classification algorithm. Note that most of this information will also be obtained from the accession files. The corresponding EMF variables will be used to check or verify the accession data. After completing this check, only one copy of this information will be retained. The background data elements from the EMF that are needed to either add or verify essential information include:

E1SEX	SEX
E1RACE	POPULATION GROUP
E1REDCAT	RACIAL/ETHNIC DESCENT
E1ETHNIC	ETHNIC GROUP DESIGNATION
E1CLANG	LANGUAGE IDENTITY
E1CITIZ	CITIZENSHIP STATU
E1DOB	DATE OF BIRTH
E1MARST	MARITAL STATUS
E1NRDEP	NUMBER OF DEPENDENTS
E1CIVED	ACADEMIC EDUCATION LEVEL
E1MADCD	COLLEGE MAJOR
E1STRD	STATE OF RESIDENCE AT ENLISTMENT

2.1.3.3. Enlistment Conditions

The enlistment data required by this project include physical and mental test scores and information on the terms or conditions of enlistment. The test scores are the primary predictive measures currently available. The information on enlistment conditions is essential to understanding the relationship between the test scores and subsequent performance in the Army. As with background data, much of the enlistment information will also be obtained from accession files. Again, only one copy of this information will be retained after any inconsistencies are resolved. The required enlistment variables include:

E1ASVBXX	ALL ASVAB AREA COMPOSITE SCORES
E1AFQSC	ARMED FORCES QUALIFICATION TEST SCORE
E1AFQG	AFQT GROUP
E1DLAB	DEFENSE LANGUAGE BATTERY SCORE
E1PHYPR	PHYSICAL PROFILE
E1PHYCA	PHYSICAL LIMITATION CATEGORY
E1XFACT	WEIGHT-LIFTING CAPACITY
E1COMPT	SERVICE COMPONENT
E1ENLOP	ENLISTMENT OPTION CODE
E1MORWA	ENLISTED/REENLISTMENT WAIVER
E1TERMS	TERMS OF SERVICE OR ENLISTMENT
E1BASD	BASIC ACTIVE SERVICE DATE
E1BONIN	BONUS INDICATOR
E1RPFLG	RECRUITER FLAG (PROMOTED OR SEPARATED)
E1RCRCD	RECRUITER CODE
E1PLOEN	STATE OF ENLISTMENT
E1TYPLA	TYPE OF LAST ACCESSION

E1DATLA	DATE OF LAST ACCESSION
E1ETSDT	DATE OF EXPIRATION OF LAST TERM OF SERVICE

2.1.3.4. Basic Progress in the Army

A major outcome to be predicted at the time of selection is the applicant's probable rate of basic progress in the Army. Several EMF variables are needed to chart the progress of the soldiers in the research samples for use in validating new and existing predictor measures. These include:

E1GRTIT	GRADE IN WHICH SERVING
E1DOR	DATE OF RANK
E1PAYGR	PAYGRADE
E1PAYSX	PAYGRADE & SEX
E1GRDDT	DATE OF LAST GRADE CHANGE
E1BPEDT	BASIC PAY ENTRY DATE
E1GRDTT	TYPE OF LAST GRADE CHANGE
E1NCOES	NCO EDUCATION SYSTEM (LEVEL ATTAINED)
E1PROPT	CURRENT PROMOTION POINTS
E1PROPDT	CURRENT PROMOTION POINT DATE
E1PRVPT	PREVIOUS PROMOTION POINTS
E1PRVPDT	PREVIOUS PROMOTION POINTS DATE
E1PROPA	PROFICIENCY PAY STATUS

E1AITDT	AIT GRADUATION DATE
E1PACE	SELF-PACED AIT FLAG
E1EERWA	EER WEIGHTED AVERAGE
E1TUREL	TOUR ELIGIBILITY
E1SECCLR	PERSONNEL SECURITY CLEARANCE
E1SGTID	DRILL SERGEANT QUALIFICATION
E1ADPAY	ELIGIBILITY FOR ADDITIONAL PAY
E1VEAP	VETERANS EDUCATION ASSISTANCE PROGRAM CODE

2.1.3.5. Performance in a Particular MOS

Since much of military performance is specific to particular occupational specialties, many of the criteria used in evaluating new and existing predictor measures will concern progress and performance within an MOS. The specific EMF variables required to track this information are:

E1CMF	CAREER MANAGEMENT FIELD
E1PRMOS	PRIMARY MOS
E1DMOS	DUTY MOS
E1SMOS3	SECONDARY MOS CURRENT (3-POS)
E1PMOTT	TYPE OF LAST PMOS CHANGE
E1PMODT	DATE OF LAST CHANGE TO PMOS
E1PGMOS	PRIMARY PROGRESSION MOS

E1BOMOS	MOS OF BONUS
E1DDSID	ADDITIONAL SKILL INDICATOR, DUTY MOS
E1ADSID2	ADDITIONAL SKILL INDICATOR, PREVIOUS
E1ADSID3	ADDITIONAL SKILL INDICATOR, 2ND PREVIOUS
E1PQDES	PRIMARY MOS, IN WHICH TESTED, SQT DESIGNATOR
E1PQSCR	PRIMARY SQT SCORE (FOR PQDES)
E1PQPER	SKILL QUALIFICATION PERCENTILE (FOR PQDES)
E1PMOST	PRIMARY MOS IN WHICH TESTED
E1PSQDT	DATE OF LAST CHANGE ON PMOS TESTED (SQT)
E1PMOST1	PRIMARY MOS IN WHICH TESTED, FIRST PRIOR
E1PMOST2	PRIMARY MOS IN WHICH TESTED, SECOND PRIOR
E1PRQDT	DATE OF PREVIOUS CHANGE IN PMOS TESTED
E1PRDES	PREVIOUS PRIMARY MOS IN WHICH TESTED
E1PRQSC	SQT SCORE FOR PREVIOUS MOS (PQDES)
E1PRPER	PREVIOUS SQT PERCENTILE (FOR PQDES)
E1SQDES	SECONDARY MOS SQT
E1SSQDT	SMOS SQT DATE
E1SQSCR	SMOS SQT SCORE

2.1.3.7. Reenlistment Eligibility and Conditions

A final indicator of each soldier's value to the Army is whether the soldier is eligible for reenlistment and in fact does reenlist. The specific EMF variables required to capture this information include:

E1EREUP	REENLISTMENT ELIGIBILITY
E1EREUPP	REENLISTMENT ELIGIBILITY BAR
E1VRPMO	SELECTIVE REENLISTMENT BONUS MOS
E1VRMUL	SELECTIVE REENLISTMENT BONUS MULTIPLIER
E1VRGRD	SELECTIVE REENLISTMENT BONUS PAY GRADE
E1VRRDT	ENLISTMENT/REENLISTMENT BONUS DATE
E1VRPNR	ENLISTMENT/REENLISTMENT BONUS PAYMENT NO.
E1VRTRM	ENLISTMENT/REENLISTMENT BONUS PAYMENT TERM
E1PSVCI	NUMBER OF TIMES ENLISTED/REENLISTED

2.1.4. SKILLS QUALIFICATION TEST (SQT) DATA

Special datafiles will be obtained containing SQT score information for soldiers in the FY81/82 accession cohort. These data will significantly expand the SQT information available in the EMF by adding scores on tests not released for operational use and by adding information on the particular form (skill level, test year and track) completed by each soldier. The specific data elements to be included, apart from identifying information used only for linkage, are:

P1MOS	MOS IN WHICH TESTED
P1SKLLVL	SKILL LEVEL TESTED
P1TRACK	FORM OF SQT AT THIS LEVEL
P1YEAR	SQT TEST YEAR
P1TESTDT	DATE OF TESTING
P1SQTSCR	SQT SCORE

2.2. FY83/84 Cohort Data

The data assembled for the FY83/84 cohort will include all of the data assembled for the FY81/82 cohort from existing sources plus a considerable number of new measures developed by the project. It is not possible to specify the exact variables at this time, but a summary of these new measures is included below.

2.2.1. INITIAL PREDICTOR DATA

All of the application and accession variables collected for the FY81/82 cohort will also be assembled for the FY83/84 cohort. One significant change in these data is that Forms 11, 12, and 13 of the ASVAB will have been introduced. In addition, Task 2 will develop and administer batteries of additional predictor measures.

2.2.1.1. Preliminary Battery

A preliminary battery of predictor measures will be administered to special samples of about 2,100-4,600 FY83/84 accessions from October 1983 to June 1984 in each of four MOS:

MOS	Title	Training Site
05C	Radio TT Operator	Ft. Gordon, GA
19E/K	Tank Crewman	Ft. Knox, KY
63B	Vehicle and Generator Mechanic	Ft. Dix, NJ Ft. Leonard Wood, MO
71L	Administrative Specialist	Ft. Jackson, SC

These data will be collected during the first week of the soldiers' advanced training (AIT). Training school achievement measures (developed in Task 3) will also be collected as enlistees pass through these training courses and will be used as criteria in the initial analysis of Preliminary Battery measures.

This preliminary battery will focus on types of predictors not currently in use. Analysis of these measures will allow an early determination of the major human attributes not assessed by the current pre-induction battery, and whether the measurement of these attributes significantly increases the accuracy with which performance is predicted. This information will be useful for guiding the development of new predictors into areas most likely to increase the accuracy of prediction and classification.

The Preliminary Battery must necessarily be made up of "off-the-shelf" instruments, because there is too little time prior to the scheduled administration of the Preliminary Battery to develop and pilot test new measures of constructs deemed potentially useful. The testing will probably be done within a four-hour time, since soldier time at AIT schools is generally allocated in four-hour blocks. That time period is sufficient to administer "off-the-shelf" measures of biographical information, vocational interest, motivation, and cognitive ability. Psychomotor measures will probably not be included in the Preliminary Battery because of the time constraints.

2.2.1.2. Trial Predictor Battery

A trial battery of predictor measures, following pilot testing for practice effects, fakeability, and motivational set (with the pilot test administered to samples of the FY81/82 cohort) will be administered to an average of 500 soldiers in each of the 19 MOS. These data will be collected between June and October 1985 from FY83/84 cohort members who will generally be in the third year of their first term of enlistment. Job-performance criterion data will be collected for these same soldiers for use in a concurrent validation of the Trial Battery. The current plan is to develop a Trial Battery that will require a maximum of four hours to administer, including computer-administered and apparatus measures.

In addition to the collection of these primary data, four research projects will be undertaken. First, to measure test-retest reliability, the predictor battery will be readministered to a subsample of 500 soldiers 30 days after the initial administration. Second, to measure practice effects, a subsample of 115 soldiers will be readministered the battery in the week following the first testing. Third, to measure fakeability, 115 soldiers will be instructed to "fake good" and another 115 soldiers will be instructed to "fake bad" on the non-cognitive portions of the paper-and-pencil battery. Finally, to measure score differences between "early career" soldiers (i.e., new recruits) and the primary sample (later career soldiers) in examining maturational effects, a sample of 1,000 new recruits will receive the battery.

2.2.2. TRAINING MEASURES

Currently available training measures will be obtained from the school records for input into the LRDB. In addition, scores from job knowledge tests, MOS content tests, performance ratings, and end-of-course knowledge tests (EOCKT) will be added to the file.

2.2.2.1. Available Records

Training performance measures that have been identified in Task 3 as adequate indicators, based on interview data and on qualitative analyses, will be obtained from school records for all recruits in the FY83/84 cohort who receive training in one of the 19 MOS selected for this research project. Task 3 staff will arrange for the school to provide the required training data to be input into the LRDB on a continuing basis from July 1983 to September 1984, as each new class completes training.

2.2.2.3. Prototype Measures

Preliminary and revised prototype performance measures will be administered to samples averaging 575 soldiers from four MOS: 05C, 63B, 19E/K, 71L. Different test formats will be examined, including free response measures and synthetic hands-on performance measures. In addition, measures of general performance in training and new indices from existing measures will be obtained for samples from all 10 MOS. The data collected on these prototype measures will be analyzed to determine the relative feasibility and value of the administration of each type of measure.

2.2.2.4. End-of-Course Knowledge Tests (EOCKT)

Revised EOCKT will be gathered on samples averaging 500 soldiers from the 19 MOS. These will be obtained at the same time that the other performance measures are obtained for the FY83/84 cohort.

2.2.3. FIRST TOUR PERFORMANCE MEASURES

Concurrent with the administration of the Trial Predictor battery (during the latter half of 1985, see section 2.2.1.2), Army-wide performance measures will be collected from the same 19 MOS samples. These data will include rating scale measures and behavioral indices generated from records of commendations, disciplinary problems, and attrition. For half of these samples (9 MOS), MOS specific performance measures will also be administered. The tentative list of MOS includes the following:

11B	Infantryman
13B	Cannon Crewman
19E/K	Tank Crewman
05C	Radio TT Operator
63B	Vehicle and Generator Mechanic
64C	Motor Transport Operator
71L	Administrative Specialist
91B	Medical Care Specialist
95B	Military Police

The measures used will include hands-on task performance tests as well as job knowledge tests and supervisor and peer ratings.

2.2.4. SECOND TOUR PERFORMANCE MEASURES

Army-wide and MOS-specific performance measures will also be collected from the FY83/84 cohort during their second tour (June 1988 through September 1988). Samples of about 100 soldiers are expected for each of 10 different MOS (05C, 63B, 71L, 19E/K and 64C, 76Y, 91B, 94B, 11B, 13B) for which first tour MOS-specific performance measures are obtained. The measures will be revised versions of the first tour performance measures.

2.3. FY86/87 Cohort Data

The data collected on the FY86/87 cohort will be parallel to the data collected on the FY83/84 cohort, except that concurrent predictor measures will not be collected on the FY86/87 cohort. Data from existing accession and EMF records will be gathered along with data from the predictor and criterion measures developed by this project.

2.3.1. EXPERIMENTAL PREDICTOR BATTERY

From March 1986 through February 1987, the revised predictor battery will be administered to samples of recruits at the beginning of AIT. Current plans call for testing an average of 2,200 recruits in each of the 19 focal MOS. (Data will be collected from additional MOS if preliminary analyses indicate that other MOS are required to assure sufficient validity generalization.)

2.3.2. TRAINING DATA

Training performance data will be obtained from schools for the FY86/87 cohort sample who receive training in the 19 focal MOS between March 1986 through May 1987. The measures collected will include the EOCKT as well as those prototypical measures that prove feasible and valid.

2.3.3. ARMY-WIDE PERFORMANCE MEASURES

The Army-wide (Task 4) and the MOS-specific (Task 5) performance measures administered to the FY83/84 cohort will be revised on the basis of analyses of these data. The revised performance measures will be administered to analogous samples of the FY86/87 cohort for use as final validation criteria.

2.3.4. SECOND TOUR DATA

Again, revised Army-wide and MOS-specific performance measures will also be administered to samples of the FY86/87 cohort who remain for a second tour of duty.

3. EDITING SPECIFICATIONS

For each set of data to be entered into the database, a detailed set of editing specifications will be developed, reviewed, revised, and implemented. These specifications will give procedures for linking the new data to existing records, identifying erroneous or improbable values, correcting these values, and replacing missing values where appropriate. Editing specifications for the FY81/82 cohort training data are given below as an example.

3.1. Editing Specifications for FY81/82 Training Data

3.1.1. LINKAGE TO OTHER FILES

Prior to the detailed editing of each field, the 1981 training data will be linked to the FY81/82 Accession data file and to the 1982 year-end EMF file. The reason for this prior linkage is two-fold. First, the 1981 training data file contains records on some number of soldiers who are not of interest to the current study. These include soldiers not in the regular Army, soldiers who actually entered prior to FY81, and soldiers who are actually reenlistees. By eliminating these soldiers first, editing resources can be concentrated on the cases of primary interest. The second reason for prior linkage is that the information from the Accession and EMF files will provide important checks on the reasonableness of the training data fields and will provide information essential to the correction of missing or invalid values.

The linkage of additional data will be accomplished in two stages. The first stage will involve matching the training records to a special "Link" file which contains identifying information on the soldiers of interest. (See the discussion of the "Link" file in Section 4.3.) For the training records that match a record in the Link file, identifying information will be stripped and replaced with the scrambled identifier from the Link file. This scrambled identifier serves as the primary key for matching data already in the database. The second stage will be to merge the training data with other information in the database using the scrambled identifier.

Two passes will be used in the initial match to the Link file. The initial pass will match on SSN. For each training record which does not match a Link file record, a second match will be attempted. The purpose of this match is to identify errors in the coding of SSNs. In this second pass, the training records will be matched to the

Link file on the basis of the name field (actually the first 5 characters of name) and on MOS. (It is expected that each of the initially unmatched training records may match many Link file records.) For each match, the SSNs will be compared and a new variable, NMATCH, will be computed as the number of matching digits. A frequency distribution will be run on NMATCH to determine an appropriate cutoff point for accepting a match. (We currently expect to accept matches with 7 or more digits in common.) Accepted matches will have all identifying information replaced with the scrambled identifier and will be merged with the main datafile by scrambled identifier. In addition, a dummy record with the alternate SSN will be inserted into the Link file for use in future matches.

3.1.2. ELIMINATION OF DUPLICATE RECORDS

The training datafile is known to contain both exact duplicate records and also valid instances of multiple records for the same soldier due to recycling. The next step in the editing process will be to eliminate the exact duplicates and create a CYCLENO variable (which numbers the training courses taken by an individual soldier sequentially beginning with 1 for the course with the earliest enrollment date) for other instances of multiple records. The CYCLENO variable, together with the soldier's scrambled SSN, will uniquely identify each valid record in the training file.

The first step in this process is to eliminate all records where the preceding record contained identical values for all fields of interest; in this case, all fi except name. After this has been completed, a second p will be made to identify obviously valid recycles. The file will be sorted by ID and by T1GRDDTE (graduation/recycle date). The first record for each soldier will have CYCLENO set to 1. Subsequent records will be accepted as valid and have the CYCLENO variable increased by 1 if the following conditions are met:

(1) the disposition variable (T1DISP) in the preceding record has a value of A or B (recycle or transfer)

(2) the T1GRDDTE variable for the current record is at least 10 days greater than the T1GRDDTE value for the preceding record. (A frequency distribution on this difference will be run to check the reasonableness of this cutoff date.)

All duplicates not meeting these two criteria will be sent to an error file, printed, and inspected by hand for further resolution. It is expected that these records will either be true duplicates with data entry errors or valid recycles with errors in T1DISP or T1GRDDTE.

3.1.3. INDIVIDUAL FIELD EDITS

The editing specifications for each field are given below. (See Section 2.1.2. for a list of the variable names to be edited.) In each case, error records are to be listed individually and manually inspected for error resolution. Where a large number of errors occur in a given field, machine corrections will be developed as appropriate. In each case, a default procedure for the imputation of missing or invalid data is given.

a. T1MOSAWD: values must match the list of valid MOS for this field. A cross-tabulation of T1MOSAWD by A1TRNMOS (training MOS from the accession file) is to be run to resolve invalid values. For records where the T1MOSAWD code is invalid or missing and an EMF record has been linked, the E1PMOS and E1DMOS variables will also be used in error resolution.

b. T1SCHOOL: must be a valid school code for this MOS.

c. T1COURSE: must be a valid code for this school and MOS.

d. T1CLASS: must be a valid code for this course and school.

e. T1SKLLVL: must be a valid code for this MOS.

f. T1ENRDTE: must be a valid date, less than T1GRDDTE, and greater than or equal to A1ENTDTE. A distribution will be run on the number of days between A1ENTDTE and T1ENRDTE to establish an appropriate cutoff for the identification of outliers. (Note that this edit may also catch errors in A1ENTDTE.) In most cases, T1ENRDTE must be identical for specific course and class codes. In such cases, the modal value will be substituted for missing or invalid values.

g. T1GRDDTE: must be a valid date and greater than T1ENRDTE. Graduation date values will also be compared with the modal value among graduates of the same course and class. For recycles and attritions, the value must be less or equal to the modal value except in the case of self-paced classes.

h. T1DISP: must be a valid code for this field. A table of T1DISP by T1ATTRIT will be examined to determine valid combinations. Basically, T1ATTRIT should be blank for graduates (F, G, or H) and for progressive transfers (E) and nonblank for recycles, attrition transfers, and relief (A,B, or C).

i. T1ATTRIT: must be a valid code and consistent with T1DISP as specified above. If attrition is indicated prior to 30 September 1982 and an EMF record is matched, the attrition code will be compared to E1CHSEP (character of separation) and E1SEPTT (type of separation) and T1GRDDTE compared to E1SEPDT separation date. Frequencies and cross-tabulations will be run to determine which combinations are to be treated as errors.

j. T1SCORE1: Frequency distributions will be run for each school, MOS awarded and course to determine cutoff values for the identification of outliers. For some MOS, the scores will be compared to the T1DISP and T1ATTRIT values to assure that scores are either missing or below a cutoff value for recycles or academic attritions. Existing documentation and subsequent inquiries will be required to complete the specification of the treatment of the field and to create a type of score value, T1STYPE1, that allows for proper interpretation of this field.

k. T1SCORE2: For certain MOS, a second score was recorded. This field will be created with appropriate analyses of outliers in accordance with existing documentation. A second score type variable, T1STYPE2, will also be created. Two additional variables will be generated from data initially in the T1SCORE2 field. For MOS 05B and 05C, a variable T1MORSE will indicate completion of Morse code training. For MOS 11X, the actual MOS awarded will be determined and a variable, T1SELECT, created to indicate whether the awarded MOS had been originally guaranteed.°

3.1.4. MACHINE CORRECTION OR IMPUTATION

After manual inspection of all error records, resolvable cases will be updated and the initial edit will be rerun. For cases where missing or invalid values remain, imputed values will be substituted. (Each variable imputed will also be flagged with a binary flag so that imputed values can be identified and, if desired, deleted in later analyses.)

For the categorical variables, "predictor" variables are already indicated in the above consistency edits. In each case, imputed values will be generated randomly with probabilities proportional to the conditional distribution of the variable in question (conditioned on the values of the predictor variable(s)). In many cases, this simply means substituting the one school code where this MOS is taught if the school code is missing, or making the course code consistent with the school and MOS codes. In other cases, values may actually be generated probabilistically.

For continuous variables (T1SCORE1 and T1SCORE2), the SAS procedure PROC IMPUTE will be used to generate imputed values from initial ASVAB test scores.

In all cases, the exact details of the machine procedures for error resolution will be refined using information from the outcome of the editing procedures.

3.2 Editing Other FY81/82 Data

The editing of other FY81/82 data (Accession, Applicant, EMF, and SQT) will proceed in a similar fashion. After initial linkage, editing will proceed variable-by-variable, using the best available information to test or correct the data in each field. Copies of the appropriate Army Regulations will be obtained to aid in the editing as well as the documentation of each field.

4. DATABASE STORAGE AND ACCESS PROCEDURES

4.1 The Use of RAPID

At the time that the proposal for this project was developed, RAPID was identified as the most cost-effective database management system (DBMS) that meets Project A needs. This decision was based on three important features of RAPID. The first was the storage and access mode employed by RAPID. RAPID uses a "transposed file" organization, which means that it stores together all the information on a single variable rather than all of the information on a single "case" or respondent. It stores the data in a direct access file with appropriate indices so that it can read selected variables without having to read through the entire file. The standard statistical packages, in contrast, employ a sequential access mode and store data by case. Even when only a few cases and variables are required, the entire file must be read in order to select the desired information. Most other common DBMSs do use direct access files, but still store information by case so that they only add additional overhead in accessing selected variables.

The second important feature of RAPID is that it provides for a significant degree of data compression. This means that it will be feasible to store much more of the data on mass storage units, greatly increasing the speed with which these data can be retrieved in comparison to tape storage.

The final advantage of RAPID is that it provides convenient interfaces with both SAS and SPSS (as well as other) statistical packages. This facilitates the creation of special analysis files and the use of SAS to manipulate data to be loaded into the database.

4.2. Anticipated File Structure

RAPID is a "relational" database system. It processes a series of "relations" which may be viewed as data tables where the columns are different variables and the rows are different observations. Each row is "identified" by one or more columns which provide the keys for accessing the information in the table. Each row must have a unique combination of key values.

Relations are normalized if they contain no "redundant" information. This frequently means creating several subfiles with different fields. In the FY81/82 training file, for example, the MOS, school, course, and class information are constant for all soldiers in the same class. A more efficient storage of information would result from maintaining course and class information in a separate relation (file) with only one entry (row) for each course and class and then keeping only an index to this information on the individual soldier records.

Determining the "optimum" arrangement of data into separate files or relations requires analysis of the trade-off between reduction in data storage requirements and reductions in processing costs, when only the smaller file(s) need to be accessed, and the corresponding increase in processing costs, when it is necessary to join information separated into different files. At present, we can only forecast requirements approximately so an exact optimization is not possible. During the course of the project, statistics on actual access requirements will be used to reevaluate our file and subfile design.

The organization of data into files currently planned is given below along with some discussion of the rationale behind the proposed organization. Table 1 summarizes the different file types that are planned and gives a three-character designator for each type that will be used as a prefix in the file name.

TABLE 1

LRDB FILE DESIGNATORS

PSF - Primary Soldier Files, one file for each cohort, one record for each soldier in the corresponding cohort, keyed by scrambled ID.

APF - Applicant File, one file for each cohort, one record for each application not leading to accession, keyed by scrambled ID and application number.

SSF - Sample Soldier Files, a separate file for each of the MOS selected for special data collection (FY83/84) and FY86/87 cohorts only), keyed by scrambled ID within file.

SPF - Soldier Progress File, one file for each cohort, one record for each EMF record pulled (tentatively 4 EMF records per year) for each individual in the corresponding Primary Soldier File, keyed by scrambled ID and month of enlistment.

FTF - Field Test Files, one file for each field test event, one record for each soldier tested, keyed by scrambled ID.

MOS - MOS Files, one file, one record for each MOS, keyed by MOS.

TSK - MOS/TASK Files, one file, one record per MOS and Task, keyed by MOS and task code.

4.2.1. PRIMARY SOLDIER FILES

There will be three primary soldier files, one for each of the three main cohorts. These files will contain all of the "constant" information on each accession in the cohort (i.e., each accession during the period that defines the cohort). This information will include all information from the current accession record, information on the completion of training, and information on reenlistment decisions. This file will be keyed by soldier identifier (scrambled SSN).

An abbreviated primary soldier file will be maintained for each of the gap periods between the three main cohorts. These files, which will contain only accession information, will be of primary use to Project B in the development of forecasting models.

4.2.2. APPLICANT FILES

A separate applicant file will be maintained for the accession period corresponding to each of the three cohorts. This file will be keyed by the same scrambled identifier used in the Primary Soldier File and by an occurrence number within each individual ID. There will be one record for each application of each individual. In order to avoid duplication, only application information not leading to an accession of interest will be kept here. By concatenating these files with the Primary Soldier Files, however, a complete set of application data can be obtained.

Each record will contain test scores and other information relating to the particular application including background data that ought not to change from one application to another but might change anyway. These data will be useful in establishing overall base rates for applicants and for looking at the level of consistency in different variables across applications.

4.2.3. SAMPLE SOLDIER FILES

For the FY83/84 and FY86/87 cohorts, there will be Sample Soldier Files consisting of all of the soldiers sampled for special data collection. Currently, we plan to maintain 19 separate files corresponding to the 19 different MOS sampled for new data collection. This will facilitate the creation of separate analysis files for each MOS. It will, of course, be a simple matter to concatenate these files for across-MOS analyses.

The Sample Soldier Files will also be keyed to an alternate identifier defined as the index number of the corresponding Primary Soldier File record. They will not contain any other variables stored in the Primary Soldier Files, but they will be directly linkable to the Primary Soldier Files by this index number (without further sorting). These files will contain all of the new measures collected on each soldier in the selected samples. Some of the measures collected will vary from one MOS to the next, particularly the MOS performance measures and the job knowledge and hands-on measures collected during training. (This is a major reason for maintaining separate files by MOS.) It is likely that these files will also be further divided by data collection period. The FY83/84 second-tour sample, for example, will be only a subset of the concurrent validation samples, and the concurrent validation samples will also be different from the samples receiving the Preliminary Predictor Battery.

4.2.4. SOLDIER PROGRESS FILES

Separate Soldier Progress files will be used to store recurring information on each soldier's progress in the Army. There will be separate Soldier Progress files for each cohort. These files will be keyed by soldier ID and a generated variable TOURMON which gives the number of months since the beginning of active service. The contents of these files will come primarily from the EMF, which will be accessed at regular intervals, and from special SQT files. The primary purpose of these files is to provide the basis for time series or career trajectory analyses in which each soldier's progress is charted as a function of time in service.

4.2.5. FIELD TEST FILES

A separate datafile will be created for each field test of each new instrument or battery. These files will be keyed by alternate identifier (index number in the relevant Primary Soldier File) so as to be readily linkable to all other information on the same soldiers. The contents of each file will be highly specific to the related field test.

4.2.6. MOS FILES

A separate file will be maintained which will contain information on the characteristics of each MOS. This file will be keyed by the three-character MOS code. The specific contents of the file are not fully known at this time. Some information on qualifications for each MOS, workforce size and requirement forecasts, training location(s), and utility measure data collection will be included.

4.2.7 TASK FILES

Information on specific tasks performed within each MOS is available from several sources (e.g., the Army Occupational Survey Program, the Soldier's Manual, the RCA study of prerequisite competencies using TRADOC sources). In developing both training and MOS-specific performance measures, it will be desirable to maintain a file of these tasks for at least the MOS selected for special data collection.

4.3. Updating Procedures

Formal updating of the LRDB will be carefully controlled by the LRDB manager. It is essential that this be an orderly process to protect the integrity of the database. Consequently, the procedures for modifying the LRDB will be made available only to the database manager and to the ARI database monitor. Other requests to use these procedures for creating/updating other (non-LRDB) files will be evaluated on merit and granted only with the approval of the ARI monitor and the database manager.

In many instances, file updates will involved adding derived variables or indices. In the course of analyses, a large number of such variables will be added to workfiles. Where the general applicability of such variables is judged to warrent the increase in storage space, these variables will be added to the master database.

The process of adding new data to the file will involve several steps. These steps are designed to minimize the need for further changes or corrections once the data become available. Insofar as possible, such changes will be strictly avoided so as to eliminate the need for rerunning significant numbers of analyses to reflect corrected data. The steps to be followed in updating the file include the following:

4.3.1. IDENTIFICATION AND ACQUISITION OF NEW DATA AND RELATED DOCUMENTATION

In the case of the acquisition of existing data, this step will be relatively simple. For new measures to be collected, however, the database staff will expect to play a more significant role in the design of the data collection instruments to facilitate data entry.

For data that are not now in machine-readable form, the DB staff will provide for data entry. Plans call for the use of the interactive entry/edit system (FORMSPEC) available at AIR's Washington Office. If similar software can be installed at NIH, we will switch to entering data directly into NIH.

4.3.2. LINKING IN RELATED DATA

A separate Link file will be maintained to facilitate the addition of new data. This file will contain basic identifying information (SSN, name, birthdate, primary MOS, race, sex) and pointers to (index numbers) records in each of the relevant relations (files). Each new dataset will be passed against the Link file. For each matching record, all identifying information will be deleted and replaced with the appropriate pointers. For initial nonmatches, a second attempt will be made to match to the Link File on the basis of secondary identifiers, including, if necessary, manual inspection of "close" matches. For many new data sources, there will be a number of cases that are not already in the Link file. Where it is desired that such cases be kept, relevant information will be added to the Link file and the data will be retained. This will be the case only if a new relation is being established so that it should cause no problem with the index values stored in the Link file.

Once all links to existing data have been established, the existing data needed for editing will be pulled out of the database and merged with the new dataset. It is expected that this merge will be accomplished using SAS, since the edit procedures are designed as a SAS application.

4.3.3. EDITING

The editing procedures are described in detail in Section 3. They will typically involve two passes. In the first pass, specifications to detect errors and improbable values will be developed and implemented. After inspecting the results of this editing pass, error resolution specifications will be developed and implemented as a second pass.

4.3.4 DOCUMENTATION

Following the completion of the editing process, the documentation of the new data will be accomplished. A central part of this activity will be establishing the codebooks including frequencies and descriptive statistics as described in Section 5 of this plan.

4.3.5. MERGING THE NEW DATA

After the documentation has been completed, reviewed, and revised as necessary, the new data will be formally merged into the DBMS and appropriate backup tapes will be created (using the RAPID UNLOAD procedure).

4.3.6. DISSEMINATION

The final step in the addition of new information to the LRDB will be to inform potential users of the availability of the data and the documentation for the data. This will be accomplished through the electronic bulletin board implemented as part of the project sign-on procedures and through mailing to a list of ARI and project staff designated to receive information on the database. This mailing list will be established and reviewed by the Project Director and Principal Investigator and by senior

ARI staff assigned responsibility for monitoring this activity.

4.3.7. EXCEPTIONS

It is expected that there will necessarily be exceptions to this orderly process. The most common form of exception is when quick analyses are required even though the data have not been completely edited. In most cases such analyses can proceed with the completion of step 2 (linking) and run in parallel with the editing. In a few cases, it may be necessary to strip identifiers and proceed with a copy of the input data only. In any event, the establishment of an orderly process makes the exceptions clearer. If preliminary analyses are run, they will be designated as such and checked as needed once the full update process has been completed.

4.4. Access

Primary access to the LRDB will be through the SAS interface procedure, PROC RAPRD. The Task 1 staff are all experienced SAS users and plan to conduct most of the analytic investigations using the SAS package. SAS is also the package of choice because of its capacity for merging and transforming data easily. We intend to further simplify access to the data by creating a WYLBUR Command Procedure that will take a file name and variable list and create the SAS set up to read the requested variables into SAS and attach all of the appropriate variable labels and formats (for value labelling). This procedure will greatly simplify authorized access to the database and will also contain a log file to monitor such accesses. (As discussed below, in Section 6, we will also place a logging procedure within the catalogued procedure that accesses the database, as a further control on access.)

After the first portions of the database are loaded, we will conduct a small cost-analysis to determine the relative efficiency of using RAPID operations to join information stored in different relations in comparison to using SAS merge operations to accomplish the same objective. This issue is not of major concern since SAS will accomplish this objective with reasonable efficiency, but it is of interest in cases where access from other procedures is required or where very large datasets are being created.

In addition to providing access to the database through SAS, we will implement procedures for generating SPSS systems files and also raw data files. A TPL interface is available, and we will install it if any need becomes apparent.

In general it is expected that requests for analysis files will be channelled through either the Project's Database Coordinator or ARI's monitor for this activity (who may also then pass the request on to the Database Coordinator). This pattern is expected to act both as a means of assuring a close monitoring of access to the data and also to insure reasonable efficiency since the database staff will be most knowledgeable about the database contents and access procedures. Except in very high priority cases, it is expected that workfile creation runs will be created overnight during the discount period. With the WYLBUR command procedures in place, we expect that workfiles can be created with only a 24-hour turnaround in most cases.

5. DATA DOCUMENTATION AND DISSEMINATION

5.1. Documentation Formats and Standards

Because the project involves the simultaneous collection and analysis of many interrelated sets of data by different teams of researchers, it demands particular effort in clear and complete documentation of the database. This effort is complicated by the fact that the database will not be constant, but rather will grow throughout the project as new measures are developed and new data are collected. It is essential, therefore, that the system for data documentation be carefully developed and strictly enforced from the outset of the project.

We will implement a multilevel system of interrelated data documentation documents that together will allow users easily to gain complete information on the data they may need to use. The key elements of this "metasystem" of documentation include:

- An Event File that documents each data collection "event";

- An Instrument File that contains copies of the data collection instruments (including both questionnaires/tests and answer sheets where separate;

- A Sample Structure File giving a list of the different samples used in the database and showing their relationship in Venn Diagram type format;

- A Dataset Log that shows the name, characteristics, and location of each data set in the database and refers to the appropriate codebook documentation of the data set;

- SAS Codebooks for each data set, including frequency distributions for each discrete variable and complete summary statistics for continuous variables; for derived variables, the computational formula will be indicated; for other variables, the source file will be shown;

- Variable Cross-Reference Files listing each of the variables in the database topically and by variable name and giving a list of all of the data sets for which the variable is available;

o <u>Data History Documentation</u> for each data set, including an overall flowchart showing the steps and workfiles in the file creation/editing process and the printed output from each step in this process.

Each of these logs or files is described more fully below. We plan to use the WYLBUR text entry/editing capabilities to maintain "on-line" versions of all but the instrument file (where only the contents and index will be on-line). Hardcopy versions of each of these text files, as well as the instrument file, will also be maintained to facilitate the production and distribution of new copies of the complete documentation package for staff and others requiring such documentation.

5.1.1. EVENT FILE

The event file gives the basic "who, what, when, why, where" of each data collection effort. Specifically, each entry will include:

1. The date(s) and place(s) of the data collection events (e.g., FY81, at all MEPS; or June 23, 1983, at Fort Knox);

2. The sample(s) from whom data were obtained including the identifier used to access the Sample Structure File;

3. The instrument(s) used (including the instrument "identifier" used in accessing the Instrument File); and

4. A concise description of the purpose and intended use of the data. (For data collected by project staff, this will be a summary of the justification statement developed prior to the data collection and will refer to the more complete statement.)

5.1.2. INSTRUMENT FILE

Researchers occasionally need access to the original data collection instruments for such purposes as checking the actual wording of particular questions, checking potential skip patterns, and generating hypotheses concerning oddities in the responses. In many systems of data documentation, codebooks are constrained by variable and option labelling that must fit the format of the particular system employed. As a result, the full text of the question of the response alternatives is not available to the analyst. In addition, the "context" of the question is not apparent in most codebooks. We will maintain a complete file of all of the instruments used, organized by an instrument identifier and accessed by instrument name and by a topical index of the instruments. Security restrictions may apply to some test instruments (e.g., ASVAB forms). We will investigate ways of satisfying security concerns in such cases.

The instrument file will be maintained in hardcopy form suitable for efficient copying on a Xerox 9400 as copies are needed for new project staff members or other researchers.

5.1.3. SAMPLE STRUCTURE FILE

Each time a new data set is received, the sample (and subsamples where appropriate) on which the data are based will be identified and assigned a sample identifier. This identifier, together with a more complete labelling of the sample(s), will be entered into the Sample Structure File that logs each sample and points to the relevant data set(s). The degree of overlap with every other sample will be ascertained and recorded in a sample structure matrix.

5.1.4. DATASET LOGS

Each of the raw and SAS data files comprising the database will be listed in a data set log. This log will show all versions or generations of each data set beginning with the initial tape(s) or card(s) received from the field or from the data entry vendor. For each data set, the location (e.g., NIH tape library, NIH disk pack, backup facility tape library) will be indicated along with the primary data set characteristics (storage mode, block size, and record size where appropriate) and pointers to relevant entries in the Sample and Instrument Files. Much of this information will be maintained in the operating system's on-line catalog for the current, operative version of each file.

5.1.5. CODEBOOKS

Detailed information on each variable in the database will be provided in SAS codebooks. The codebooks will be organized by dataset (relation) and data collection instrument with file and instrument identifiers indicated on the heading of each page. The specific contents of the codebooks will include:

1. a variable name and a more complete description for each variable;

2. a summary of the characteristics of the variable (character or numeric, number of characters/number of decimal places);

3. the number of cases with valid responses and the number for which the variable was omitted or missing;

4. a label for each response option for all discrete variables; and

5. the actual frequency distribution for each discrete variable and appropriate summary statistics (mean, standard deviation, median, qua tile points, minimum and maximum) for each continuous variable.

5.1.6. VARIABLE CROSS-REFERENCE FILE

The Variable Cross-Reference File will contain an alphabetic and a topical listing of all of the variables in the entire database. For each variable, the appropriate instruments, data sets, and samples will be indicated. The topical index will be of particular importance in providing researchers a means of getting an efficient overview of the system and determining the availability of data to meet specific needs. During the planning phase, an initial variable taxonomy will be developed, and this taxonomy will be expanded during the project as appropriate. In developing this taxonomy, multiple listings for variables will be assumed (e.g., initial ASVAB scores might be listed under "Accession Data," under "Aptitude Measures," and also under "Performance Predictors").

5.1.7. DATA HISTORY DOCUMENTATION

Data history documentation will make it possible to examine each step in the creation and editing of the final datasets. This history documentation of each dataset will consist of a flowchart showing the files and programs used at each step in the creation/editing process and the output from each computer run in this process. The output will show both all of the program statements used and any printed results (e.g., record counts or warning messages). This documentation will be maintained on-line while the datasets are active.

5.2. Dissemination

Several different methods will be employed to make information on the database available to appropriate individuals. News of immediate importance will be placed in an on-line electronic bulletin board with headlines announced through each user's logon profile. Similar on-line aids, accessible from Project accounts, will be used to point to the WYLBUR versions of the data documentation described in Section 5.1.

As the data from each new data collection become available, an informal workshop will be held. Printed copies of the documentation will be distributed to authorized users, and special characteristics of the data will be discussed. An initial workshop, held in May 1983, covered data storage and access procedures and information on RAPID, WYLBUR, and NIH computer facilities in general, as well as the detailed contents of the FY81/82 cohort files.

6. DATABASE SECURITY

6.1. The Need for Security

Whenever a large amount of data on individuals is maintained and stored, it is necessary to develop procedures to protect that data from compromise. The security of the Project A and B LRDB is particularly important for a number of reasons. Some of the data collected on individual soldiers, such as promotions, paygrade, or disciplinary actions, will be private in nature, and the privacy of that information must be maintained. Since many researchers will be accessing the LRDB for a variety of uses, the integrity of the data must be maintained to insure that the data remain accurate and consistent across uses. Finally, it is necessary to secure the data base to insure that the Army maintains ownership of the data. In other words, to insure that the data within the LRDB are used only for authorized Project A and B research.

6.2. Security Procedures

The security of the LRDB will be protected in a number of ways. Soldier social security numbers (SSN) will be routinely encrypted to insure the privacy of each soldier's records. Access to the LRDB will be controlled both to further protect soldier privacy and to insure proper use of the data. To provide further physical security, a log will be maintained for the LRDB system that will note each attempted access of the LRDB and whether the access was authorized or not. Finally, a set of data processing practices will be established to provide security for the information managing aspects of data within the LRDB. Each of these procedures will be detailed in the subsections that follow.

6.2.1. SSN ENCRYPTION

The key aspect to guaranteeing the privacy of individual soldier data will be the coding or encrypting of each soldier's identifier. This encryption will be accomplished by scrambling each soldiers SSN in an unpredictable way. The algorithm that will do the encrypting (and if needed, decrypting) will be known only to the LRDB manager and ARI in-house data base administrators. A printed copy of the algorithm will be securely maintained by the Project A COTR. All of the data files of the LRDB that can be routinely accessed and any project workfiles generated from the larger LRDB files will use only this encrypted SSN as the soldier identifier.

6.2.2. CONTROLLED FILE ACCESS

The integrity and accuracy of the LRDB data will be maintained by controlling the access to the large files or relations within the data base. This procedure will also further contribute to the privacy protection of individual soldier records. In general, the system to be adopted will use the RACF procedure available at NIH to allow the access of particular files to authorized users. Under RACF, different levels of access can be granted to different users. By specifying a "universal access" of "NONE," access can be restricted to only those users granted specific exceptions. In most cases, such users will be allowed "READ" access only. Such users will have to provide an eight-character RACF password (different for each user) in order to read the datafiles for which they have been given access. Using the provisions of RACF, a series of access "levels" will be developed which should provide timely access to relevant data needed by Project A and B researchers and yet protect the security and integrity of the data.

Level 1. At the highest level of access will be the data base administrators. Currently these individuals are Dr. Lauress Wise and Ms. Winnie Young of AIR and Dr. Paul Rossmeissl and Ms. Frances Grafton of ARI. Level 1 personnel will have access to all of the files and relations within the data base. Furthermore, only Level 1 personnel will be able to enter data into the data base or modify data already stored in the data base. Thus, the data base administrators must assume responsibility for data entry, editing, and the storage of original data materials (i.e., tapes, punched cards) in a secure location. In addition, it will be the duty of the Level 1 personnel to create Level 4 workfiles as they are needed by other project researchers.

Level 2. Personnel at the second access level will be able to directly read data from all of the files in the data base with the exception of the Link File (see Section 4.3.2), which will contain basic soldier identifying information. This exception is made to maintain soldier privacy. It is planned that two members of the Project A staff will have Level 1 access to the LRDB. Dr. Ming-mei

Wang, the deputy Task 1 leader for statistical analyses, will need to be able to quickly access all data files, since the validation analyses of Task 1 span all tasks and data sets collected within Project A. Dr. Lawrence Hanser, the Task 4 monitor, will also be provided with Level 2 access. Dr. Hanser has been associated with Projects A and B since their inception and will backup the ARI in-house Level 1 LRDB staff in insuring that ARI has complete access to the LRDB for in-house research.

Level 3. Most project personnel will have some Level 3 LRDB access. Researchers at this level will have direct access to all files that are generated by the particular tasks they are investigating. Furthermore, they will have direct access to the files created by other tasks that directly impact their work. For example, Task 2 researchers will have direct access to the task analysis data collected by Task 5 so that the new predictors that are developed will address areas of the criterion space not currently covered by ASVAB.

Level 4. The most common way in which project researchers will access the LRDB is through the creation of workfiles (see Section 4.4.). By requesting the creation of a work file, a researcher will be able to obtain data from all of the large files in the data base except the Link File which will contain soldier identifying information and will always be kept private and secure. The key aspect of workfiles relevant to LRDB security is that the researcher will only receive the data that he or she requested and there will be a precise record of who requested what data. When a project scientist requires a workfile, he or she will submit a data request form to either the contractor or ARI data base administrators. This request form will ask:

(1) Who wants the data?

(2) What variables are needed?

(3) What sample is needed?

(4) Which LRDB file or files contain the data being requested?

(5) Why are the data needed?

(6) Will the data be downloaded to hardware other than the NIH computer facility?

(7) What will be done with the data after its current use is completed (i.e., file will be scratched or saved for future use)?

In addition, each data request form will remind the researcher seeking data that LRDB is the property of the Army to be used only for Project A and B research and that all publications, papers, and briefing charts based upon these data must be submitted to ARI for clearance before they are presented to the public.

Paper copies of the workfile request will be available to all Project A and B scientists, but it is expected that most researchers will make use of a request form that will be stored on-line at NIH and can be quickly sent to a data base administrators using WYLBUR electronic mail. It is expected that most work file requests will be filled using overnight runs at NIH (see Section 4.4.) and should be ready within twenty-four hours of the original request for data.

6.2.3. LRDB LOG

The procedure used to execute the RAPID data base management systems' data retrieval programs has been modified to log a record of each access or attempted access to the data base. This access log file will be reviewed weekly to assure that no inappropriate access has been attempted. In addition, the monthly accounting information of each project user will be monitored for any indication of unauthorized access to the LRDB. These audit trails will serve as a second level of protection against unauthorized use of the data by anyone who manages to obtain the necessary RACF passwords. They will not directly prevent unauthorized access to the LRDB, but the threat of exposure should serve as a significant deterrent to attempts at unauthorized LRDB access. The log will also help the data base administrators decide which project files should be stored on disk rather than tape by providing information as to how frequently data are requested from any given file.

6.2.4. OTHER PHYSICAL SECURITY PROCEDURES

Much of the data that will be entered into the LRDB will come from existing Army sources, such as the EMF. Additional precautions beyond those mentioned above will be taken to secure the information on these data tapes. The key aspect of this additional security is to collect and store information from these sources only if it is essential to the goals of Project A and B. For example, with regard to the EMF, this LRDB plan indicates specifically which variables will be needed. Other variables, in particular, each soldier's location and assigned unit, will not be acquired in any form. In addition to limiting the data elements to be stored, the number of soldiers for whom any data will be retained will be limited. As indicated in Section 3 above, the LRDB will not obtain and keep information on all active service personnel. Only data from personnel selected for Project A and B research will be maintained.

6.2.5. DATA BASE ENTRY AND EDITING SECURITY PROCEDURES

In addition to providing for the physical security of the data base, procedures have been established to maintain soldier privacy within the areas of data base information management. Included within the broad topic of information management are such specific areas as handling of raw data, maintenance of raw data forms, and procedures for dealing with processed data (such as printouts or written reports). This section presents the procedures that will be used to provide security during data entry and editing, while the following section presents procedures that will be followed to protect soldier privacy in the analysis and reporting of data.

Data entry. All forms for data collected in the field will be shipped to the data entry station at AIR in sequentially numbered packages via certified mail. Within 24 hours of their receipt, an entry in a data entry log will be noted for each package. This log will be maintained on-line, but will be backed up by a hardcopy following each log update. The log will contain identification of each package received, the number and type of the documents included, and the current status (entry preparation, entry, verification, editing, or shredded) of the data.

Data editing. While the data is being edited, the data collection forms (the raw data) will be stored in a locked room at a site removed from any post where the individual responses should not be of interest to anyone. Data integrity of the new data will be insured through thorough editing of the data. This editing will include: complete verification of all entered data, a reconciliation of the resultant record counts against the initial document counts, and relational editing of all new data to appropriate existing sources (e.g., the SSNs and birth dates match the master link file). Once the data have been entered into the LRDB and completely edited, the AIR data base administrator will review the completeness of the data entry/editing process. In performing this review, he will consult with the task leader and ARI task monitor responsible for the data collection. Following any further revisions resulting from this review and a final approval of the editing, a back-up copy of the the resulting datafiles will be created that does not contain any personnel identifier other than the encrypted SSN. This tape will be removed from the NIH facilities and stored at a separate location.

Once the data have been backed up onto tape, all of the input documents will be shredded. A final count will be made of the number of documents shredded and this count will be checked against the initial document counts and the data entry log. At the same time that the data input forms are destroyed, all printouts generated during the editing of the data will be reviewed. Edit Run summaries and other general information will be found together to form the detailed documents of the editing process. Any other printouts, including any with potentially identifying information, will be destroyed along with the input documents. Likewise, any computer workfiles containing possible identifying information (excluding the master link file), along with all summary files not needed as backup or documentation, will be deleted from the system and then overwritten. The ARI data base administrator (or someone he delegates) will oversee this entire process.

6.2.6. DATA ANALYSES AND REPORTING SECURITY PROCEDURES

All workfiles, printouts, and analyses produced by the project data base personnel will contain a header indicating that the products were based on personnel data that the product should therefore be handled in an appropriate manner. When researchers are finished with any data, they will be required to specify the disposition of all workfiles and computer printouts that were created during the analyses. If work is of a contuining nature, a list of the workfiles and printouts will be retained for verification at the final completion of the analyses. When all analyses are completed, the ARI reviewer (the data base administrator or the appropriate task monitor) will approve the contents of any workfiles or printed documents that are to be retained. The primary purpose of this review is to assure that no information that might be used to infer individual soldier identities is retained.

All reports, journal articles, and conference papers, based on Project A and Project B research must be cleared by ARI before publication. This clearance process is primarily concerned with the political and scientific sensitivity of the research and typically is composed of three levels of clearance (team chief or task monitor, tech area, and research laboratory). In the case of reports based on LRDB data, these reviews will be expanded to assure that information is not included in the reports that might eventually be used to ascertain the identify of individual soldiers.

6.3. Summary

Any set of procedures designed to store data electronically needs to balance the ease with which data can be accessed against the security of the data base. The procedures presented in this section tend to favor the security aspect of this balance. The number of data files that most project researchers will be able to access directly will be quite limited. Furthermore, only the data base administrators will be able to add or modify data, and access the true soldier identifying information. However, efficient use and rapid creation of the workfiles should provide any project scientist with the data that he or she needs to perform the required research.

IV. VALIDITY ANALYSIS: PRELIMINARY ANALYSIS OF THE FY81/82 DATA FILE

Donald H. McLaughlin (AIR)
Ming-Mei Wang (AIR)
Paul G. Rossmeissl (ARI)

The analyses of the FY81/82 cohort data file will serve several purposes.

(1) The validity and differential validity of the existing predictors (ASVAB 8/9/10) against existing criteria (training grades and SQT) will be determined on all MOS for which there are sufficient data. These results will serve as a benchmark against which the subsequent validations using new and/or improved predictors and criterion measures can be compared.

(2) The validity of alternative composites of ASVAB subtests can be compared with the validity of the existing composites.

(3) The validity generalization pertaining to both existing and alternative ASVAB composites can be modeled using the MOS clusters developed in Task 5.

(4) Alternative analytic methods for estimating the prediction parameters required by the classification system can be developed and evaluated using this data base. That is, what method should be used to combine predictor information into a predicted score so as to maximize classification validity, minimize "shrinkage," and maximize robustness against cohort charges.

(5) The psychometric and distributional properties of the existing criterion measures can be determined so as to better describe their strengths and weaknesses.

The degree to which these objectives can be accomplished is a function of the size and completeness of the FY81/82 cohort data file. A more detailed discussion of these analyses is presented in ARI Research Report 1332.

As noted in the previous section describing the development of the FY81/82 data file, the effort required to accumulate and merge the records, edit the files, and resolve discrepancies has been considerable and is not yet complete. However, it has been possible to begin the preliminary analyses of these data and the initial steps are described below.

Data Usability

The major purpose of the data usability analyses was to determine the degree to which the statistical assumptions underlying the validation analyses might be violated. The univariate distributions of the training and SQT criterion variables were obtained as well as the bivariate distributions of each criterion with each ASVAB subtest. Also, possible transformations of the criterion variables were evaluated using the same methods.

MOS with 100 or more observations on both a training outcome criterion and SQT were selected for analysis. The distributions of training and SQT scores were examined by MOS (by course within MOS for training scores) in terms of historgrams, normal probability plots, and cumulative normal plots to assess the extent to which the criterion scores conform to normal distributions. Scatterplots of ASVAB subtest scores with each criterion score were also inspected to assess the bivariate relationships.

The findings for the univariate analyses are similar across MOS. In general, both criterion variables are negatively skewed, i.e., there are ceiling effects, with the SQT score typically less skewed than the corresponding training score. Figure 5 illustrates a typical normal probability plot as obtained for MOS 94B with SQT score as criterion. (Note that if a distribution is, in fact, normal, the points should lie along a straight line running from the bottom left to the top right of the grid.)

In the case of the bivariate analyses, a consistent pattern emerged. When the plot diverged from a bivariate normal distribution, the pattern is that of a so-called "reverse fan spread," that is, the conditional variance of the criterion variable for high values of the predictor is less than the conditional variance for low-to-moderate values of the predictor. And so, if a recruit did well on the ASVAB he/she is likely to do well on the criterion measures. But if a recruit did not do well on ASVAB, his or her criterion performance is less predictable. This observation suggests possible violation of the homoscedasticity assumption underlying the Ordinary Least Square (OLS) regression method. Figure 6 illustrates the heteroscedasticity of the SQT scores for MOS 94B with mathematical knowledge score on ASVAB as the predictor.

Exploratory analyses were carried out to investigate possible criterion transformations that would alleviate the ceiling effects and heterogeneity of variances. According to the so-called "ladder" of transformations presented by Tukey (Mosteller & Tukey, 1977, p. 79f), bivariate distributions such as these can be "straightened out" by a transformation in the direction of a square or a cube of the skewed variable. Such transformations exaggerate the differences between high values of the variable, placing the modal value closer to the center of the range of scores. The square of the criterion variable was plotted for some selected MOS and those plots were

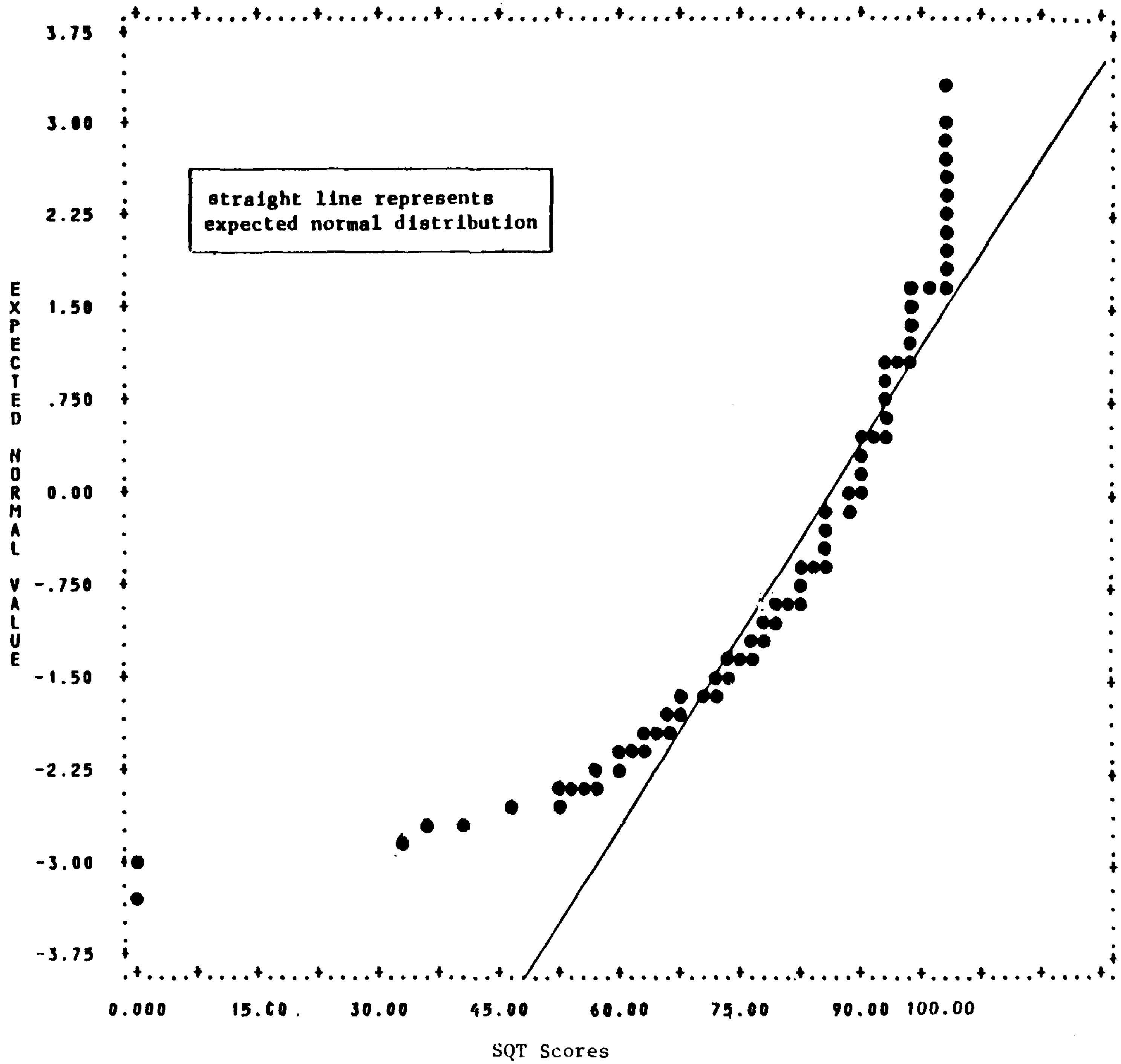


Figure 5. Normal Probability Plot of SQT Scores for FY81 Accessions in MOS 94B.

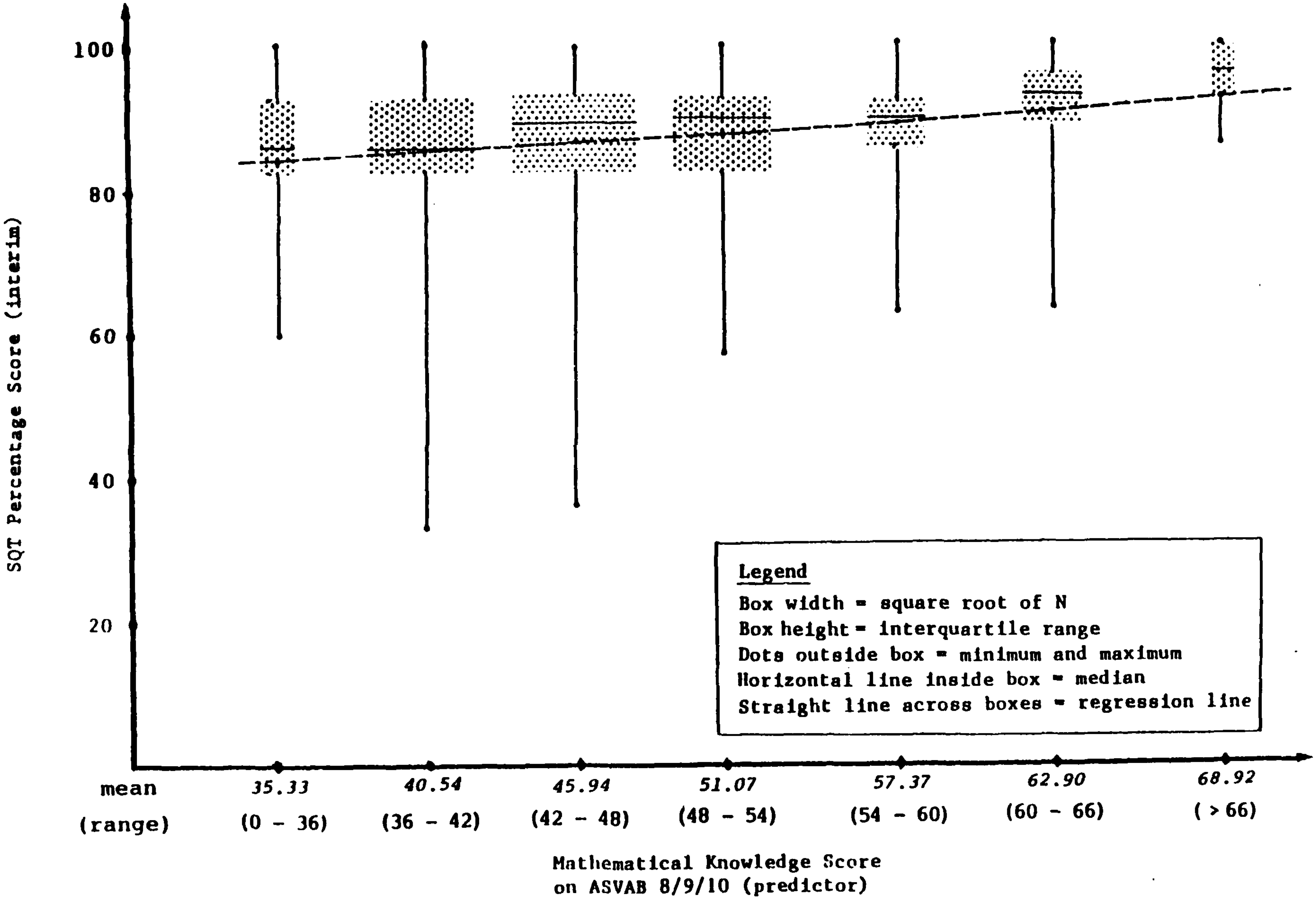


Figure 6. Heteroscedasticity of SQT Scores for FY81 Recruits of MOS 94B.

compared with the plots of the untransformed data. The transformations resulted in little change in the form of the distributions. However, it should also be noted that gross violations of model assumptions are rare. The "reverse fan spread" phenomenon may be a reflection of some kind of systematic error in criterion measurement.

Sample Sizes

A major concern for the 81/82 file data is that the proportion of the 300+ MOS in the enlisted occupational structure should contain sufficient cases to permit appropriate analyses. Therefore, we have attempted to include both training and SQT criterion measures so that separate validation analyses using each criterion can be compared. The training scores were collected as a part of ARI's preparation for this project, and we have been in the process of acquiring SQT data from Army computer files. The sample sizes shown here must be considered approximate, because if the schedule and availability of files permit, we will add (1) more SQT file records, and (2) additional accessions not on the current FY81 accessions file.

The current state of our data base for purposes of ASVAB validation, modeling classification and validity generalization, and determining criterion interrelationships is summarized in Table 4. The size of the sample for a given MOS was governed by the number of nonprior service enlistees who took one of the ASVAB 8/9/10 forms.

In preparing this table, we included MOS for which we have adequate data on at least 100 enlisted personnel, and we separated the data for different

Table 4

Feasibility of Subgroup Validation and Total Sample Size for FY81 Accessions, by Prior Service and Criterion Availability

(Nonprior-Service, Both Training and SQT Criterion Available)

MOS	TCRS	AA	QS	GRPT	TOTTR	GRPS	TOTSQ
05B	2A	SC	90	RACE AND SEX	517	RACE AND SEX	457
05C	2D	SC	95	RACE AND SEX	605	RACE BY SEX	889
11B	IN	CO	85	RACE ONLY	976	RACE ONLY	3430
11C	IN	CO	85	RACE ONLY	557	RACE ONLY	817
11H	IN	CO	85		428	RACE ONLY	572
12B	AB	CO	85		131	RACE ONLY	1015
12F	AF	CO	85	RACE ONLY	198		140
13B	3B	FA	85	RACE ONLY	639	RACE ONLY	1830
13E	3E	ST	95	RACE ONLY	449		280
13F	3F	FA	100	RACE ONLY	659	RACE ONLY	465
19D	9D	CO	85		186	RACE ONLY	527
19E	9E	CO	85		166	RACE ONLY	926
31M	4D	EL	95	RACE AND SEX	586	RACE AND SEX	653
31V	1V	EL	95	RACE ONLY	457	RACE ONLY	303
36C	AA	EL	90	RACE ONLY	179	RACE ONLY	195
36K	AC	EL	90	RACE ONLY	659	RACE ONLY	726
55B	5B	GM	85	RACE AND SEX	191	RACE ONLY	221
57H	G1	GM	85	RACE ONLY	169	RACE ONLY	123
62B	CB	MM	85		221		159
63B	3B	MM	85	RACE AND SEX	899		117
64C	EC	OF	85		199	RACE AND SEX	1403*
64C	4C	OF	85	RACE ONLY	398	RACE AND SEX	1403*
67N	65	MM	100		150		181
73C	5R	CL	95	RACE AND SEX	194	RACE AND SEX	242
75B	5E	CL	95	RACE ONLY	483	RACE ONLY	398
75D	5D	CL	95	RACE ONLY	228	RACE AND SEX	338
75E	5E	CL	95	RACE ONLY	268	RACE ONLY	165
76W	DB	CL	90		132		142*
76W	PW	CL	90	SEX ONLY	204		142*
82C	2C	ST	95		369		188
94B	KA	OF	85	RACE AND SEX	621	RACE BY SEX	1170*
94B	4B	OF	85	RACE AND SEX	625	RACE BY SEX	1170*
95B	SB	ST	100	SEX ONLY	716	RACE AND SEX	1370

(Nonprior-Service, Only SQT Criterion Available)

MOS	TCRS	AA	QS	GRPT	TOTTR	GRPS	TOTSQ
12C	AC	MM	85		66		166
52D		GM	95		0		121
63B	DB	MM	85		3		117
75C	5D	CL	95		85		108

Table 4 (Cont'd)

Feasibility of Subgroup Validation and Total Sample Size for FY81 Accessions, by Prior Service and Criterion Availability

MOS	TCRS	AA	QS	GRPT	TOTTR	GRPS	TOTSQ
(Nonprior-Service, Only Training Criterion Available)							
11O	IN			RACE ONLY	326		0
15D	5D	OF	95		281		95
15E	5E	OF	95		264		74
16B	BA	OF	85		111		3
16E	EB	OF	95		117		0
16R	RA	OF	85	RACE ONLY	296		0
16S	SA	OF	85	RACE ONLY	506		1
17C	7C	SC	95	RACE AND SEX	189		71
17K	GA	EL	85		128		89
19F	9F	CO	85		123		6
27E	7E	EL	95		127		96
31N	4C	EL	95	RACE ONLY	159		84
32D	80	EL	95		116		1
45K	K9	GM	95		121		3
51K	BK	GM	85	RACE ONLY	155		0
54C	SS	GM	95		124		72
61C	H1	MM	100		123		0
63D	SA	MM	100		268		3
63G	M7	MM	100		100		0
63H	H1	MM	85		217		22
63N	TS	MM	95		274		10
63T	FI	MM	100		472		0
63W	W1	MM	85	RACE ONLY	276		4
63Y	TV	MM	100		130		8
67U	P1	MM	100		175		80
67V	18	MM	100		153		78
67Y	S1	MM	100		122		47
68J	W6	GM	95		102		34
68M	W8	GM	90		103		31
71N	L1	CL	95		117		68
76C	EC	CL	95	RACE BY SEX	1137		10
76P	5F	CL	90	RACE AND SEX	557		9
76V	EV	CL	90	RACE AND SEX	362		0
76Y	EY	CL	95	RACE AND SEX	377		10*
76Y	5G	CL	95	RACE AND SEX	297		10*
76Y	6Y	CL	95	RACE AND SEX	461		10*
91B	01	ST	95	RACE AND SEX	724		0
91C	02	ST	95	RACE AND SEX	220		0
91E	05	ST	95	SEX ONLY	154		5
92B	25	ST	95	SEX ONLY	121		44

* Number of FY81 accessions having SQT scores for the MOS disregarding availability of training criterion for the individual; thus same for different courses of an MOS.

Note. AA = Current aptitude area composite;
QS = Current qualifying score;
TCRS = Training course;
GRPT = Possible subgroup validation with training criterion;
GRPS = Possible subgroup validation with SQT criterion;
TOTTR = Number of useable training records;
TOTSQ = Number of SQT scores available.

schools within the same MOS. Although the threshold of 100 is small for multivariate analyses, there is a trade off between sampling error and being able to include a sufficient number of MOS to model validity generalization. At N = 100, the standard error of the correlation coefficient is approximately .10.

Table 4 also indicates the MOS for which we have adequate data to perform subgroup validations. We tentatively consider availability of criterion scores for 50 or more enlistees within group as sufficient to support subgroup analysis. The lower threshold (50 instead of 100 previously used to determine data sufficiency for validation with the entire MOS) was adopted because we plan to employ simultaneous estimation technique to conduct the subgroup analysis. This approach uses both the within and between subgroup information to estimate subgroup parameters and thus tends to provide more stable estimates. Therefore, smaller sample sizes for each subgroup may be tolerated. However, we plan to emphasize only those subgroup analyses that are based on at least 100 cases. Of the 67 MOS with an adequate data base for the training criterion, 32 have sufficient numbers of both Blacks and Whites, 16 have sufficient numbers of both men and women, and one (76C) has sufficient numbers of all four combinations to support separate validation analyses. Of the 33 MOS with an adequate data base for the SQT criterion, 20 have sufficient numbers of both Blacks and Whites, 8 have sufficient numbers of both men and women, and two (05C and 94B) have sufficient numbers of all four combinations to support separate validation analyses.

It seems clear that we will not have sufficient data to examine the interactions between race/ethnicity and sex effects as regards the validity of ASVAB subtests or composites. Also, due to sample size limitations, we cannot separately perform validation analyses for racial groups other than Blacks and Whites.

ASVAB Population Intercorrelations

Although criterion data are available only for recruits who actually enlist and are assigned to MOS, the selection and classification decisions must be applied to the entire population of applicants. To develop procedures when criterion data are missing for a large, nonrandom segment of the population, it is necessary to adjust for restriction in range. A key component of this adjustment is the population covariance/correlation matrix of ASVAB subtests.

The report on the 1980 youth profile (18-to 23-year-olds) provides one estimate for the potential applicant population. However, for purposes of this project, the ASVAB covariance/correlation matrix for FY81 nonprior service applicants has been estimated from a large sample (17,500) of applicants taken from the total population of approximately 500,000 FY81 applicants. Table 5 presents means and standard deviations of ASVAB subtest scores for this sample. As shown, the applicants to the Army, on the average, score a half standard deviation below the norm (score of 50, see the unweighted means). The unweighted variance-covariance estimates are provided in Table 6a.

Table 5

Summary Statistics for ASVAB Forms 8/9/10, Based on a 5% Sample of FY81 Nonprior-Service Applicants (Standard Scores are Used, N = 17,521)*

	ASVAB SUBTEST										
STAT	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
UNWEIGHTED ESTIMATES											
MEAN	44.1	46.2	44.1	45.4	47.1	48.8	44.8	46.5	44.7	45.0	44.2
STD	10.1	9.1	10.4	10.3	10.5	9.4	10.0	8.5	9.1	9.3	10.3
WEIGHTED ESTIMATES											
MEAN	48.9	51.3	49.3	50.2	50.9	51.8	48.2	50.9	48.4	48.6	49.5
STD	10.6	10.1	10.5	10.1	10.2	9.6	10.1	10.0	9.6	9.6	10.4

* Each subtest has a mean of 50 and standard deviation of 10 for the norming population.

Table 6a

Unweighted Estimates of Covariances and Correlations
Among the Subtests of ASVAB Forms 8/9/10--Standard Scores,
Based on a 5% Sample of FY81 Nonprior-Service Applicants*

(Above Diagonal = CORR, Below Diagonal = COV, Diagonal = VAR)

SUBTEST	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
GS	102.84	.68	.81	.71	.44	.37	.65	.61	.68	.70	.82
AR	62.88	83.57	.68	.65	.56	.45	.56	.75	.65	.61	.71
WK	85.25	65.08	108.19	.78	.50	.45	.60	.61	.63	.68	.98
PC	74.34	61.64	84.31	106.83	.52	.47	.55	.58	.59	.61	.89
NO	46.26	53.63	53.97	56.23	109.37	.65	.30	.53	.37	.36	.53
CS	35.12	39.10	43.78	45.67	64.42	89.12	.25	.44	.33	.30	.47
AS	65.93	51.16	62.93	56.76	31.77	23.54	100.16	.44	.71	.74	.62
MK	52.67	58.31	53.88	51.04	47.18	35.41	37.62	72.00	.58	.53	.63
MC	62.27	54.28	59.52	55.35	35.42	28.08	64.83	44.51	82.17	.70	.65
EI	65.60	51.37	65.15	58.22	34.93	26.56	68.43	41.82	58.92	85.90	.69
VE	85.52	66.88	105.25	95.28	57.06	46.31	63.72	55.33	60.80	65.78	106.82

* The 5% sample includes 17,521 nonprior-service applicants who took ASVAB form 8/9/10. Because standard scores are used, each subtest has a variance of 100 for the norming population.

Table 6b

Weighted Estimates of Covariances and Correlations Among the Subtests of ASVAB Forms 8/9/10--Standard Scores, Based on a 5% Sample of FY81 Nonprior-Service Applicants*

(Above Diagonal = CORR, Below Diagonal = COV, Diagonal = VAR)

SUBTEST	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
GS	111.51	.74	.83	.75	.51	.42	.67	.68	.72	.74	.84
AR	78.39	102.00	.74	.71	.62	.52	.59	.81	.69	.65	.76
WK	92.46	78.33	110.23	.82	.57	.50	.61	.67	.65	.70	.98
PC	80.28	73.03	87.52	102.33	.58	.51	.57	.64	.62	.64	.91
NO	54.33	64.10	60.72	60.07	103.36	.68	.34	.58	.42	.41	.60
CS	42.96	50.23	50.07	49.40	65.85	91.75	.27	.50	.36	.34	.52
AS	71.43	60.04	64.83	58.21	35.28	26.20	102.81	.48	.73	.76	.62
MK	72.20	81.53	69.86	64.72	59.20	47.98	48.38	99.83	.63	.59	.68
MC	72.43	66.83	65.79	60.49	40.94	33.33	71.26	60.51	91.58	.74	.67
EI	75.24	63.30	70.37	62.27	40.13	31.01	74.11	56.31	67.89	92.60	.71
VE	92.63	80.09	107.75	96.26	63.16	52.07	65.59	71.23	67.00	70.86	108.95

* The 5% sample includes 17,521 nonprior-service applicants who took ASVAB form 8/9/10. Because standard scores are used, each subtest has a variance of 100 for the norming population.

Table 6c

Estimated Correlation Matrix of ASVAB Tests (Form 8/9/10), Based on 1980 Youth Population, 18-to-23 Years Old

TEST	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
GS											
AR	72										
WK	80	71									
PC	69	67	80								
NO	52	63	60	60							
CS	45	51	55	56	70						
AS	64	53	53	42	30	22					
MK	69	83	67	64	62	52	41				
MC	70	69	60	52	40	34	74	60			
EI	76	66	68	57	41	34	75	59	74		
VE	80	73	98	90	62	57	52	70	60	67	

To compare the correlation matrix obtained for the FY81 Army applicants with that for the youth population, we weighted the sample to match the deciles of the AFQT. The weighted estimates are given in Table 6b, while the corresponding estimates from the 1980 youth population are shown in Table 6c. The weighted estimates are quite similar to the estimates from the youth population. The only difference between the two estimates that is larger than .10 is the correlation between PC (paragraph comprehension) and AS (auto shop information), where the youth population estimate is .42 and the weighted estimate is .57.

The unweighted estimates of correlations are consistently lower than the corresponding weighted values. The average difference, however, is small (.04). We plan to use the unweighted estimates for the FY81 nonprior service applicant population to adjust for range restriction, because the present validation is specifically aimed to facilitate the selection and classification of applicants to the Army.

Pass/Fail Rates

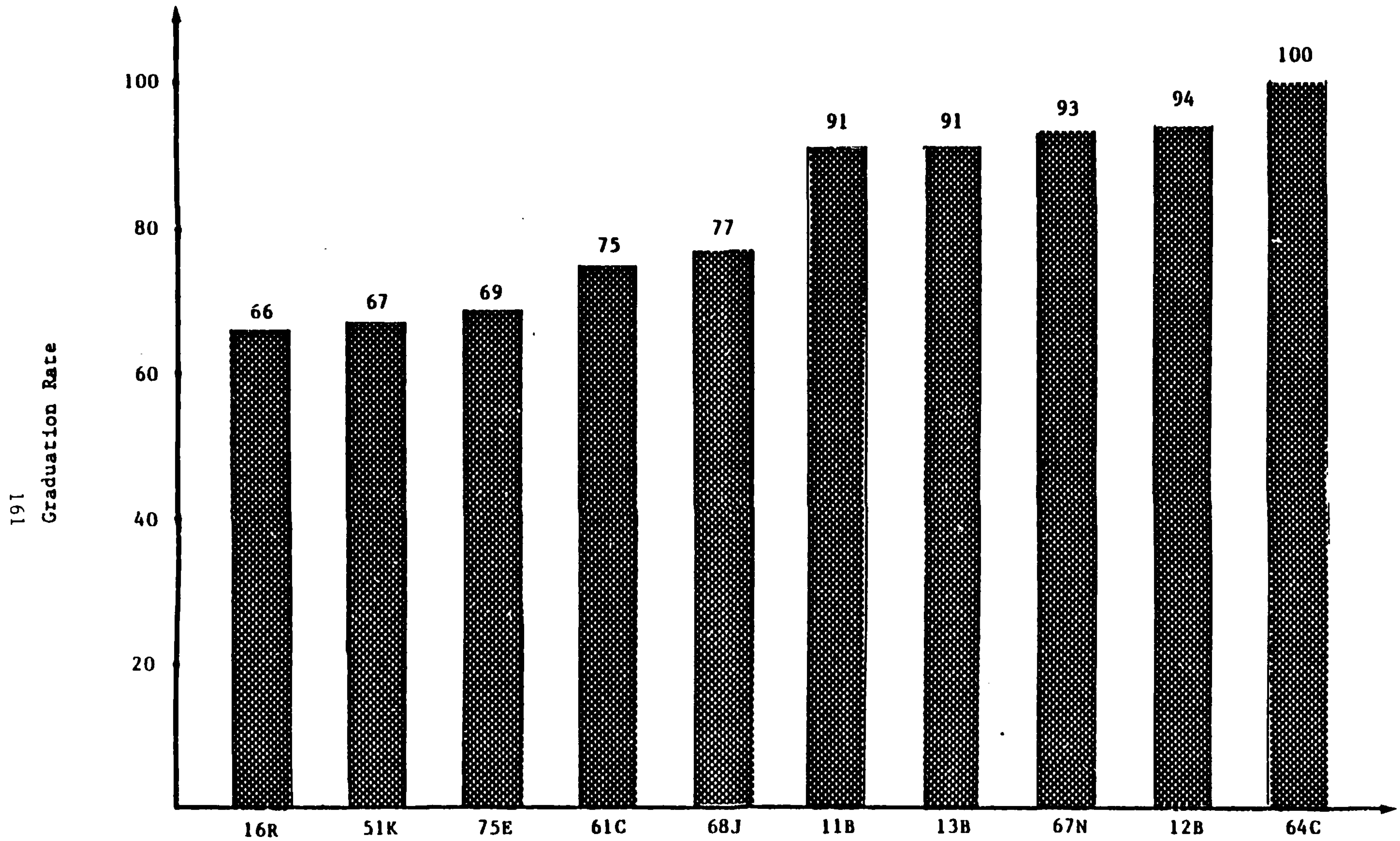
The major criteria for the present validation are training success scores, e.g., end-of-course grade, and SQT score. In addition to the training course grade, the data also include indicators of end-of-course disposition: whether the soldier graduated and if not, the cause for nongraduation. The graduation rates for training courses with at least 100 students are presented in Appendix C.

While the pass rates were generally quite high, there were significant differences in graduation rates among MOS/courses. For example, the graduation rates in Infantry (11B), Combat Engineer (12B), Cannon Crewman (13B), Motor Transport Operator (64C), and Utility Helicopter Repairer (67N) ranged from 91 percent to 100 percent; while for Short Range Gunnery Crewman (16R), Technical Engineering Supervisor (51K), Watercraft Engineer (61C), Aircraft Repairer (68J), and Personnel Actions Specialist (75E), the graduation rate ranged from 66 percent to 77 percent. Figure 7 highlights the differential graduation rates among the courses for these MOS.

For the SQT, the passing score is set at 60 percent. Table 7a shows the SQT passing rates for FY81 nonprior service enlistees in MOS that have at least 100 scores available. Table 7b presents similar data for prior service enlistees. The tables reveal that nearly all soldiers (96.7 percent) who take these tests achieve passing scores. The only MOS for which the pass rate is less than 90 percent are Terminal Operations Coordinator (57H), Finance Specialist (73C), Personnel Specialists (75B, 75C, 75D, 75E), and Petroleum Supply Specialist (76W).

Some Preliminary Validation Results

To serve as examples of validation analyses, we selected nine MOS representing each of the nine operational aptitude composites for preliminary analysis. These examples are based on samples of FY81 accessions who have both training and SQT scores available in the current data base. Because the current data base does not include the scores for a large number of soldiers who have taken an SQT this year, the sample sizes are smaller than



16R Short Range Gunner Crewman
51K Technical Engineer Supervisor
75E Personnel Actions Specialist
61C Watercraft Engineer
68J Aircraft Repairer

11B Infantry
13B Cannon Cunman
67N Utility Helicopter Repairer
12B Combat
64C Motor Transport Operator

Figure 7. Differential graduation rates among training courses for selected MOS (FY81 Accessions).

Table 7a

SQT Passing Rates for MOS With at Least 100 Cases
FY81 Nonprior-Service Accessions*

FAIL/PASS
FREQUENCY
ROW PERCENT

MOS	FAIL	PASS	TOTAL
05B	21 4.60	436 95.40	457
05C	55 6.19	834 93.81	889
11B	18 0.52	3412 99.48	3430
11C	8 0.98	809 99.02	817
11H	3 0.52	569 99.48	572
12B	12 1.18	1003 98.82	1015
12C	2 1.20	164 98.80	166
12F	0 0.00	140 100.00	140
13B	15 0.82	1815 99.18	1830
13E	7 2.50	273 97.50	280
13F	0 0.00	465 100.00	465
19D	2 0.38	525 99.62	527

* SQT percentage score of 60 or higher is passing.

Table 7a (cont'd)

SQT Passing Rates for MOS With at Least 100 Cases
FY81 Nonprior-Service Accessions*

FAIL/PASS
FREQUENCY
ROW PERCENT

MOS	FAIL	PASS	TOTAL
19E	5 0.54	921 99.46	926
31M	14 2.14	639 97.86	653
31V	9 2.97	294 97.03	303
36C	9 4.62	186 95.38	195
36K	10 1.38	716 98.62	726
52D	10 8.26	111 91.74	121
55B	9 4.07	212 95.93	221
57H	23 18.70	100 81.30	123
62B	3 1.89	156 98.11	159
63B	6 5.13	111 94.87	117
64C	48 3.42	1355 96.58	1403
67N	1 0.55	180 99.45	181

* SQT percentage score of 60 or higher is passing.

Table 7a (cont'd)

SQT Passing Rates for MOS With at Least 100 Cases
FY81 Nonprior-Service Accessions*

FAIL/PASS
FREQUENCY
ROW PERCENT

MOS	FAIL	PASS	TOTAL
73C	56 23.14	186 76.86	242
75B	108 27.14	290 72.86	398
75C	36 33.33	72 66.67	108
75D	87 25.74	251 74.26	338
75E	40 24.24	125 75.76	165
76W	20 14.08	122 85.92	142
82C	2 1.06	186 98.94	188
94B	13 1.11	1157 98.89	1170
95B	7 0.51	1363 99.49	1370
TOTAL	659 3.32	19178 96.68	19837 100.00

*SQT Percentage score of 60 or higher is passing.

Table 7b

SQT Passing Rates for MOS With at Least 100 Cases
FY81 Prior-Service Accessions*

FAIL/PASS
FREQUENCY
ROW PERCENT

MOS	FAIL	PASS	TOTAL
11B	7 1.24	557 98.76	564
11C	1 0.64	156 99.36	157
11H	1 0.81	122 99.19	123
12B	7 3.54	191 96.46	198
13B	6 2.19	268 97.81	274
19E	0 0.00	195 100.00	195
64C	15 5.64	251 94.36	266
94B	2 0.73	272 99.27	274
95B	4 2.27	172 97.73	176
TOTAL	43 1.93	2184 98.07	2227 100.00

*SQT percentage score of 60 or higher is passing.

will be used in actual validation. (The SQT data for FY81 accessions will be increased substantially when the update tapes that ARI has requested arrive.)

Table 8 presents the summary statistics for ASVAB scores by MOS/course. The sample correlations of the subtest scores and the uncorrected simple validities (with training and SQT score as separate criterion) are shown in Table 9.

The sample validity coefficients are adjusted for the restrictions of range employing the classical multivariate correction method (Lawley, 1943). As noted earlier, the unweighted estimates of covariances for FY81 nonprior service applicants (see Table 6a) are used as the population matrix in the correction. Tables 10a and 10b give the corrected subtest validities with training and SQT score as the criterion, respectively.

The uncorrected validities of the operational aptitude composites are also computed and presented in Tables 11a and 11b for the two criteria. The corrected validities for these composites are given in Tables 11c and 11d. In these tables, the validities for the composites that are used in the current selection and classification system (REQUEST) for each particular MOS are noted with asterisks. Note that the validity coefficients have not been adjusted for the unreliability of the criterion. When this adjustment is made, the validation will be proportionately higher.

Caution should be exercised in interpreting these provisional results. In particular, the sample sizes for the subgroups are small and their results

Table 8

Summary Statistics of ASVAB and Criterion Scores for FY81 Accessions in Nine MOS (Sample includes only soldiers with both training and SQT scores)

MOS	TSCHL	TCRS	STAT	GS	AR	WK	PC	VE	NO	CS
05C	113	2D	MEAN	51.1	53.3	53.2	54.0	53.4	55.7	55.8
05C	113	2D	STD	7.9	8.1	6.1	5.8	6.1	6.1	6.8
11B	809	IN	MEAN	50.1	51.3	50.4	51.6	50.8	50.6	51.5
11B	809	IN	STD	8.3	7.6	7.6	7.4	7.2	7.6	6.4
13B	810	3B	MEAN	44.2	48.8	44.6	46.7	44.9	50.3	50.8
13B	810	3B	STD	9.4	7.3	8.8	8.5	8.6	6.8	5.9
31M	113	4D	MEAN	51.6	53.4	51.2	52.5	51.7	50.5	51.3
31M	113	4D	STD	6.2	6.4	6.8	6.4	6.4	8.4	7.6
55B	093	5B	MEAN	47.1	43.0	46.9	47.2	46.6	43.8	46.2
55B	093	5B	STD	6.2	6.9	6.4	7.2	6.3	7.0	7.2
62B	807	CB	MEAN	47.6	48.4	46.9	47.9	46.6	49.2	49.6
62B	807	CB	STD	8.1	7.9	8.0	8.2	8.1	7.6	7.5
75B	121	5E	MEAN	45.7	49.3	48.2	49.3	48.5	56.3	56.4
75B	121	5E	STD	8.7	8.5	8.2	8.3	8.0	5.1	6.0
94B	101	KA	MEAN	47.0	48.7	48.7	50.4	49.0	51.1	49.6
94B	101	KA	STD	7.4	7.3	6.6	6.4	6.2	7.3	8.3
95B	813	SB	MEAN	54.9	55.1	55.5	55.4	55.7	53.5	53.7
95B	813	SB	STD	5.7	7.3	4.9	5.1	4.4	7.4	8.3

MOS	TSCHL	TCRS	STAT	AS	MK	MC	EI	TSCR	SQT	N
05C	113	2D	MEAN	51.4	52.1	50.0	50.5	88.9	76.2	343
05C	113	2D	STD	8.3	8.3	8.1	7.5	6.0	9.8	
11B	809	IN	MEAN	51.9	49.5	50.9	50.5	94.8	87.0	575
11B	809	IN	STD	7.5	8.1	7.2	7.5	5.0	7.6	
13B	810	3B	MEAN	43.7	48.3	46.3	45.2	78.5	86.5	374
13B	810	3B	STD	10.1	6.7	7.4	8.6	18.5	9.1	
31M	113	4D	MEAN	48.2	51.8	48.8	51.3	93.0	86.6	272
31M	113	4D	STD	8.7	6.4	6.8	6.7	4.9	9.9	
55B	093	5B	MEAN	45.1	45.3	41.5	47.0	85.8	79.1	100
55B	093	5B	STD	6.0	5.6	6.1	5.0	4.6	10.0	
62B	807	CB	MEAN	53.8	47.1	51.4	50.9	91.5	79.8	121
62B	807	CB	STD	7.3	7.1	7.8	7.4	6.7	9.4	
75B	121	5E	MEAN	44.5	49.6	44.7	46.3	86.5	66.9	263
75B	121	5E	STD	8.7	7.6	8.0	8.0	11.6	17.8	
94B	101	KA	MEAN	48.1	47.2	48.4	46.9	86.6	87.5	320
94B	101	KA	STD	7.5	6.8	7.0	7.7	1.6	9.0	
95B	913	SB	MEAN	54.0	53.5	53.6	52.8	81.3	85.9	449
95B	813	SB	STD	7.4	7.5	6.4	6.9	6.6	9.2	

Note. TSCHL = Training school code; TCRS = Training course code; ASVAB subtest scores are standardized so that the mean for the norm population is 50 and standard deviation is 10.

Table 9

Sample (Uncorrected) Validity Coefficients of ASVAB Tests by MOS and Course

GROUP	COURSE	N	CRIT	GS	AR	VE	NO	CS	AS	MK	MC	EI	TSCR	SQT
05C:	113: 2D	243	TSCR	.18	.26	.15	-.01	.02	.14	.22	.23	.20		.10
05C:	113: 2D		SQT	.26	.29	.21	-.06	-.05	.25	.26	.26	.28	.10	
05C: W	113: 2D	239	TSCR	.13	.23	.15	.03	.13	.10	.19	.19	.16		.13
05C: W	113: 2D		SQT	.17	.25	.22	-.01	.01	.19	.24	.23	.21	.13	
05C: B	113: 2D	86	TSCR	.13	.16	.09	-.00	-.11	.07	.18	.18	.09		-.07
05C: B	113: 2D		SQT	.13	.13	.06	.01	-.08	.06	.16	.05	.18	-.07	
05C: F	113: 2D	58	TSCR	.13	.25	.18	.18	-.03	.24	.15	.18	.30		.10
05C: F	113: 2D		SQT	.01	.24	.10	.05	.13	.03	.26	.02	.06	.10	
05C: M	113: 2D	285	TSCR	.19	.26	.14	-.04	.04	.13	.23	.24	.18		.10
05C: M	113: 2D		SQT	.29	.27	.23	-.04	-.06	.24	.24	.27	.30	.10	
11B:	809: IN	575	TSCR	.22	.24	.19	.12	.10	.18	.27	.27	.15		.16
11B:	809: IN		SQT	.25	.27	.27	.11	.07	.23	.30	.25	.26	.16	
13B:	810: 3B	374	TSCR	.10	.12	.13	.07	.01	.15	.04	.15	.18		.11
13B:	810: 3B		SQT	.29	.25	.29	.12	.07	.32	.19	.29	.22	.11	
31M:	113: 4D	272	TSCR	.12	.26	.16	.15	.17	.01	.25	.12	.08		.23
31M:	113: 4D		SQT	.19	.14	.09	-.03	-.04	.07	.15	.11	.16	.23	
31M: W	113: 4D	171	TSCR	.22	.29	.18	.19	.17	-.04	.29	.11	.20		.27
31M: W	113: 4D		SQT	.22	.16	-.00	-.08	-.13	.01	.17	.08	.11	.27	
31M: B	113: 4D	87	TSCR	-.20	.13	-.00	.09	.13	-.19	.22	-.01	-.18		.20
31M: B	113: 4D		SQT	.14	.13	.27	.09	.15	.13	.17	.13	.20	.20	
31M: F	113: 4D	76	TSCR	.14	.35	.06	.13	.25	.03	.16	.05	.12		.14
31M: F	113: 4D		SQT	.14	.21	-.10	-.16	-.05	.18	.01	.10	.05	.14	
31M: M	113: 4D	196	TSCR	.14	.25	.16	.12	.09	.07	.26	.22	.14		.25
31M: M	113: 4D		SQT	.22	.12	.14	-.01	-.06	.07	.19	.13	.22	.25	
55B:	093: 5B	100	TSCR	.22	.28	.21	.04	.11	.20	.19	.07	.15		.29
55B:	093: 5B		SQT	.17	.19	.12	.17	.13	.06	.35	.07	-.00	.29	
62B:	807: CB	121	TSCR	.32	.29	.27	-.03	.27	.42	.29	.32	.21		.22
62B:	807: CB		SQT	.36	.30	.33	-.04	.13	.39	.31	.39	.19	.22	
75B:	121: 5E	263	TSCR	.17	.34	.17	.04	.01	.23	.27	.17	.15		.34
75B:	121: 5E		SQT	.35	.45	.37	.06	.05	.27	.38	.34	.26	.34	
94B:	101: KA	320	TSCR	.13	.18	.15	.02	.11	.19	.18	.07	.18		.11
94B:	101: KA		SQT	.18	.19	.17	-.16	.02	.18	.12	.19	.17	.11	
94B: W	101: KA	204	TSCR	.09	.13	.14	.13	.13	.11	.13	-.03	.12		.04
94B: W	101: KA		SQT	.17	.12	.18	-.20	-.02	.15	.11	.15	.14	.04	
94B: B	101: KA	109	TSCR	.05	.11	.05	-.16	.01	.19	.20	.13	.19		.26
94B: B	101: KA		SQT	.01	.22	.04	-.04	.02	.08	-.02	.13	.10	.26	
94B: F	101: KA	60	TSCR	.23	.21	.29	-.14	.19	.37	.30	.33	.19		.29
94B: F	101: KA		SQT	.18	.04	.20	-.12	-.06	.04	.06	.08	.07	.29	
94B: M	101: KA	260	TSCR	.13	.18	.08	.02	.02	.28	.17	.10	.27		.09
94B: M	101: KA		SQT	.18	.21	.17	-.17	.04	.22	.13	.22	.19	.09	
95B:	813: SB	449	TSCR	.30	.22	.30	.07	.10	.21	.16	.21	.32		.14
95B:	813: SB		SQT	.17	.24	.15	.15	.11	.22	.23	.24	.23	.14	

Note. GROUP = MOS name followed by subgroup identification (W for White, B for Black; and F for Female, M for Male).
TSCR = Training course grade. SQT = SQT percentage score.
The last two columns provide the intercorrelations between the training and SQT criterion scores.

Table 10a

Corrected Validity Coefficients of ASVAB Tests (Form 8/9/10) by MOS Groups, Criterion is Training Score (Corrections based on unweighted covariance matrix for FY81 nonprior-service applicants)

MOS_RS	N	GS	AR	VE	NO	CS	AS	MK	MC	EI
05C:	343	.26	.31	.26	.19	.20	.23	.27	.29	.26
05C: W	239	.29	.35	.33	.25	.31	.27	.30	.32	.30
05C: B	86	.15	.16	.10	.05	-.00	.12	.18	.19	.13
05C: F	58	.51	.55	.56	.64	.42	.59	.49	.53	.60
05C: M	285	.21	.25	.18	.10	.15	.17	.23	.25	.19
11B:	575	.32	.35	.29	.24	.23	.29	.34	.36	.25
13B:	374	.12	.14	.15	.14	.06	.19	.08	.19	.22
31M:	272	.46	.51	.48	.37	.35	.31	.48	.40	.41
31M: W	171	.55	.57	.54	.44	.39	.35	.53	.42	.54
31M: B	87	-.15	-.03	-.09	-.01	.06	-.22	.06	-.08	-.16
31M: F	76	.45	.57	.38	.38	.35	.30	.47	.31	.39
31M: M	196	.46	.51	.47	.32	.26	.36	.47	.46	.44
55B:	100	.61	.57	.57	.27	.24	.54	.54	.49	.60
62B:	121	.43	.40	.41	.27	.36	.51	.38	.43	.40
75B:	263	.23	.33	.22	.13	.13	.28	.28	.23	.24
94B:	320	.30	.34	.35	.24	.26	.34	.32	.27	.34
94B: W	204	.32	.35	.39	.34	.33	.33	.31	.24	.33
94B: B	109	.34	.38	.36	.17	.23	.41	.41	.39	.41
94B: F	60	.45	.50	.56	.21	.28	.60	.52	.59	.56
94B: M	260	.27	.30	.25	.18	.14	.38	.26	.27	.37
95B:	449	.59	.52	.61	.33	.32	.48	.45	.49	.57

Note. MOS_RS = MOS name followed by subgrou_ identification (W for White, B for Black; and F for female, M for male).

Table 10b

Corrected Validity Coefficients of ASVAB Tests (Form 8/9/10) by MOS Groups, Criterion is SQT Score (Corrections based on unweighted covariance matrix for FY81 nonprior-service applicants)

MOS_RS	N	GS	AR	VE	NO	CS	AS	MK	MC	EI
05C:	343	.36	.37	.36	.23	.19	.36	.34	.34	.38
05C: W	239	.34	.38	.40	.26	.22	.37	.35	.35	.37
05C: B	86	.20	.17	.15	.11	.05	.16	.21	.13	.26
05C: F	58	.06	.24	.16	.04	.20	.08	.27	.07	.12
05C: M	285	.42	.39	.42	.25	.19	.40	.35	.39	.43
11B:	575	.36	.38	.38	.26	.21	.34	.37	.36	.36
13B:	374	.34	.33	.35	.25	.19	.35	.29	.34	.27
31M:	272	.42	.39	.36	.19	.13	.30	.38	.32	.38
31M: W	171	.31	.27	.19	.05	-.01	.17	.26	.18	.23
31M: B	87	.58	.58	.62	.42	.37	.47	.55	.52	.58
31M: F	76	.12	.20	-.10	-.15	-.16	.23	.03	.12	.10
31M: M	196	.45	.41	.40	.18	.10	.35	.40	.39	.46
55B:	100	.50	.49	.47	.32	.28	.40	.59	.42	.44
62B:	121	.48	.44	.48	.29	.26	.49	.42	.51	.40
75B:	263	.54	.61	.60	.48	.45	.39	.55	.49	.45
94B:	320	.24	.25	.25	.01	.08	.26	.18	.25	.24
94B: W	204	.19	.17	.17	-.10	-.00	.18	.15	.17	.17
94B: B	109	.44	.52	.51	.37	.31	.48	.34	.51	.48
94B: F	60	.33	.21	.33	.09	.06	.21	.22	.25	.21
94B: M	260	.24	.26	.23	.00	.08	.28	.17	.27	.25
95B:	449	.40	.43	.41	.33	.27	.42	.41	.44	.43

Note. MOS_RS = MOS name followed by subgroup identification (W for White, B for Black; and F for female, M for male).

Table 11a

Sample Correlations Between Current Aptitude Composites and Training Score

MOS_RS	N	AFQT	GM	EL	CL	MM	SC	CO	FA	OF	ST
05C:	343	.23	.24	.26	.08	.23	.17*	.26	.27	.23	.25
05C: W	239	.22	.19	.22	.15	.19	.21*	.25	.26	.20	.21
05C: B	86	.15	.18	.20	-.02	.14	.03*	.15	.18	.16	.20
05C: F	58	.28	.28	.28	.19	.37	.29*	.28	.23	.32	.21
05C: M	285	.22	.24	.27	.07	.22	.15*	.27	.28	.21	.25
11B:	575	.25	.26	.28	.20	.26	.23	.30*	.31	.27	.29
13B:	374	.14	.15	.14	.11	.19	.14	.15	.12*	.17	.13
31M:	272	.29	.16	.29*	.22	.14	.19	.21	.30	.18	.24
31M: W	171	.33	.23	.38*	.24	.19	.20	.20	.32	.19	.29
31M: B	87	.11	-.16	-.02*	.11	-.11	.02	.02	.18	-.04	.01
31M: F	76	.27	.19	.32*	.22	.15	.20	.26	.30	.11	.15
31M: M	196	.27	.22	.32*	.18	.20	.17	.23	.30	.22	.29
55B:	100	.28	.35*	.35	.17	.19	.22	.26	.25	.22	.26
62B:	121	.31	.40	.36	.26	.38*	.38	.46	.40	.40	.38
75B:	263	.31	.27	.30	.17*	.22	.24	.30	.30	.23	.25
94B:	320	.20	.23	.23	.14	.20	.23	.22	.20	.21*	.18
94B: W	204	.20	.15	.16	.20	.14	.24	.14	.13	.16*	.11
94B: B	109	.05	.24	.22	-.05	.19	.05	.22	.17	.16*	.18
94B: F	60	.25	.40	.33	.18	.32	.31	.40	.37	.40*	.43
94B: M	260	.16	.28	.25	.06	.28	.19	.22	.16	.22*	.16
95B:	449	.26	.35	.34	.17	.30	.24	.28	.24	.30	.33*

Note. MOS_RS = MOS name followed by subgroup identification (W for White, B for Black; and F for female, M for male).

* Figures are correlations between the required aptitude composite and the criterion (training score).

Table 11b

Sample Correlations Between Current Aptitude Composites and SQT Score

MOS_RS	N	AFQT	GM	EL	CL	MM	SC	CO	FA	OF	ST
05C:	343	.28	.34	.33	.05	.30	.22*	.31	.29	.30	.31
05C: W	239	.26	.27	.27	.09	.25	.20*	.27	.26	.27	.26
05C: B	86	.11	.20	.20	-.01	.13	.03*	.08	.12	.09	.15
05C: F	58	.21	.12	.19	.18	.06	.17*	.19	.27	.08	.12
05C: M	285	.27	.35	.33	.06	.31	.20*	.29	.27	.31	.32
11B:	575	.31	.33	.33	.22	.30	.27	.31*	.31	.31	.32
13B:	374	.30	.32	.29	.25	.32	.32	.33	.30*	.34	.32
31M:	272	.12	.21	.26*	.00	.12	.04	.10	.13	.10	.20
31M: W	171	.06	.18	.25*	-.10	.04	-.09	.04	.10	-.00	.18
31M: B	87	.28	.27	.32*	.24	.23	.27	.22	.22	.27	.30
31M: F	76	.01	.16	.18*	-.14	.06	-.06	.14	.08	-.00	.06
31M: M	196	.15	.25	.29*	.03	.14	.06	.10	.14	.12	.25
55B:	100	.22	.27*	.30	.20	.15	.20	.19	.28	.19	.26
62B:	121	.35	.40	.37	.22	.38*	.34	.43	.39	.44	.44
75B:	263	.49	.40	.47	.36*	.35	.38	.44	.46	.40	.46
94B:	320	.16	.22	.22	.00	.16	.09	.23	.18	.18*	.22
94B: W	204	.10	.19	.18	-.04	.09	.03	.16	.13	.11*	.21
94B: B	109	.16	.08	.13	.01	.15	.06	.22	.14	.17*	.07
94B: F	60	.10	.12	.12	-.01	.02	.01	.03	.03	.08*	.19
94B: M	260	.17	.24	.23	.01	.19	.11	.26	.21	.20*	.23
95B:	449	.25	.31	.31	.17	.32	.25	.32	.29	.31	.29*

Note. MOS_RS = MOS name followed by subgroup identification (W for White, B for Black; and F for female, M for male).

* Figures are correlations between the required aptitude composite and the criterion (SQT score).

Table 11c

Corrected Validity Coefficients of Current Aptitude Composites: Criterion is Training Score (Corrections Based on Unweighted Covariance Matrix for FY81 Nonprior-Service Applicants)

MOS_RS	N	AFQT	GM	EL	CL	MM	SC	CO	FA	OF	ST
05C:	343	.31	.30	.32	.26	.30	.28*	.33	.33	.30	.31
05C: W	239	.37	.34	.36	.35	.36	.37*	.40	.39	.36	.36
05C: B	86	.13	.17	.18	.06	.15	.09*	.15	.16	.14	.18
05C: F	58	.66	.65	.63	.65	.73	.71*	.66	.62	.72	.60
05C: M	285	.22	.24	.26	.17	.22	.20*	.26	.27	.22	.25
11B:	575	.34	.35	.36	.30	.35	.33	.39*	.39	.36	.37
13B:	374	.16	.18	.17	.14	.23	.17	.18	.15*	.20	.16
31M:	272	.54	.49	.54*	.48	.47	.49	.50	.54	.49	.53
31M: W	171	.61	.58	.64*	.55	.54	.55	.54	.59	.55	.59
31M: B	87	-.06	-.15	-.09*	-.02	-.15	-.09	-.09	.00	-.12	-.08
31M: F	76	.52	.47	.55*	.44	.43	.45	.48	.53	.43	.46
31M: M	196	.52	.51	.55*	.42	.49	.45	.50	.52	.50	.54
55B:	100	.58	.68*	.68	.43	.59	.52	.59	.57	.58	.64
62B:	121	.43	.51	.47	.41	.50*	.50	.54	.49	.50	.48
75B:	263	.28	.30	.32	.19*	.27	.25	.31	.30	.27	.28
94B:	320	.37	.38	.38	.34	.37	.38	.38	.37	.37*	.36
94B: W	204	.42	.38	.38	.42	.39	.45	.40	.38	.41*	.36
94B: B	109	.38	.46	.45	.30	.43	.38	.45	.43	.41*	.43
94B: F	60	.53	.63	.59	.42	.60	.53	.63	.58	.60*	.61
94B: M	260	.29	.38	.35	.23	.37	.31	.35	.30	.34*	.30
95B:	449	.59	.62	.62	.50	.58	.56	.57	.55	.59	.62*

Note. MOS_RS = MOS name followed by subgroup identification (W for White, B for Black; and F for female, M for male).

* Figures are correlations between the required aptitude composite and the criterion (training score).

Table 11d

Corrected Validity Coefficients of Current Aptitude Composites: Criterion is SQT Score (Corrections Based on Unweighted Covariance Matrix for FY81 Nonprior-Service Applicants)

MOS_RS	N	AFQT	GM	EL	CL	MM	SC	CO	FA	OF	ST
05C:	343	.39	.43	.42	.31	.40	.37*	.40	.38	.40	.40
05C: W	239	.42	.43	.42	.35	.42	.40*	.42	.41	.43	.42
05C: B	86	.17	.25	.24	.12	.20	.15*	.16	.17	.17	.20
05C: F	58	.19	.15	.19	.15	.09	.15*	.19	.24	.11	.16
05C: M	285	.43	.48	.47	.34	.46	.41*	.44	.41	.45	.46
11B:	575	.41	.42	.43	.34	.41	.38	.41*	.41	.42	.42
13B:	374	.37	.37	.36	.32	.38	.37	.39	.36*	.40	.38
31M:	272	.38	.43	.46*	.27	.36	.32	.36	.37	.36	.43
31M: W	171	.22	.29	.31*	.09	.19	.13	.19	.21	.18	.27
31M: B	87	.64	.65	.67*	.56	.62	.61	.61	.62	.63	.66
31M: F	76	-.00	.15	.13*	-.16	.09	-.06	.12	.06	.02	.05
31M: M	196	.41	.49	.51*	.28	.42	.34	.40	.40	.41	.48
55B:	100	.51	.56*	.58	.43	.49	.47	.50	.54	.50	.57
62B:	121	.49	.53	.51	.41	.52*	.49	.54	.50	.54	.55
75B:	263	.66	.57	.62	.61*	.57	.62	.61	.65	.61	.63
94B:	320	.23	.27	.27	.14	.23	.19	.27	.24	.24*	.27
94B: W	204	.13	.21	.20	.03	.13	.08	.17	.15	.13*	.20
94B: B	109	.55	.52	.52	.48	.57	.54	.58	.52	.58*	.52
94B: F	60	.27	.29	.28	.19	.23	.22	.23	.22	.27*	.33
94B: M	260	.22	.28	.27	.13	.25	.19	.28	.24	.24*	.26
95B:	449	.46	.49	.49	.41	.51	.46	.50	.48	.50	.48*

<u>Note</u>: MOS_RS = MOS name followed by subgroup identification (W for White, B for Black; and F for female, M for male).

* Figures are correlations between the required aptitude composite and the criterion (SQT Score).

may be quite unreliable. It should also be noted that the variances of some of the training scores, e.g., 31M, are extremely small.

Next Steps

During the next contract period a number of very important analyses will be carried out using the 81/82 data file. The most important of these follow:

(1) Validation of ASVAB subtests and current composites.

(2) Evaluation of alternative methods for adjusting selectivity bias, for example, the design of paradigms that will allow the application of Heckman's procedure in the estimation of regressions from biased samples.

(3) Determination of appropriate number of composites and formation of homogeneous MOS groups for the development of new composites.

(4) Evaluation of the discriminant validity of the new composites employing optimal assignment algorithms.

(5) Investigation of moderator effects and differential validities among subgroups of enlisted personnel.

(6) Cross-validation of the newly developed composites.

(7) Modeling the generalization of validation results to the population of Army jobs.

Each of these activities will produce a technical report. A final report on the recommendation for new composites and qualification scores for ASVAB forms 11/12/13 will also be prepared.

Associated Reports

In the past year a number of reports and papers have been produced that deal directly with Task 1 activities. These papers fall into three broad categories. First research was done investigating the factor structure of ASVAB the Army's current selection and classification instrument. The second group of papers reported preliminary investigations into the validity of ASVAB as a predictor of training. Finally, research was also done examining the validity of the Military Applicant Profile (MAP), a supplemental selection device for applicants who do not have a high school degree, and reading assessment.

The investigation of the factorial structure of ASVAB was conducted by Hanser and Mitchell. This paper is relevant to the goals of Project A for two reasons. First a clearer understanding of the content or structure of ASVAB should help illustrate areas of the possible predictor/criterion space where improvements are possible. Second, Hanser and Mitchell's careful analysis of the various subgroups within the Army's applicant population should suggest whether or not a single selection test is appropriate for all of these subgroups. The work done by Hanser and Mitchell is also noteworthy in that it used the new LISREL V program developed by Joreskog, rather than traditional factor analytic techniques.

Two research efforts over the past year investigated the validity of ASVAB as a predictor of training success. The first of these, conducted by Weltin and Popelka, was a detailed investigation of the validity of the CL or clerical composite. The second effort was undertaken by Rossmeissl, Martin, Wing and Wang, and analyzed a larger number of MOS and composites, but with less detail. Both of these investigations found high correlations between ASVAB composites and training scores. This work will provide an important baseline or background to the large project A undertaking of determining the best set of AA composites for use with ASVAB 11/12/13.

A next piece of research relevant to Task 1 did not look at ASVAB scores at all. This research was conducted by Eaton, Weltin, and Wing, and investigated the validity of the Military Applicant Profile or MAP. MAP had been used since 1979 to screen 17 year old men applicants who did not have a high school degree. On the basis of the work done by Eaton et al, the use of MAP has been extended by the Army to all non-high-school graduates. This work shows the importance of nontraditional measures for selection and classification and illustrates some of the techniques that can be used in the validation of those measures.

The last paper, by Oxford-Carpenter and Schultz, reviews the properties of available reading tests. Reading is an important issue in testing, as a mediator of the validity of many tests, as an educational criterion, and as a potential criterion for reenlistment and other Army selection and classification needs. The paper describes these issues in the context of a cognitive model of reading.

Factorial Invariance of the Armed Services Vocational Aptitude Battery*

Lawrence M. Hanser and Karen J. Mitchell

U.S. Army Research Institute
for the Behavioral and Social Sciences

Introduction

The Armed Services Vocational Aptitude Battery (ASVAB) serves as the primary cognitive test battery for screening applicants for entry into military service. A complete description of the development, content, and psychometric properties of ASVAB can be found elsewhere (c.f., Maier and Grafton, 1981). A number of composites are derived from the ten subtests which comprise ASVAB. For example, the Armed Forces Qualification Test (AFQT) score, which is used by all the Services, provides the initial determination of whether an individual meets minimum mental ability standards to enter military service. Individuals scoring below the 16th percentile are currently ineligible for entering any of the Services. Each Service also uses a number of other ASVAB composites for determining eligibility for entry into specific training programs. For example, individuals desiring to become military police in the Army must, among other requirements, score at or above a certain percentile on the Skilled Technical composite. These composites, with the exception of AFQT, are specific to each Service.

Previous research (Kass, Mitchell, Grafton, & Wing, 1982; Ree, Mullins, Mathews, & Massey, 1981) on the factor structure of ASVAB has demonstrated that the constructs measured by ASVAB can be reasonably described by four factors. With minor variation, these are: Verbal (General Science, Word Knowledge, and Paragraph Comprehension), Speed (Coding Speed and Numerical Operations), Math (Arithmetic Reasoning and Mathematical Knowledge) and Technical (Auto/Shop Information, Mechanical Comprehension, and Electronics Information). While the research by Ree et al. did not address the issue of population subgroup differences in factor structure, Kass et al. (1982) examined subgroup differences, but collapsed across race on one hand and across sex on the other. They did not examine the factor structure within each separate race/sex subgroup.

There is some evidence which indicates that there may be important differences between individual race/sex subgroups. For example, white females exhibit the highest rate of attrition from the Army, yet black females exhibit the lowest rate. While this specific kind of difference may not be reflected in constructs measured by ASVAB, data should be analyzed in such a way as not to obscure subgroup differences where they may exist. The purpose of this research was to examine the factorial invariance of the ASVAB for four race/sex subgroups: 1) white males, 2) black males, 3) white females, and 4) black females. The three hypotheses which were examined were: 1) equality of subgroup correlation matrices, 2) invariance of the number of factors across subgroups, and 3) invariance of factor loadings across subgroups.

---

*Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983.

Method

Subjects

Subjects were all Army applicants in FY 1981 who had been tested with ASVAB Form 8a. This sample consisted of 12,635 black males, 4,429 black females, 6,272 white females, and 28,275 white males.

Analyses

A matrix of the intercorrelations of the 10 ASVAB subtests was calculated for each of the above groups. These four correlation matrices served as input to the remainder of the analyses. Each of the subgroup matrices was factor analyzed using an unrestricted maximum likelihood technique with an oblique rotation. Two, three, and four factor solutions were obtained for each group. The purpose of these initial factor analyses was to attempt to replicate solutions found by other investigators and to provide a basis for the restricted maximum likelihood analyses which were to follow.

The three hypotheses noted above were tested using restricted maximum likelihood analysis (Joreskog & Sorbom, 1983). The first of these analyses simply tested the equality of the subgroup correlation matrices. The second analysis tested the hypothesis that the number of factors was invariant across groups. A factor pattern was specified to be invariant, but the factor loadings were allowed to vary. The third analysis tested whether the factor loadings themselves were invariant across groups. For all analyses the factors were allowed to correlate, thus allowing for oblique solutions.

Results

The results of the initial four factor unrestricted maximum likelihood solutions are provided in Table 1. Only minor differences exist between these and the four factor solutions reported by Kass et al. (1982) and Ree et al. (1981). These four factors can be interpreted as Verbal, Speed, Math, and Technical, as noted in the introduction. It is clear that there were no substantial subgroup differences evident in these solutions. It was not certain, however, that four factors were necessary to describe the data. The initial eigenvalues dropped rapidly and became negative after the fourth factor, and, on examination, indicated that perhaps two factors would have been satisfactory.

For the above reasons, we also examined the two and three factor solutions for each subgroup. On the basis of eigenvalues and interpretability, we judged that a two factor solution provided a more parsimonious explanation for the data. The two factor solutions are reported in Table 2. These two factors are a General factor and a Speed factor. In the two factor solution, however, we found a slight difference between the solutions for the minority groups and the solution for white males. For all three minority groups these factors were relatively clean, but for white males the Mathematics Knowledge and Arithmetic Reasoning subtests had substantial loadings on the Speed factor. It was decided to examine these differences more closely in the restricted models which followed.

When the four correlation matrices were tested for equality using the

LISREL VI program (Joreskog and Sorbom, 1983), the chi-square was 408.2 with 165 degrees of freedom. Clearly, the matrices were unequal. This was not surprising, given that this was the strictest possible test of subgroup equality. This result did not preclude the more relaxed model of a common factor structure across subgroups. Rather, had this hypothesis not been rejected, there would have been no need to continue with the remaining hypotheses.

The second hypothesis was that the number and pattern of the factors were equivalent across subgroups. When the four factor solution identified previously was tested, the program was unable to compute a maximum likelihood solution. Upon inspection, it was determined that there were too many constraints on the model. That is, the four factor solution could not be fit to the data when only two or three variables were allowed to load on each factor. When the four factor unrestricted solution was examined to guide the choice of other variables which might have been allowed to load on the four factor restricted model, the meaning of the factors became clouded. On the basis of these results, the original four factor model was rejected.

We continued to test the hypothesis that the number and pattern of factors were equal across subgroups with a two factor model which consisted of a General factor (all subtests except Numerical Operations and Coding Speed) and a Speed factor (Numerical Operations and Coding Speed). This model was chosen for its straightforward interpretability and the fact that it had emerged for three of four groups in the unrestricted solutions. This two factor solution had also been previously found by Kass et al. (1982). This time, maximum likelihood estimates could be computed. The statistics provided by LISREL VI which were used to judge the fit of the model indicated that if the Arithmetic Reasoning and Mathematics Knowledge subtests were allowed to load on the factor with the speeded subtests, the fit would improve. The earlier unrestricted two factor solutions had also suggested this arrangement, more strongly for white males, but for the other subgroups as well. That these subtests would load on the Speed factor also seemed theoretically reasonable. Whereas the necessary adjustments to the four factor model had yielded a logically unclear set of factors, the modified two factor solution remained interpretable. As expected, when these subtests were allowed to load on the Speed factor, the fit of the model improved significantly for all subgroups.

An examination of the statistics accompanying the revised two factor model indicated that the fit of the model for white males would improve significantly if Paragraph Comprehension were also allowed to load on the Speed factor. While this model would have fit all subgroups slightly better, it was judged that loading Paragraph Comprehension on the Speed factor did not make theoretical sense. In the interests of completeness, this model was computed to determine its impact on model fit. The improvement in fit was deemed insufficient given the overall impact on interpretability of the factors. The unrestricted and restricted two factor solutions converged, and the revised two factor solution was chosen as best representative of the data.

The test which remained was the test of the invariance of the factor structure across the subgroups. Whereas previously only the pattern of loadings had been constrained, it remained to be seen whether the structures themselves were invariant. A model which constrained the loadings to be equal across subgroups was fit. The difference between the chi-squares for these two models (pattern invariance vs. loading invariance) was computed to be 63.83 with 30 degrees of freedom. This significant difference meant that while the pattern may have been acceptable across groups, the structure was not invariant. That is, while the same variables loaded on the same factors, the loadings differed across subgroups.

Upon examining the results of this model, it appeared that the variant group may have been white males. The third hypothesis was tested again using only three subgroups, excluding white males, and the results were quite different. The difference between the two chi-squares calculated on the three remaining minority subgroups was 23.84 with 20 degrees of freedom. This difference was not significant, and indicated that not only could the the pattern of loadings be considered invariant across subgroups, but the loadings themselves were essentially invariant across subgroups. Thus the final interpretation was that the pattern of loadings was appropriate for all subgroups, but there were slight differences in the loadings themselves between the white male subgroup and the other three subgroups. These differences revolve around the loadings of the math related subtests and are evidenc in Table 2.

Discussion

Previous researchers have concluded that ASVAB is best explained by a four factor solution and that the four factor solution holds across race and sex subgroups. Our results indicated that a four factor solution could be obtained from the data, but that a two factor solution was probably more appropriate, and certainly more parsimonious. With regard to race and sex subgroups, we found that there were only minor differences between them. Thus the major distinction between this and previous research is in the number of factors retained.

The Army currently uses a system of ten composites to qualify applicants for specific Military Occupational Specialities. It could be argued that since ASVAB only measures two constructs, as indicated in this paper, there is no need for more than two composites. In fact, the Air Force uses a system of only four composites. However, this by itself is insufficient evidence with which to change the Army system of ten composites which has a strong scientific as well as historical basis. It should be changed only on the basis of strong empirical data which includes an investigation of the criterion-related as well as construct validity of ASVAB. Current work being done by the U.S. Army Research Institute is focusing on these very issues. In the meantime, since the primary purpose of ASVAB is to select and classify individuals for placement into military training and service, perhaps the most important result of this research is that it appears that ASVAB measures similar constructs regardless of race or sex.

References

Joreskog, K.G., & Sorbom, D. (1979). Advances in Factor Analysis and Structural Equation Models. Cambridge: Abt Books.

Kass, R.A., Mitchell, K.J., Grafton, F.C., & Wing, H. (1982). Factor Structure of the Armed Services Vocational Aptitude Battery (ASVAB), Forms 8, 9, and

10: 1981 Army Applicant Sample. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Maier, M.H., & Grafton, F.C.(1981). Aptitude Composites for ASVAB 8, 9, and 10 (ARI Research Report 1308). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ree, M.J., Mullins, C.J., Mathews, J.J., & Massey, R.H. (1981). Armed Services Vocational Aptitude Battery: Item and factor analyses of Forms 8, 9, and 10 (AFHRL-TR-81-55). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Table 1. Four Factor Oblique Unrestricted Maximum Likelihood Solutions.

Subtests	White Males I	II	III	IV	Black Males I	II	III	IV
General Science	38	-07	44	20	32	-04	11	49
Arithmetic Reasoning	15	15	07	62	07	02	74	04
Word Knowledge	16	04	81	01	09	02	-01	87
Paragraph Comprehension	25	19	48	05	14	15	12	53
Numerical Operations	-02	81	03	09	-02	68	16	04
Coding Speed	06	73	01	00	05	82	-06	00
Auto/Shop Information	90	05	01	-12	73	04	-07	07
Mathematics Knowledge	-01	06	01	86	02	04	70	05
Mechanical Comprehension	74	02	-06	21	61	04	19	-05
Electronics Information	63	-01	23	08	54	-03	05	26
	White Females				Black Females			
General Science	37	-09	19	43	56	-04	03	25
Arithmetic Reasoning	12	07	70	08	06	-03	85	00
Word Knowledge	12	00	-00	85	86	-01	01	07
Paragraph Comprehension	11	16	08	57	64	13	10	00
Numerical Operations	-08	72	13	09	02	85	04	-01
Coding Speed	10	76	-04	-04	01	64	03	03
Auto/Shop Information	76	07	-05	03	10	04	-01	50
Mathematics Knowledge	01	06	81	03	05	11	60	09
Mechanical Comprehension	50	00	36	-03	-01	04	20	39
Electronics Information	55	00	06	23	24	-02	03	44

Table 2. Two Factor Oblique Unrestricted Maximum Likelihood Solutions.

Subtests	White Males I	II	Black Males I	II
General Science	81	10	83	-02
Arithmetic Reasoning	36	56	47	31
Word Knowledge	73	21	82	08
Paragraph Comprehension	61	30	67	21
Numerical Operations	-07	86	-01	86
Coding Speed	-03	74	02	71
Auto/Shop Information	87	-12	70	-02
Mathematics Knowledge	27	59	42	31
Mechanical Comprehension	76	07	62	07
Electronics Information	87	-01	79	-04
	White Females		Black Females	
General Science	89	-06	81	-07
Arithmetic Reasoning	56	35	51	27
Word Knowledge	81	07	87	-01
Paragraph Comprehension	63	22	68	16
Numerical Operations	-04	85	01	84
Coding Speed	-03	72	-04	68
Auto/Shop Information	73	-06	52	00
Mathematics Knowledge	48	39	45	32
Mechanical Comprehension	68	07	44	09
Electronics Information	80	-05	64	-06

Evaluation of the ASVAB 8/9/10 Clerical (CL)* Composite for Predicting Training Performance

Mary M. Weltin and Beverly A. Popelka

This report documents test validation research undertaken by the U.S. Army Research Institute at the direction of the Army Deputy Chief of Staff for Personnel to evaluate the effectiveness of the aptitude area composite (CL) currently used to classify entering soldiers into clerical Military Occupational Specialties (MOS). The clerical aptitude area is a composite of subtests appearing on the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8/9/10.

The ASVAB is a battery of ten subtests (see Table 1) that are combined in different ways for use as selection and classification composites. One set of four subtests is known as the Armed Forces Qualification Test (AFQT), and is used to screen applicants for eligibility for military service. Having met the AFQT criterion for selection, an Army applicant must next achieve a passing score on an aptitude area composite for placement into a particular training specialty.

For clerical training, the Army uses the clerical composite (CL) which consists of the four subtests of Word Knowledge (WK), Paragraph Comprehension (PC), Numerical Operations (NO), and Coding Speed (CS). Sample items for each of the ten ASVAB subtests can be found in Appendix A.

Recent emphasis in Army training has been on criterion-referenced testing, which has produced dichotomous pass/fail scores (Maier, 1981). Such scores do not display adequate variance among individuals to reflect differences in levels of performance. The training criteria used in this report were part of a special data collection undertaken to obtain continuous criterion scores suitable for validation.

The purpose of this research was:

1) to validate the operational clerical composite of ASVAB 8/9/10 subtests against school performance in Army entry-level enlisted clerical (MOS), and

2) to identify and evaluate possible alternative composites predictive of clerical training school performance.

---

*To be published as ARI Technical Report 567.

Background

The current CL composite has evolved from a series of research projects on enlisted classification conducted since World War II. Table 2 presents a summary of relevant validation research on clerical MOS. The earliest composite used to classify Army clerical positions consisted of a reading vocabulary test, an arithmetic reasoning test, and a coding speed test. This reflected job analysis (Trump, Marion, & Karcher, 1957) which found that perceptual speed, number facility, memory and perceptual patterning were the important components in clerical jobs.

At the beginning of World War II, a general ability construct was thought to be most predictive of performance in training (PRS 808, 1947). Later, as the available manpower pool decreased, test developers attempted to design composites that would tap aptitudes such as perceptual speed that would be relevant to specific occupational skills.

When the aptitude area system was reconstituted in 1958, each composite contained only two tests, one measuring general ability and one measuring a specialized aptitude. For clerical positions, the composite was Verbal (VE) + Army Coding Speed (ACS) (Maier & Fuchs, 1969). When the composites were again reevaluated in 1969, additional tests were added to improve the validity of the composites, making them again more general in content (Maier & Fuchs, 1972). The testing philosophy of this time reflected a desire to distribute general mental ability equitably across all the MOS groups and to use the classification composites as a secondary screen for applicants of marginal ability.

Although the composition of the subtests and their relative weighting in the clerical composite varied over the years, the same aptitude dimensions (verbal, math, clerical speed) were included until about 1973. At that time, a new classification system was inaugurated. For the new battery, the Army Classification Battery-1973 (ACB-73), the aptitude area composites were constructed to have maximum absolute validity for predicting training performance. Each composite had at least one test of general ability; at least two of the subtests in each composite required the ability to read. However, the cost of building literacy into each composite was that the composites were highly intercorrelated.

The same aptitude composites developed for ACB-73 were also used for ASVAB 6/7, introduced in 1976. For the current forms (ASVAB 8/9/10), Maier and Grafton (1981) built new aptitude area composites by using parallel predictor subtests from ASVAB 6/7 and validating against measures of job proficiency and training success.

Research on clerical jobs in the civilian sector has found results similar to those from the military research. In the Federal civil service, for example, clerical workers have been selected using a test that included vocabulary, paragraph comprehension, alphabetizing, simple mathematics, and typing scales (USCSC, 1973). In an early review, Bennett and Cruikshank (1949) concluded that the best predictor for clerical occupations was a test of general mental ability in combination with a test of perceptual speed. Validation research in the clerical area has generally found validity coefficients between such test scores and training/job success to be quite high. Ghiselli (1966) reported that test validities for proficiency and training criteria in clerical occupations were on the order of .75 across test types. Pearlman (1979) collected nearly 3,400 validity coefficients from 700 studies involving a total sample size of over 470,000. For training performance criteria, he found higher corrected mean validities for measures of verbal ability ($r=.65$) quantitative ability ($r=.71$), reasoning ability ($r=.40$), memory ($r=.59$) and perceptual speed ($r=.40$) than for measures such as motor ability ($r=.35$) and spatial ability ($r=.38$). This may be due to the relatively academic orientation of most clerical training programs.

The impetus for the current validation effort stems from recent emphasis on optimizing utilization of the Army's personnel resources. Because of the large number of clerical soldiers the Army trains each year (18,000 in FY 81), even modest increments in validity can have dramatic payoff in monetary savings attributable to higher performance and lower training attrition costs.

Method

Subjects. Subjects were 3,984 new trainees entering the Army for clerical training in Fiscal Year 1981 (FY81). Only subjects having ASVAB 8/9/10 scores on record were used since previous forms of the ASVAB were not being validated in this research. Twelve of the twenty MOS that use CL for classification were represented in this sample; the remaining eight MOS had to be excluded either because of small sample size or inadequate variance in the criterion measures. Table 3 describes the twelve MOS subsamples showing sizes, and predictor and criterion means and standard deviations (SD).

Measures.

Predictors. Predictor measures were the ten ASVAB subtest scores, which were transformed to standard scores using a standard Army conversion table. A description of these subtests is found in Table 1.

Criteria. End-of-course grades were used as measures of training performance. These scores were standardized within MOS to permit comparison across MOS. Disposition categories, e.g., graduates, recycles (who retake the same course), reliefs (who attempt training in a different course), and failures, were used to set replacement scores as defined below. The graduate/non-graduate dichotomy was used as a secondary performance criterion for validation.

Procedures. Data were edited to remove certain cases: scores for students who did not complete training for medical or disciplinary reasons, scores that were out-of-range, and repeated measures on students who recycled through training. To reduce the error variance in criterion scores, a score replacement technique was used following a rationale developed by Maier (1968). Since the score of record for a student who failed to complete the course may have represented only a partial score achieved up to that point in training, scores for 'recycles', 'transfers', and 'reliefs' were replaced. A 'recycle' is a student who attempts training in the same course a second time; a 'transfer' is moved to a different course of training, and a 'relief' is discharged without further attempts at training.

For recycles, a score one-half standard deviation below the mean of the graduates was substituted for the score of record. For reliefs and transfers, a score one standard deviation below the mean was substituted. About 5% of the total sample was involved in a substitution of this type.

Having standardized and edited the scores, we then computed covariance matrices and corrected the subtest-course grade correlations for restriction in range of the predictor scores using a method described by Lord and Novick (1968). For an unrestricted population reference, the correlation matrix derived from the 1980 Profile of American Youth (OASD, 1982) was used (See Appendix C). No other corrections were made.

A final procedural decision was how best to combine the information from the twelve individual MOS regressions. Two methods of weighting the subsamples were compared: one weighted the 12 MOS equally, the other weighted by the MOS sample size, which reflected MOS accession size.

<u>Analyses</u>. Descriptive statistics (means, standard deviations) for each MOS were calculated for predictor (CL) and criterion (course grade). Frequency distributions were examined for normality, skew, and kurtosis. Multiple correlation and regression analyses were performed for each MOS and for the total combined sample to predict course grade from ASVAB subtest scores. The corrected covariance matrices were used to derive regression equations using stepwise procedures (Hull and Nie, 1981). The multiple correlation between the current composite (WK+PC+CS+NO) and criterion scores was compared with that for an alternate unit-weighted, revised composite suggested by the regression analyses. The stability of the regressions was tested by dividing the two largest MOS in half randomly, developing a prediction equation on one half and cross-validating on the other half-sample.

Differential prediction by sex was examined by comparing the slopes, intercepts, and standard errors of estimate of regression lines developed separately for males and females. These lines were computed from a unit weighted composite of predictors Arithmetic Reasoning (AR), Paragraph Comprehension (PC), and Mathematics Knowledge (MK) regressed against course grade. Subgroup analyses by race were not performed due to ambiguity in the coding of racial categories.

For the pass/fail criterion, simple chi-square analyses were performed to evaluate the predictors' separation of students who successfully complete training from those who do not.

Utility analyses were modeled after Brogden (1946) and Cronbach and Gleser (1965) following an equation developed by Schmidt and Hunter (1981) for computing the potential utility of a selector, using a continuous criterion. Here, the value of an increase in performance (ΔU) due to use of a valid test is a function of the number selected (N), the validity coefficient (R_{xy}), the average standard score on the test for those selected ($\bar{Z}_x$), and the standard deviation of criterion performance in dollars (SD_y) (Hunter & Schmidt, 1981). The assumption is made that the relation between the test scores and training performance is linear.

The total productivity gain is:

$$\Delta U = N\ R_{xy} SD_y\ Z_x .$$

Results

Inspection of the frequency distributions of standardized course grades revealed that, as might be expected in criterion-referenced testing, the grade distributions were negatively skewed. This skewness may have resulted from the schools using criteria that maximized the proportion of students who successfully completed training. No attempt was made to normalize the distributions; the obtained validities may thus underestimate the true relationships.

Results of the correlation, regression analyses, and composite comparisons are presented in Table 4. All coefficients reported have been corrected for restriction in range. The upper portion of this table lists the correlation between each of the individual subtests and the course grade criterion for the designated MOS. The highest single predictor validities were observed for the AR and MK subtests. For the total sample, AR alone predicted training school course grades as efficiently (r=.69) as the four subtests of the current composite (r=.68). Lowest subtest validities were found for Auto-Shop Information (AS) and Electronics Information (EI).

Results of the multiple regressions (SPSS forward regression method) are displayed in the middle section of Table 4. For example, for MOS 71D, WK correlated .49 with course grade, next MK entered the regression, increasing the multiple correlation to .52, followed by AS, etc. The multicollinearity of ASVAB sub-tests makes interpretation of the regressions problematical. However, a common result across nine of the twelve individual MOS regressions was that a mathematics (either AR or MK) subtest consistently accounted for the most course grade variance. Simply adding AR to the current composite raised its validity from .68 to .73.

In an effort to locate the best composite for all twelve of these MOS, we combined the results from the individual MOS validations. Since the relative importance of the individual MOS was not determined a priori, two weighting methods were compared for pooling data across MOS: weighting MOS by sample size, and weighting the MOS equally. Results of both of the pooled-sample regressions (SPSS Stepwise method) were quite similar for either method of weighting MOS subsamples. A composite of three subtests, AR, PC, and MK, predicted approximately 50% of the variance in course grades, with multiple R = .74 . This multiple correlation depends on equalizing the means of the twelve MOS and using beta-weighted predictors in the regression equation.

The subtests of the current composite (WK, PC, CS, NO), by comparison, correlated significantly ($p<.01$) lower (.68) with course grades.

Using the utility formula, $\Delta U = N R_{xy} SD_y \bar{Z}_x$, the savings attributable to using an improved CL composite can be estimated by substituting,

$R = .68$ (for existing CL composite)

$R = .74$ (for improved composite)

$SD_y = \$1,000$; $\bar{Z}_x = 1$; $N = 18,000$

where N is the number of clerical accessions in FY81. The value of performance SD_y, was not determined empirically but was inferred to be $1,000 based on estimates of training performance in similar clerical positions (Hunter & Schmidt, 1982). Assuming selection is made at .3 SD above the mean ($Z = .3, \phi=.381$, $p=.382$, $\bar{Z}_x = 1$). Projecting these figures on future clerical accessions, the potential savings to the Army for increasing validity from the .68 of the current composite to .74 for the new composite is more than one million dollars, each year the improved composite is used.

Operationally, unit-weighted composites have generally been preferred to the less stable beta-weighted composite scores. To evaluate the prediction using unit-weighted subtests, the correlation between obtained course grades and those predicted from a unit-weighted composite of AR, PC, and MK were compared to similar correlations using the CL subtests (WK, PC, CS, NO). These correlations are listed in the lower section of Table 4 to show the composites' relative predictive efficiency across the 12 MOS samples. MOS 75B for example, had the lowest multiple correlation coefficient; 76J had the highest. The revised composite improved prediction in seven of the 12 MOS. Changing the composite does not however significantly alter the rank order of the 12 MOS' multiple correlation coefficients. Cross-validation in two large-sample MOS obtained corrected correlations of .52 for 76C and .68 for 76P.

The check for differential prediction by sex revealed that at lower composite scores, the regression line of a unit-weighted composite of AR+PC+MK against course grade for females lay above the regression line for males (Fig. 1), suggesting possible underprediction for women scoring below the composite mean. The following male and female subgroup statistics were significantly different ($p<.01$): validities ($r_m=.396$, $r_f=.316$), predictor SD's ($\sigma_m=10.1, \sigma_f=9.08$) and criterion SD's ($\sigma_m=19.93$, $\sigma_f=19.68$) . The difference in standard errors of estimate of the two groups ($\epsilon_f=18.68; \epsilon_m=18.29$) surpassed the chance level ($F=1.043$, df=

3145,835, $p<.01$), due perhaps to the large sample size. However, the two regression lines did not differ significantly in slopes or intercepts. Stepwise regression for females (n=836) produced the predictors MK, AR, and NO for a corrected multiple R of .40, while the regression for males (n = 3147) produced AR, AS, and MK for a multiple R of .42. Two of the subtests, AR and MK, are common to both sex's regressions. Differences in the third predictor as well as the absence of PC may be due to subtest multicollinearity.

The second performance criterion, the dichotomous pass/fail separation, was analyzed using chi-square tests. Schools were found to vary greatly in their attrition rates (See Table 5). Some reported virtually no attrition (76W, 73C) while others (75D, 75E) had more than 20% attrition. Schools also differed in disposition methods, for example, whether a student having been unsuccessful in completing course requirements recycled through the same course or attempted other training.

To evaluate whether composite scores could distinguish between those who successfully complete training and those who fail, composite score distributions for graduates and nongraduates were compared. Here, graduates include both regular and accelerated graduates and nongraduates include any students who recycled, transferred, or were relieved from duty. Using the current CL composite subtests, significant differences between graduates and nongraduates were found for two MOS: 71D (χ^2=17.4, $p< .003$) and 75E (χ^2=12.7, $p<.03$). When an alternate composite of AR, PC, and MK was used, significant differences were found for five of the twelve MOS: 71D (χ^2=22.8, $p<.0001$), 71M (χ^2=26.6, $p<.0001$), 75D (χ^2=12.5, $p<.01$), 75E (χ^2=28.2, $p<.0001$), and 76C (χ^2=45.9, $p<.0001$). Despite these significant chi-squares, the overlap of composite distributions for total sample graduates and nongraduates is high (see Fig. 2).

Discussion

This research has shown the importance of mathematics aptitude in clerical training school performance. The results are consistent with previous validation research on clerical MOS. In constructing the ASVAB 8/9/10 composites, Maier and Grafton (1981) found that AR was among the four subtests most predictive of training and job performance in Army clerical occupations. However, at that time it was considered important that all the military services use the same predictors whenever feasible and the other services preferred to avoid a heavy quantitative loading for the clerical areas. The decision was made for the Army to use the same subtests that the Air Force and Marine Corps used to classify clerical personnel (i.e., WK, PC, NO, CS). The reported loss in validity related to Skill Qualification Test (SQT) and training performance criteria was three points, from

.58 to .55 excluding AR (Maier & Grafton, 1981). More recent ASVAB validation research has also found AR to be important in predicting clerical performance (Maier, 1982).

The speeded tests, CS and NO, were not found to be as predictive of clerical performance as would be expected from a conceptual analysis of the job tasks. This may result from attenuation due to unreliability of these subtests in operational use. Compared to power tests, the speeded tests are subject to additional error variance due to timing and practice effects which tend to reduce their test-retest reliability. McCormick and his associates (1982), found that when applicants were allowed to retest repeatedly, scores of speeded tests showed the greatest improvement. An ongoing research project at the Army Research Institute will investigate methods that would permit accurate and reliable measurement of job relevant perceptual skills.

Although the utility analyses revealed substantial potential savings might be possible from modifying the CL composite, the design of all Army aptitude area composites must be considered within the total context of all the positions to be filled. While a composite of AR, PC, and MK is suggested as efficient for predicting training performance, possible opportunity costs of loss in differential prediction must also be considered.

While the Arithmetic Reasoning test is predictive of training success generally, it may be less suitable for classification purposes than a specific predictor such as clerical speed or psychomotor skill since the AR test would probably predict success equally well in many different areas. ASVAB AR items appear to tap a general problem-solving aptitude in which the arithmetic operations are not explicit but are left to the subject to choose. These skills may be more general than the arithmetic operations called for in the mathematics knowledge subtest. However, adding AR would increase the intercorrelations among the existing classification composites since it is already part of four of the composites in addition to the selector composite. While overall systems optimization cannot be effected in any single research project, ongoing ARI research is designed to address the best structure of all ASVAB Aptitude Area Composites using a large sample of Army MOS and a variety of performance criteria. In the interim, it appears the composite identified by this research (AR + PC + MK) is the most valid ASVAB composite for clerical soldiers.

REFERENCES

Bennett, G. K. & Cruikshank, R. M. A summary of clerical tests. New York: The Psychological Corporation, 1949.

Brogden, H. E. When testing pays off. Personnel Psychology, 1949, 2, 171-183.

Brogden, H. E. Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. Educational and Psychological Measurement 1959, 19, 181-190.

Cronbach, L. J. & Gleser, G. C. Psychological tests and personnel decisions (2nd Edition). Urbana, IL: University of Illinois Press, 1965.

Department of Defense. Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics), March, 1982.

Department of the Army. The Adjutant General's Office, Personnel Research Section. Development of aptitude areas for classification of enlisted personnel in the Army. Washington, DC: PRS Report No. 808, Oct. 1949.

Frankfeldt, E. Comparison of ACB and ASVAB clerical tests for use in the Army Qualification Battery (Research Memorandum 70-6). Arlington, VA: U.S. Army Behavior and Systems Research Laboratory, December 1970.

Ghiselli, E. E. The validity of occupational aptitude tests. New York: Wiley, 1966.

Helme, W. H., Denton, B. & Anderson, A. A. Evaluation of tests to predict success in Army clerk course. Washington, DC: U. S. Army Personnel Research Office, Technical Research Note 130, January, 1963.

Helme, W. H. Differential validity of the ACB for courses in seven job areas. Washington, DC: Personnel Research Branch, PRB, Technical Report 1118, April 1960.

Hull, C. H. & Nie, N. H. SPSS Update 7-9. New York: McGraw-Hill, 1981.

Hunter, J. E. & Schmidt, F. L. Fitting people to jobs: The impact of personnel selection on national productivity. In Dunnette, M. D & Fleishman, E. A. (Eds.) Human Performance and Productivity Vol I. New Jersey: Erlbaum, 1982.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley Publishing Company, 1968.

Maier, M. H. Procedures for assigning criterion grades to failures and turnbacks in Army school courses. Washington, DC: U. S. Army Behavioral Sciences Research Laboratory, Technical Research Note 197, April, 1968.

Maier, M. H. Validation of selection and classification tests in the Army. Alexandria, VA: Army Research Institute, Personnel Utilization Working Paper 82-2, November, 1981.

Maier, M. H. Issues for defining ASVAB 11/12/13/14 aptitude composites. Alexandria, VA: Center for Naval Analysis (CNA 82-3199/27), December, 1982.

Maier, M. H. & Fuchs, E. F. Development of improved aptitude area composites for enlisted classification. Arlington, VA: U. S. Army Behavioral Science Research Laboratory, Technical Research Report 1159, September, 1969.

Maier, M. H. & Fuchs, E. F. Development and evaluation of a new ACB and aptitude area system. Arlington, VA: Behavior and Systems Research Laboratory, Technical Research Note 239, Sept. 1972.

Maier, M. H., & Grafton, F .C. Aptitude composites for ASVAB 8,9, and 10. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, Research Report 1308, May, 1981.

McCormick, B., Dunlap, W., Kennedy, R. & Jones, M. The effects of practice on the Armed Services Vocational Aptitude Battery. Paper presented at the meeting of the Military Testing Association, San Antonio, TX, November, 1982.

Pearlman, K. The validity of tests used to select clerical personnel: A comprehensive summary and evaluation. Washington, DC: Office of Personnel Management, Personnel Research & Development Center, Technical Study 79-1, August, 1979.

Swanson, L. Armed Services Vocational Aptitude Battery, Forms 6 and 7: Validation against school performance in Navy enlisted schools (July 1976-February 1978). San Diego, CA: Navy Personnel Research and Development Center, (NPRDC TR 80-1), November 1979.

Trump, J. B., Marron, J. E. & Karcher, E. K. Test of abilities not covered by the ACB. Research Memorandum 57-24, Dec. 1957. (Cited in Helme, et al (1963) op cit.)

United States Civil Service Commission, Washington, DC: Federal Office Assistant Examination, AN 2400R4, 1973.

Vitola, B. M., Mullins, C. J. & Croll, P. R. Validity of Armed Services Vocational Aptitude Battery, Form 1, to predict technical school success (AFHRL-TR-73-7). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, July 1973.

Wagner, M. P., Dirmeyer, R. P., Means, B. & Davidson, M. K. Analysis of aptitude, training, and job performance measures Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics), Technical Memorandum 82-2, February, 1982.

Woods, I. A., Armore, S. J., Droege, R. C. & Karcher, E. K. Prediction of success in personnel and administration courses. Washington DC: Personnel Research Branch, PRB Technical Research Note 56, March 1956.

Woods, I. A., Reuder, M. E. & Bolanovich, D. J. Prediction of success in the clerical procedures and typing course Washington, DC: Personnel Research Branch, PRB Technical Research Note 47, July 1955.

Table 1

Description of the Subtests in ASVAB 8/9/10

Subtest Name	Content	Number of Items	Test Time (Min.)
* Word Knowledge (WK)	Understanding the meaning of words, i.e. vocabulary	35	11
* Arithmetic Reasoning (AR)	Word problems emphasizing mathematical reasoning rather than mathematical knowledge	30	36
* Paragraph Comprehension (PC)	Understanding the meaning of paragraphs	15	13
* Numerical Operations (NO)	A speeded test of four arithmetic operations, i.e., addition, subtraction, multiplication and division.	50	3
General Science (GS)	Knowledge of the physical an biological sciences	25	11
Electronics Information (EI)	Knowledge of electronics and radio principles	25	24
Mathematics Knowledge (MK)	Knowledge of algebra, geometry and fractions	25	24
Auto-Shop Information (AS)	Knowledge of auto mechanics, shop practices and tool functions	25	11
Coding Speed (CS)	A speeded test of matching words and numbers	84	7
Mechanical Comprehension (MC)	Understanding mechanical principles such as gears, levers, pulleys.	25	19

* Tests in Armed Forces Qualification Test (AFQT)

Table 2
SUMMARY OF MILITARY VALIDATION RESEARCH
ON CLERICAL MOS

Criterion	TEST VERSION (dates used)	PREDICTOR	CORRELATION	OCCUPATION (sample size)	REFERENCE
Training Scores	AGCT (1940-1949)	AGCT	.62	Clerk (2329)	Fuchs (1949)
		RV+AR+ACS	.45		
Final coursegrades	ACB-49 (1949-1972)	RV+AR+ACS	.73	Clerk (n/a)	Woods & Reuder (1955)
Final coursegrades	ACB-56	RV+AR+ACS	.58	Steno(1138)	Woods (1956)
			.69	Pers. Adm.(572)	
			.76	Pers. Mgmt.(1112)	
			.72	Admin.(801)	
			.70	Acctg.(738)	
			.72	Postal(588)	
Final coursegrades	ACB-58	ACS	.53	Clerical (13 Courses)	Helme, W. (1960)
		RV	.60		
		AR	.66		
Average written test score		VE+ACS/2	.80	Clerical (1043)	Helme (1963)
		Maximum validity	.85		
Performance ratings in practicum aspects		VE	.76		Helme (1963)
		AR	.68		
		VE+ACS	.82		
Typing test	AQB (1961-1972)	CS	.45	General Clerk 70A10 (460)	Frankfeldt (1970)
		ACS (total)	.35		
		Number reversal	.22		
		Coding	.37		
Final coursegrades	ACB-73 (1973)	CL= AR+WK+AD+CA	.68	Clerical (25,000)	Maier & Fuchs (1972)
		Maximum validity	.78		
Final coursegrades	ASVAB 1 (1968)	Administrative Index	.86	Clerical (Air Force) (1778)	Vitola, Mullins & Croll (1973)
Time-to-train	ASVAB 6/7 (1976-1980)	WK+AD	-.27	Clerical(129) (Navy)	Swanson (1979)
		WK+AD+NO	-.36		
Final coursegrades	ASVAB 8/9/10 (1980-present	MK	.55	Clerical (704)	Maier & Grafton (1981)
		AR	.53		
		NO	.51		
		WK	.50		
		VE+CS+NO	.55		
Final coursegrades	ASVAB 8/9/10	CL= VE+CS+NO	.63	73C (188)	·yer &
			.44	75B (162)	
Time to complete			-.48	75B (162)	

AGCT = Army General Classification Test
ACB = Army Classification Battery
AQB = Army Qualification Battery
RV = Reading and Vocabulary
AR = Arithmetic Reasoning
ACS = Army Coding Speed
AD = Attention-to-Detail
CA = [illegible]

VE = Verbal
NO = Numerical Operations
MK = Mathematics Knowledge
CS = Coding Speed

TABLE 3
SAMPLE CLERICAL MOS
PREDICTOR & CRITERION SCORES

MOS	TITLE	SAMPLE SIZE	CL SCORE MINIMUM	CL MEAN	CL SD	CRITERION	CRITERION MEAN	CRITERION SD
71D	LEGAL CLERK	103	110	119	8.5	Course GPA	80	13.4
71M	CHAPEL ACTIVITIES SPECIALIST	98	95	107	10.5	Course GPA	83	10.6
71N	TRAFFIC MANAGEMENT SPECIALIST	131	95	103	8.4	Course GPA	84	10.2
73C	FINANCE SPECIALIST	214	95	106	11.3	Course GPA	93	6.9
75B	PERSONNEL ADMINISTRATION SPECIALIST	525	95	106	9.3	Course GPA	86	11.1
75C	PERSONNEL MANAGEMENT SPECIALIST	101	95	107	9.9	Course GPA	84	12.1
75D	PERSONNEL RECORDS SPECIALIST	238	95	103	8.9	Course GPA	85	8.5
75E	PERSONNEL ACTIONS SPECIALIST	296	95	108	10.5	Course GPA	84	6.7
76C	EQUIPMENT RECORDS & PARTS SPECIALIST	1215	95	99	8.5	End-of-Course Test Score	87	6.9
76J	MEDICAL SUPPLY SPECIALIST	99	95	104	8.5	Percent of total points first time tested (10 tests)	86	8.4
76P	MATERIAL CONTROL & ACCOUNTING SPECIALIST	618	90	96	7.9	End-of Course Test Score	87	5.0
76W	PETROLEUM SUPPLY SPECIALIST	340	90	91	5.2	Course GPA	92	4.5

TABLE 4

Correlation and Regression Analyses *

A. CORRELATIONS BETWEEN ASVAB SUBTESTS AND TRAINING SCHOOL COURSE GRADES BY MOS

MOS	71D	71M	71N	73C	75B	75C	75D	75E	76C	76J	76P	76W
N	103	98	131	214	525	101	238	296	1215	99	618	340
Subtest												
AR	47	69	72	51	37	46	53	66	51	78	66	65
WK	49	65	67	41	33	51	40	66	47	66	58	63
PC	47	65	67	44	30	45	40	65	44	65	57	55
NO	30	65	66	40	27	37	28	60	39	63	49	53
GS	40	64	63	38	34	46	44	61	46	69	53	59
CS	27	66	64	32	23	42	28	54	32	55	48	47
AS	16	33	38	26	29	31	38	43	37	47	31	53
MK	47	68	71	46	36	49	47	65	50	80	69	60
MC	28	48	52	33	28	42	44	48	43	54	47	57
EI	40	52	56	32	31	45	43	53	44	54	45	58

B. MULTIPLE REGRESSION R'S (SUCCESSIVE STEPWISE)

71D	71M	71N	73C	75B	75C	75D	75E	76C	76J	76P	76W
WK 49	AR 69	AR 72	AR 51	AR 36	WK 51	AR 53	WK 66	AR 51	MK 80	MK 69	AR 65
MK 52	CS 78	CS 79	PC 53	AS 39	MK 55	AS 54	MK 72	WK 53	AR 82	PC 71	WK 69
AS 54	GS 79	PC 80	NO 53	MK 40	CS 56	MK 54	NO 74	MK 54	CS 83	AR 72	AS 71
EI 56	MK 80	MK 81	EI 54	MC 40	EI 57	NO 55	PC 74	AS 55	GS 84	CS 72	CS 72

C. PREDICTOR COMPOSITES' R (UNIT WEIGHTS USED)

	71D	71M	71N	73C	75B	75C	75D	75E	76C	76J	76P	76W
CL	45	77	78	46	33	52	40	72	48	73	62	64
AR+PC+MK	52	75	78	52	38	52	52	73	54	82	67	67

* Coefficients corrected for restriction in range

TABLE 5

FAILURE RATE BY MOS

MOS	N	% ATTRITING*	RECYCLES	RELIEFS	TRANSFERS
71D	103	17.1	1.0	16.1	
71M	98	4.1			4.1
71N	131	11.5	3.4	7.6	
73C	214	0.5	0.5		
75B	525	3.8	1.0	2.8	
75C	101	---			
75D	238	21.8	5.5	15.6	0.8
75E	296	25.9	0.3	0.3	25.3
76C	1215	6.3		6.3	
76J	99	7.0	2.0	5.0	
76P	618	0.7		0.7	
76W	340	---			
TOTAL	3984	6.9			

* For inability to comprehend course material

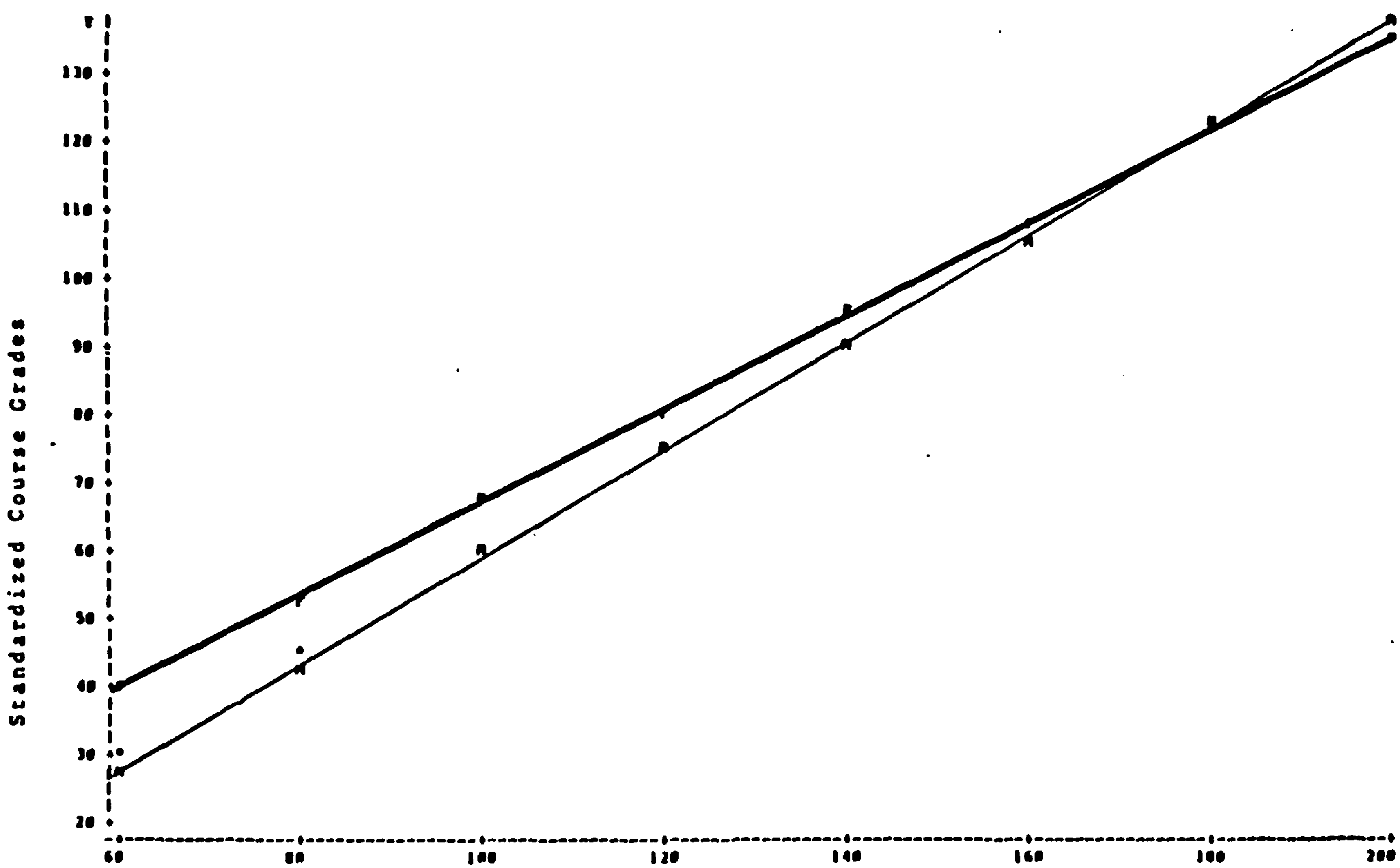


Figure 1

Regression Lines for AR + PC + MK

* = Total
M = Male
F = Female

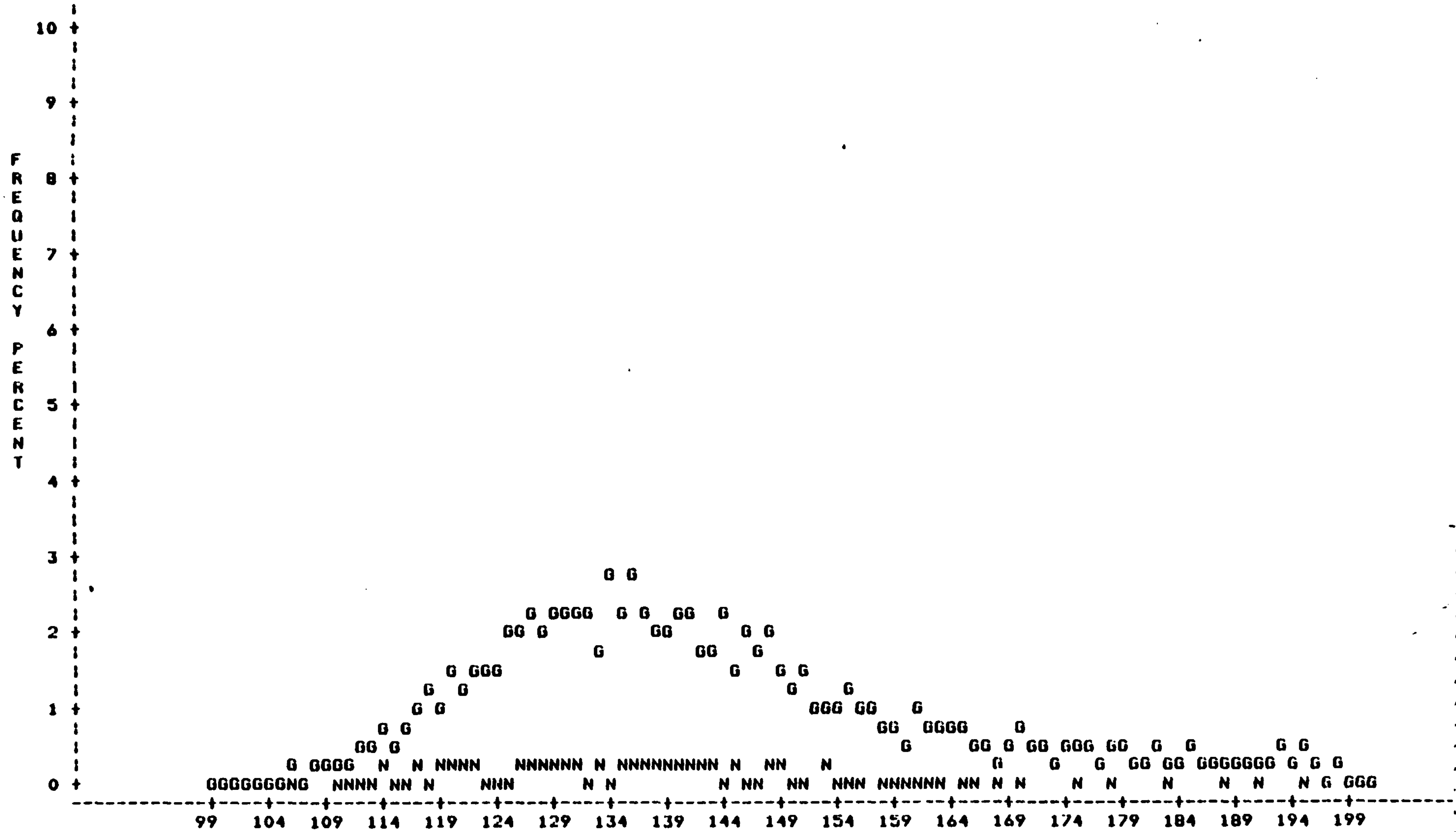


Figure 2

FREQUENCY PERCENT DISTRIBUTION OF

AR + PC + MK

G = TRAINING COURSE GRADUATES

N = TRAINING COURSE NONGRADUATES

Appendix A
(Adapted from DoD 1304.12Z)
Sample ASVAB Questions

GS General Science

Water is an example of a a) solid b) gas c) liquid d)crystal.

AR Arithmetic Reasoning

A person buys a sandwich for 50¢, soda for 25¢, and pie for 40¢.
What is the total cost?
a) $1.00 b) $1.05 c) $1.15 d) $1.25

WK Word Knowledge

Small most nearly means
a) sturdy b) round c) cheap d) little.

PC Paragraph Comprehension

The duty of the lighthouse keeper is to keep the light burning no matter what happens, so that ships will be warned of the presence of dangerous rocks. If a shipwreck should occur near the lighthouse even though he would like to aid in the resuce of its crew and passengers, the lighthouse keeper must

a) stay at his light b) rush to their aid c)turn out the light
d) quickly sound the siren.

NO Numerical Operations

2 + 3 =

a) 1 b) 4 c) 5 d) 6

CS Coding Speed

Key

green	2715	man	3451	salt	4586
hat	1413	room. . . .	2864	tree	5972

Answers

room	1413	2715	2864	3451	4586
green	2715	2864	3451	4586	5972
tree	2715	2864	5972	3451	4586

AS Auto and Shop Information

The most commonly used fuel for running automobile engines is

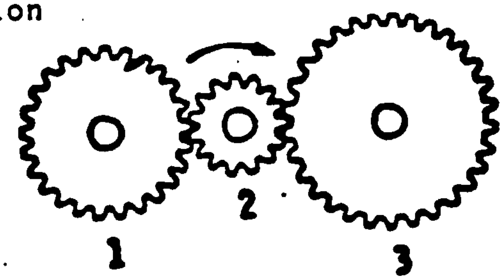
a) kerosene b) benzene c) crude oil d) gasoline.

MK Math Knowledge

If a + 6 = 7 then a is equal to

a) 0 b) 1 c) -1 d) 7/16

MC Mechanical Comprehension



Which of the other gears is moving in the same direction as gear 2?

a) Gear 1 b) Gear 3 c) Neither of the other gears
d) Both of the other gears.

EI Electronics Information

What does the abbreviation a.c. stand for?

a) additional charge b) alternating coil
c) alternating current d) ampere current

APPENDIX B

(Adapted from AR 611-201)

DESCRIPTION OF JOB DUTIES FOR SAMPLE CLERICAL MOS

71D LEGAL CLERK

Prepares legal correspondence, records, and related papers, such as courts-martial, courts of inquiry, and investigations using knowledge of Uniform Code of Military Justice, Manual for Courts-Martial, Manual of the Judge Advocate General.

71M CHAPEL ACTIVITIES SPECIALIST

Performs chapel and religious support functions, such as religious services, counseling, and education as well as general administration and typing duties.

71N TRAFFIC MANAGEMENT COORDINATOR

Assists in receiving, storing, loading, shipping, and unloading supplies, equipment, household goods, and personal effects. Prepares forms and maintains records covering inbound and outbound shipments, recording quantity and condition of property, claims for adjustments for property lost or damaged in shipment.

73C FINANCE SPECIALIST

Performs duties pertaining to pay, leave, travel and maintenance of personnel finance records of military personnel and other finance functions, such as quality edit, disbursing and travel, and general administration.

75B PERSONNEL ADMINISTRATION SPECIALIST

Performs personnel and administrative functions such as personnel actions, personnel accountability (SIDPERS), typing, and general administration.

75C PERSONNEL MANAGEMENT SPECIALIST

Participates in occupational classification and management of manpower resources, such as assignment, replacement, promotion, reduction, classification, evaluation, testing, as well as typing and general administration.

75D PERSONNEL RECORDS SPECIALIST

Maintains officer and enlisted personnel records in records section of personnel activity, to include in/out processing, personnel records maintenance, and preparation of SIDPERS input and control data.

75E PERSONNEL ACTIONS SPECIALIST

Processes personnel actions concerning service members and their dependents, counseling and referring individuals to appropriate support facilities, preparing documentation for reenlistment, discharge certificates, casualty reports.

76C EQUIPMENT RECORDS & PARTS SPECIALIST

Performs duties involving supply of repair parts and maintenance of equipment records. Receives, stores, and issues repair parts. Initiates and keeps records on equipment use and operation.

76J MEDICAL SUPPLY SPECIALIST

Performs requisitioning, receipt, inventory management, storage, preservation, issue, salvage, stock control and accounting of medical supplies and equipment.

76P MATERIAL CONTROL AND ACCOUNTING SPECIALIST

Performs management or stock record functions pertaining to receipt distribution, and issue of Class II, IV, VI, VII and IX material. Performs accounting, editing, document control, record keeping, sales, and direct exchange of such material.

76W PETROLEUM SUPPLY SPECIALIST

Operates and maintains storage and transfer equipment for petroleum products. Distributes petroleum by connecting tanks, operating pump engines, nd opening valves to transfer petroleum. Reads meters and gauges and verifies amount and type of petroleum in storage.

Appendix C

Intercorrelations Between ASVAB Subtests for Profile Study Sample

(N=9173)

	ASVAB Subtest									
	AR	WK	PC	NO	GS	CS	AS	MK	MC	EI
AR	-									
WK	.71	-								
PC	.67	.80	-							
NO	.63	.60	.60	-						
GS	.72	.80	.69	.52	-					
CS	.51	.55	.56	.70	.45	-				
AS	.53	.52	.42	.29	.64	.22	-			
MK	.83	.67	.64	.62	.69	.52	.41	-		
MC	.68	.59	.52	.40	.70	.33	.74	.60	-	
EI	.66	.68	.57	.41	.76	.34	.75	.58	.74	-

AR = Arithmetic Reasoning
WK = Word Knowledge
PC = Paragraph Comprehension
NO = Numerical Operations
GS = General Science
CS = Coding Speed
AS = Auto and Shop Information
MK = Mathematics Knowledge
MC = Mechanical Comprehension
EI = Electronics Information

Table taken from Profile of American Youth, 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery, Office of the Assistant Secretary of Defense (MRA&L) March, 1982.

Validity of ASVAB 8/9/10 as Predictors of Training Success*

Paul G. Rossmeissl
Clessen J. Martin
Hilda Wing

US Army Research Institute for the
Behavioral and Social Sciences

Ming-mei Wang

American Institutes for Research

The Armed Services Vocational Aptitude Battery (ASVAB) is a multiple cognitive abilities test battery used by all the military services for selection and classification of enlisted personnel. Subtest scores are combined in several different ways by each military service. One combination, the Armed Forces Qualification Test (AFQT), can be linked back to assignment of military personnel during the latter days of World War II. Currently, all services use the AFQT as a primary selection hurdle for initial entry. Other combinations unique to each service were developed to predict success in training in the various military schools. The Army's combinations are termed Aptitude Area (AA) composites. Individuals wishing to join the Army must achieve minimal scores in one or more AA to be eligible to begin training in a specific Army school.

Research Problem

ASVAB Forms 8/9/10 were introduced in October 1980; current plans are to have parallel Forms 11/12/13 replace them. The ASVAB high school testing program is presently using (a correctly normed) ASVAB 5, but one of the current ASVAB forms will be transferred for high school use and renumbered as ASVAB 14. The Army uses both training and job performance measures (Skill Qualification Tests - SQT) for test validation. All military services were requested to provide training validity information for the preparation of the new high school Counselor's Manual and the technical manual for ASVAB 8-14. The research described below was completed to meet that requirement.

Method

Predictors

ASVAB 8/9/10 is composed of ten subtests: General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto/Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI). For purposes of Army selection and classification these subtests are combined into aptitude area (AA) composites (see Table 1).

*Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983.

Table 1

The Composition of the ASVAB Composites

Operational Army Composites

AFQT	= VE + AR + .5NO
Electronics (EL)	= AR + EI + MK + GS
Operators/Foods (OF)	= NO + VE + MC + AS
Surveillance/Communications (SC)	= NO + CS + VE + AS
Motor Maintenance (MM)	= NO + EI + MC + AS
Clerical (CL)	= NO + CS + VE
Skilled Technical (ST)	= VE + MK + MC + GS
Combat (CO)	= AR + CS + MC + AS
Field Artillery (FA)	= AR + CS + MC + MK
General Technical (GT)	= VE + AR
General Maintenance (GM)	= MK + EI + GS + AS

Proposed High School Composites

Mechanical Trades	= AR + MC + AS + EI
Office and Supply	= VE + CS + MK
Electronics/Electrical	= AR + EI + MK + GS
Skilled Services	= AR + VE + MC
Academic Ability	= AR + VE

The AFQT is used for the initial selection of personnel and the other composites are used for the assignment of soldiers to the various Military Occupational Specialties (MOS) or jobs within the Army. The proposed ASVAB high school composites will be used for career counseling and with those students who are considering a military career. For the purposes of this research all ASVAB subtest scores were converted to standard scores. These standard scores were then combined, as in Table 1, to form the appropriate composite scores.

Criteria and Sample

The Army does not routinely record end-of-training grades on a soldier's personnel file. For this reason, during calendar year 1981 the Army Research Institute (ARI) requested detailed training data for all MOS with 100 or more entrants per year. Included in these data were end-of-course grades for each soldier. Collection was terminated at 1000 for the high density MOS, and at the end of the year for the remaining MOS. It is these end-of-course grades which formed the criterion measures for this research. It was not possible to find useful criteria for all MOS. Many did not show sufficient variance in the end-of-course grade to be useful in the assessment of predictor validities. For example, in the MOS 16E, HAWK Fire Control Crewmember, 92% of the grades reported were at the maximum value of 100. The analyses of this research were, therefore, limited to a sample of 11 MOS shown in Table 2. These MOS were selected because they all had a fairly large N (operationally

Table 2

MOS Included in the Research

MOS	Name	Army Composite
05G	Signal/Security Specialist	SC
16P	Short Range Missile Crewman	OF
16S	MANPADS Crewman	OF
32D	Tech Controller	EL
33S	Electronic Warfare Systems Repairer	ST
61B	Watercraft Operator	MM
61C	Watercraft Engineer	OF
67Y	Attack Helicopter Repairer	MM
68J	Attack Fire Control Repairer	EL
71D	Legal Clerk	CL
76P	Material Control & Accounting Specialist	CL

defined as 90 or greater) and a training score standard deviation greater than five. Summary statistics for the criterion measures from these MOS are given in Table 3.

Table 3

Summary Statistics for Training Criteria

MOS	N	Training Score Mean	Training Score S D
05G	91	84	7.3
16P	101	83	14.2
16S	514	79	8.3
32D	120	81	14.2
33S	103	82	9.0
61B	92	80	7.7
61C	150	83	6.9
67Y	137	83	6.3
68J	128	86	6.1
71D	96	73	22.9
76P	613	87	5.1

Analyses

The data for the MOS listed in Table 3 required further editing before any validation analyses were performed. First, scores for all soldiers who had attrited from training for non-academic reasons were dropped. Standard scores were then computed for those remaining. Academic attrites were assigned a score of one standard deviation below the minimum passing score and academic recycles were assigned a score that was one-half of a standard deviation below the minimum passing score. This differential score assignment to attrites and academic ~ecyles has been a conventional procedure in ARI validation es arch involving pass/ fail training criteria and does reflect raining achievement differences between these two failure groups.

Four sets of predictors were validated against the criterion measures from each MOS: AFQT, the appropriate AA composite, the appropriate proposed high school composite, and the high school composite for Academic Ability. Uncorrected validities for these predictors were obtained using standard regression analyses. In addition, a stepwise regression (Draper & Smith, 1966) based upon the ten ASVAB subtests was conducted for each MOS. The results of this analysis can be interpreted as the "best" fit of the ASVAB subtests to the criterion data and, therefore, could be used as an index for the fit of the other predictors. Validities for the composite predictors corrected for restriction in range were obtained using Lawley's (1943) general case method. This method can be shown to be mathematically identical to that proposed by Gulliksen (1950).

Results and Discussion

Table 4 presents the validity coefficients obtained from each of the 11 MOS for both AFQT and the appropriate AA composite. For each validity coefficient, the uncorrected and corrected value for restriction in range has been computed. Also reported for each MOS is the uncorrected stepwise best fit estimate based upon all 10 subtests of the ASVAB. Inspection of the uncorrected validity coefficients for the stepwise best fit analysis reveals that in all cases, these values are higher than for either the AFQT or for the corresponding AA composite. The average increment among the 11 MOS for the uncorrected stepwise values in comparison to the AA composite value was .10.

Table 4

Corrected and Uncorrected Validities for Operational Army Composites

MOS	Uncorrected Stepwise Best Fit	AFQT Uncorrected/ Corrected	Army Composite Uncorrected/ Corrected
05G	.61	.55/.81	(SC) .48/.79
16P	.28	.15/.30	(OF) .21/.36
16S	.28	.17/.40	(OF) .23/.44
32D	.46	.44/.67	(EL) .43/.67
33S	.66	.46/.84	(ST) .56/.87
61B	.51	.49/.69	(MM) .45/.65
61C	.58	.45/.73	(OF) .45/.75
67Y	.45	.29/.66	(MM) .39/.75
68J	.53	.28/.62	(EL) .44/.73
71D	.41	.38/.65	(CL) .27/.64
76P	.48	.40/.68	(CL) .26/.60

Inspection of the validity coefficients for the Army composites, corrected for restriction in range, revealed validities ranging from .36 for 16P to .87 for 33S with an average validity coefficient of .66. The average corrected validity coefficient for AFQT was .64. The largest validities were obtained for the

Skilled Technical(ST) composite (.87) and for Surveillance/ Communications(S/C) composite (.79) and the lowest average validity was for the Operators/Food (O/F) Composite (.52).

Somewhat surprising is that, for four of the 11 MOS, AFQT yielded a higher corrected validity coefficient than did the corresponding Army composite. In no instance was the increase greater than .08, as in 76P. Correspondingly, the increased predictive validity for the Army composites in relation to AFQT was greatest in 68J and 67Y, where the increases were .11 and .09, respectively.

High School Composites

This validation research offered the opportunity to analyze the validity of the recently proposed high school composites. Table 5 presents the corrected and uncorrected validities for those MOS having an AA composite either identical to or very similar to the proposed high school composites. The high school Electronics/ Electrical composite is identical to the Army's Electronics composite, thus the results for 32D and 68J are identical to the data presented in Table 4. The high school Skilled Services composite is similar to the Army's Skilled Technical composite with the exception that AR replaces MK in the Army composite and General Science is included in the Army's composite but is not present in the high school composite. The corrected validity coefficient for 33S for the high school Skilled Services composite was .85 and the corrected High School Academic Ability composite was .82. Inspection of Table 1 reveals that the only difference in the high school composite for Skilled Services and Academic Ability is the addition of the Mechanical Comprehension test in the Skilled Services composite; thus, the corrected validity coefficient of .82 for the Academic Ability composite was expected to approximate the validity of the Skilled Services composite for 33S.

e

Table 5

Uncorrected / Corrected Validities for the Proposed High School Composites

MOS	High School Composite	Composite Validity Uncorrected/ Corrected	Academic Ability Uncorrected/ Corrected
05G	N/A	N/A	.59/.82
16P	N/A	N/A	.12/.29
16S	N/A	N/A	.17/.40
32D	Electronics/Electrical	.43/.67	.43/.67
33S	Skilled Services	.49/.85	.42/.82
61B	Mechanical Trades	.47/.65	.51/.70
61C	N/A	N/A	.46/.74
67Y	Mechanical Trades	.32/.72	.25/.64
68J	Electronics/Electrical	.44/.73	.28/.63
71D	Office and Supply	.38/.67	.39/.65
76P	Office and Supply	.42/.69	.40/.68

Technical Report 567

AD-A130939

VALIDITY OF THE MILITARY APPLICANT PROFILE (MAP) FOR PREDICTING EARLY ATTRITION IN DIFFERENT EDUCATIONAL, AGE, AND RACIAL GROUPS

Newell K. Eaton, Mary Weltin, Hilda Wing

Submitted by:
Newell K. Eaton, Chief
SELECTION & CLASSIFICATION TECHNICAL AREA

Joyce L. Shields, Director
MANPOWER AND PERSONNEL
RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

December 1982

Army Project Number
2Q162722A791

Validation of Enlistment Standards

FOREWORD

The high cost of attrition in the all volunteer force led to the development in all military services of attrition screening instruments. The Army's Military Applicant Profile (MAP), based on more than twenty years of research, has been used since July 1979 to screen 17-year-old male non high school graduates. The attrition rate of this group is now equivalent to that of (unscreened) 18-20-year-old male nongraduates.

The question arose as to whether MAP would be effective for older applicants as well. In answering the question of MAP validity for different ages, the influence of education level was also explored. The results show that MAP is an effective predictor of attrition for nongraduates of all ages and races, but that it is less effective for predicting attrition of high school graduates.

EDGAR M. JOHNSON
Technical Director

VALIDITY OF THE MILITARY APPLICANT PROFILE (MAP) FOR PREDICTING EARLY ATTRITION IN DIFFERENT EDUCATIONAL, AGE, AND RACIAL SUBGROUPS

EXECUTIVE SUMMARY

Requirement:

The purpose of this research was to determine the extent to which MAP is a valid screening device for attrition for older male non-high-school-graduate applicants for Army service. The MAP is currently used to screen 17-year-old male nongraduates.

Procedure:

The 1976-77 recruit data base provided the only available data across age, education, and race groups. Chi-square analyses (two-way and multivariate) were used to evaluate the effects of MAP and three demographic variables (education, age, and race) on attrition (six-month stay-leave).

Findings:

MAP scores predicted attrition equally well for all age groups and both graduates and nongraduates. The lower overall attrition rate for graduates minimizes the utility for reducing attrition with this group. Its use for any age group of nongraduate males of either race is supported by these data and analyses.

Utilization of Findings:

This report was prepared for the use by the Army in determining how MAP should be used as a screening device for nongraduate male applicants.

VALIDITY OF THE MILITARY APPLICANT PROFILE (MAP) FOR PREDICTING EARLY ATTRITION IN DIFFERENT EDUCATIONAL, AGE, AND RACIAL SUBGROUPS

CONTENTS

INTRODUCTION[1]

Attrition refers to a soldier's leaving active duty prior to the expected completion of his or her contractual tour of duty; such attrition is a serious problem for the Army. By the time a soldier has completed advanced individual training the Army has invested an estimated $10,000 in recruiting, training and support costs (McConnell & McNichols, 1979). The individual has invested several months of his/her life as well. Efforts to preserve these investments have a long history in military behavioral research (Plag, 1964; Bell, Bolin & Houston, 1974; Bell & Holz, 1975; Bell & Houston, 1976).

Cognitive ability tests such as the Armed Services Vocational Aptitude Battery (ASVAB) are not generally predictive of motivational attrition but biographical questionnaires have been more successful. The self-description technique has been adapted to a wide variety of military screening and selection topics for prediction of job performance as well as for attrition. One such instrument is the Military Applicant Profile (MAP), an autobiographical information questionnaire comprised of 60 items relating to the respondent's family, academic and work experience, athletic/physical competence, self-concept, and social style/participation. MAP was developed from Army Research Institute (ARI) research in military delinquency dating back to the Korean War (Carleton, Burke, Klieger & Drucker, 1957; Johnson & Kotula, 1958).

Initially, weights were assigned to certain items of biographic information. These were combined with aptitude test results to yield a "whole person" score. The higher the score, the better the predicted chances for adjustment and completing Initial Entry Training. The biographic information was age, education, and record of prior civil court conviction (Seeley, 1978). In the next research step, a personal history questionnaire was added to the MAP. The Early Experience Questionnaire (EEQ) inquired about prior civilian activities such as community and extra-curricular school activities, participation in sports, reasons for dropping out of school, personal activities and civilian job experience (Bell, Kristiansen & Seeley, 1974). At this point in the development of a suitable screening device, the decision was made by the Army to emphasize the prediction of attrition among the highest risk group, the high school non graduate. Items most predictive of attrition in non graduates were collected and formed the basis of the current MAP, Forms 4A and 4B. MAP has been used operationally since July, 1979 to screen 17 year old male high school non graduates. Such selection appears to have reduced attrition for this group from 20% to 14% at the six-month point of service (Erwin, 1982).

Because of the MAP's recent success with the younger male non graduate group, expanding its use to older applicants had been proposed previously. A report upon which to base a policy decision was required during 1982. The purpose of this paper was to evaluate whether the prediction of attrition from MAP scores was confounded by education, race or age variables. The earlier research conducted during the developmental stages of MAP (Erwin & Herring, 1977; Frank &

[1]The assistance of Frances Grafton and Betty Teevan in completing many of these analyses was invaluable. Further, John Mellinger, Larry Hanser, and Joyce Shields provided many useful comments and suggestions in shaping the data analyses and interpretation. Finally, Dean Ball provided countless drafts of these pages exhibiting extraordinary persistence and adaptability. The assistance of each of these contributors is gratefully acknowledged.

Erwin, 1978) did not completely address these specific comparisons. A more recent, more extensive data collection using new and existing forms of MAP has been initiated. However, attrition data for this group had not become available at the time of this report. Consequently, the data compiled by Erwin & Herring in 1976-77 from Army recruits were reanalyzed to address the present concerns about age, education, and race effects on prediction of attrition, and provide a report in 1982.

The use of any selection procedure should provide benefits which exceed the costs of operation. Expectancy tables typically display the benefits of a given personnel procedure by incorporating false negative rates. For example, the estimated savings attributable to using MAP as a selection procedure for 17 year old male non graduates between July, 1979, through September, 1981, are over sixteen million dollars. The costs of testing all members of this group with MAP and or recruiting additional applicants to replace those screened out by MAP need to be subtracted from the estimated savings to produce the net gain (or loss) of using MAP. While the latter costs are not known with precision, it is probably true that they are less than the estimated savings.

METHOD

Subjects and Procedure

The subjects were 4,282 male Army enlistees to whom MAP instruments were administered between November, 1976, and February, 1977, and for whom attrition data were available. MAP data were collected at Reception Stations after soldiers had enlisted in the Army and were used for research purposes only, not to select or screen enlistees in any way. Attrition data reflected whether the soldiers were still in the Army 180 days after entering active duty, or had left for "failure to adapt" reasons. Reasons such as lack of motivation, failure to meet physical standards (non medical), and discharges for marginal performance were considered relevant. Soldiers who were discharged for medical, hardship, or related reasons were excluded from the sample. Subjects were classified into subgroups for the age and education analyses. Those who had earned high school diplomas were considered graduates (n=1999); those who had not were considered nongraduates (n=2279). Those who had earned General Equivalency Diplomas (GED) were considered nongraduates, consistent with current Army policy. For four individuals, education status was not known. Subjects were grouped into four age groups (17-year-olds, 18-year-olds, 19-20-year olds, and those 21 and older) and into one of two race groups, blacks (n=1105) and whites (n=2816). There were 357 who were classified as Hispanic (235), other (105), or unknown.

Multivariate Analyses

Multivariate chi-square analyses were used to evaluate the extent to which the relationship between attrition rate and MAP score differed as a function of education level, age, and race. The multivariate chi-square analyses used log-linear estimation of expected frequencies, and provided results which yield, for nominal dependent variable data such as attrition/non attrition, results allowing interpretation of main effects and interactions of independent variables like those possible with analysis of variance (ANOVA). A significant χ^2 for attrition x MAP score is somewhat analogous to a significant ANOVA main effect of MAP, while a significant attrition x MAP x education χ^2 is somewhat analogous to a significant ANOVA interaction of MAP score with education. A detailed discussion of the analysis technique and the limitations of interpretation of its results can be found in the BMDP manual (1981).

Attrition/Exclusion Tables

To address the impact on attrition of using MAP, tables were prepared to provide attrition rate and exclusion rate information as a function of possible MAP cut scores. Cut scores were defined as scores below which an individual would be excluded from service. Exclusion rate was defined as the percent of people who would be excluded with MAP scores below a particular cut score, while attrition rate was defined as the percent of people who would not be excluded by a particular cut score but who would attrite. The tables were prepared using actual data from the 4279 male enlistees in the sample.

RESULTS

Of the 4,278 enlistees in the sample, 621 (14.5%) had left the Army prior to completing 180 days service for failure to adapt. A frequency table was developed, showing attrition rate as a function of MAP score, for all the soldiers in the sample. These data are illustrated in Figure 1, which clearly shows the strong relationship between attrition rate and MAP scores (χ^2 =317., df=6, p<.001).

Multivariate chi-square analyses were used to evaluate the extent to which the relationship of attrition rate and MAP score differed as a function of education, age, and race. Although the total sample size appeared to be large, analyses using more than three variables at a time resulted in unacceptably low cell frequencies, with consequent uninterpretable results. For example, for the graduates, there are 73 seventeen-year-olds, 537 eighteen-year-olds, 793 nineteen and twenty-year olds, and 596 subjects aged 21 or older. For the non graduates, there were 933 seventeen-year-olds, 565 eighteen-year-olds, 482 nineteen and twenty-year-olds and 299 subjects aged 21 or older. Race subgroups of 1105 blacks and 2816 whites became very small when further divided by education, age, and attrition variables. This prevented the execution of a complete "factorial" analysis. Consequently, analyses were executed in logical order, evaluating the contingencies between attrition, MAP, and one other variable only. First, the attrition x MAP x education analysis was conducted, yielding both a strong attrition x MAP relationship (χ^2 =1994, df=3, p<.001), a strong attrition x education relationship (χ^2=56.4 , df = 1, p<.001), and a nonsignificant attrition x MAP x education relationship (χ^2 = 1.4 , df=3, p>.100). The strong attrition x education relationship indicated that for soldiers with the same MAP score, a different attrition rate would be expected for graduates and non graduates. Therefore, graduates and non graduates would have to be treated separately in any use of MAP in personnel decisions based on predicted attrition. The relationship between attrition rate and MAP score is shown in Figure 2 for graduates and non graduates separately. So few graduates scored in the 30-39 point range on MAP that the empirical relationship between MAP score and attrition has been drawn to begin at the 40-49 point interval for graduates.

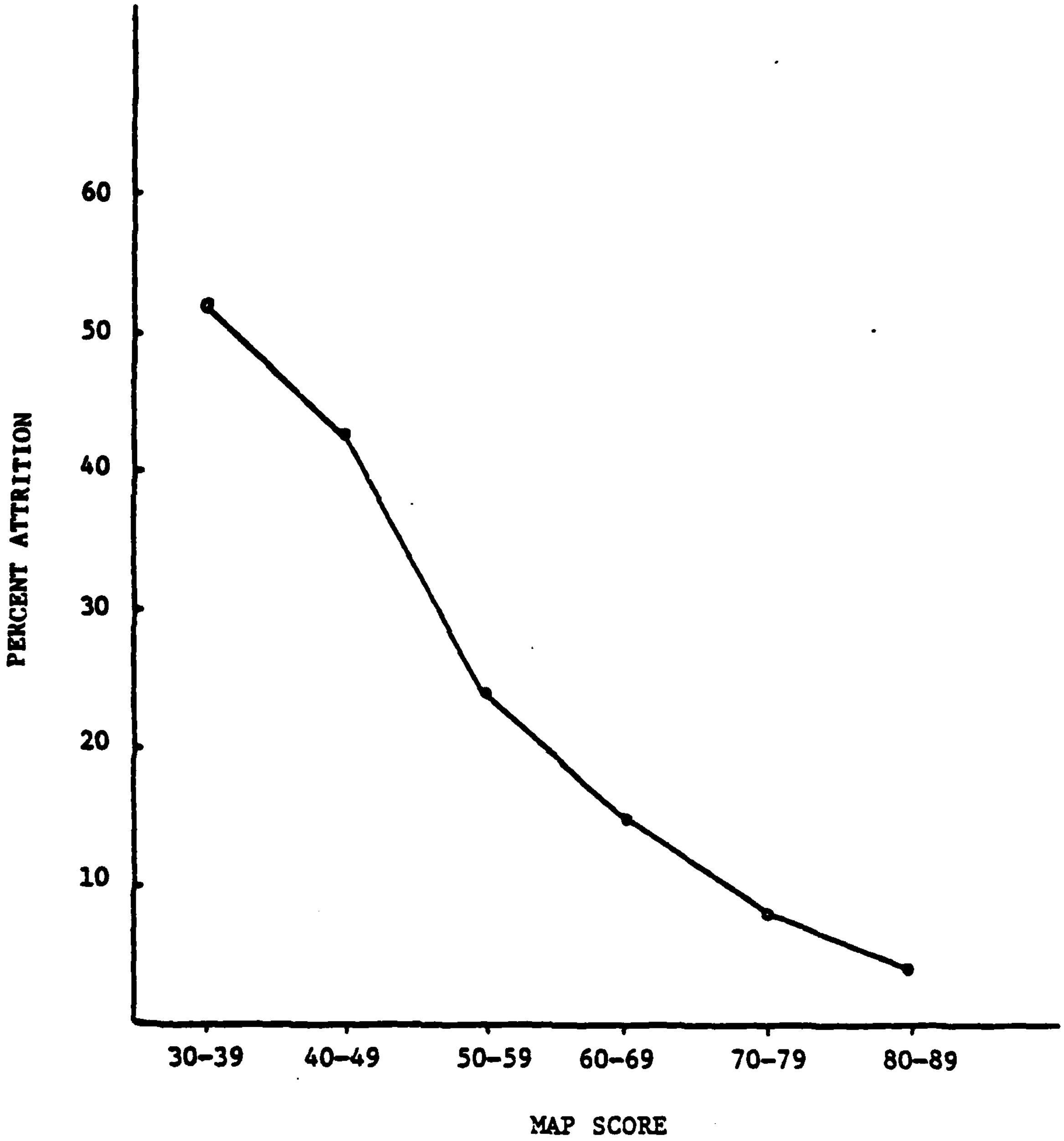


Figure 1. Percent attrition as a function of MAP score for total 1976-77 enlistee sample.

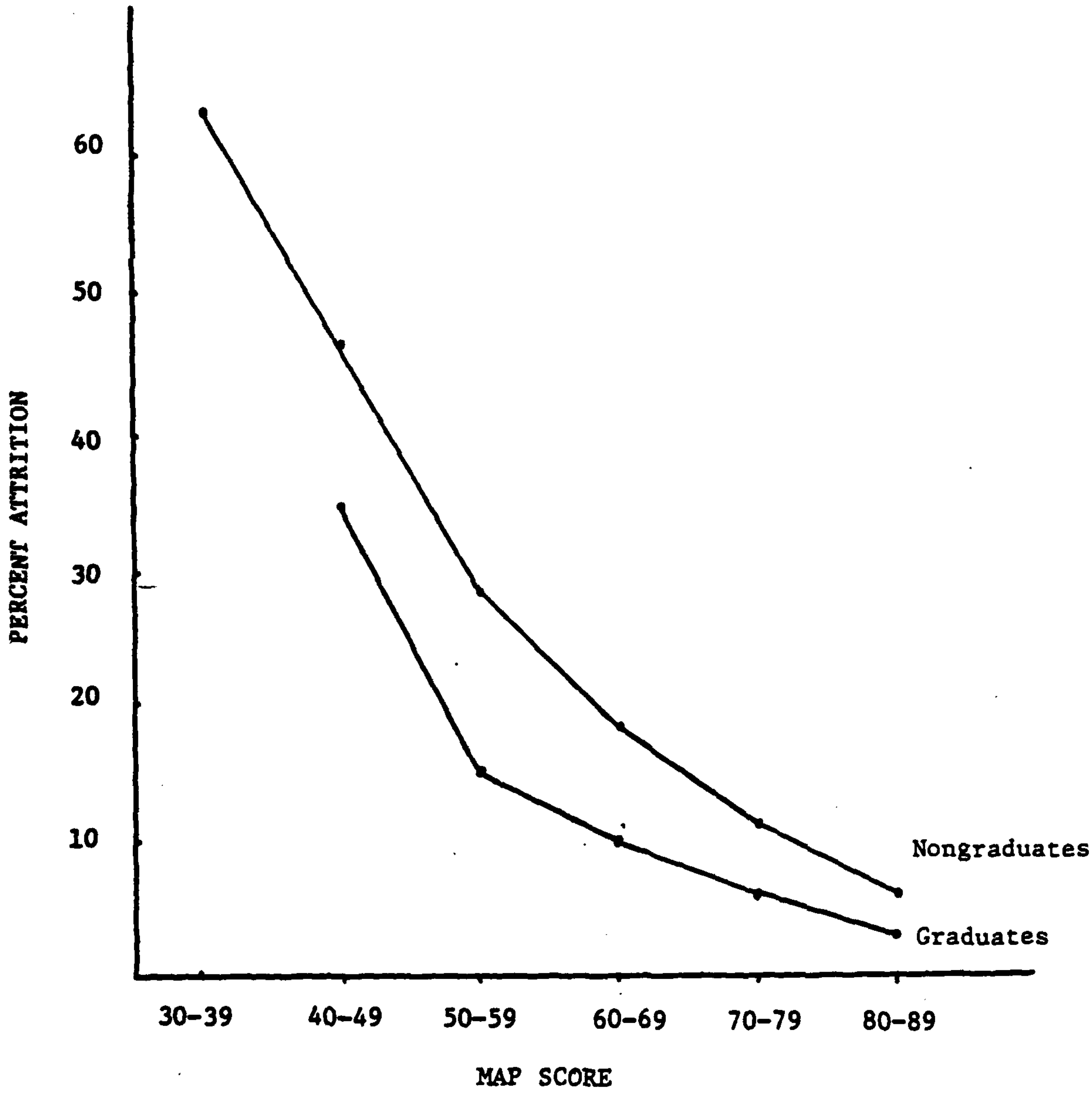


Figure 2. Percent attrition as a function of MAP score separately for high school graduates and nongraduates, for 1976-77 enlistee sample.

Because of the differential impact of education, separate analyses were used to examine attrition x MAP x race relationships for graduates (N=1939) and non graduates (N=2280). For graduates the attrition x MAP relationship was significant (χ^2=33.48, df=3, p<.001), but the attrition x race, and attrition x race x MAP relationships were not (χ^2's = 2.47 and 1.78, df's=1 and 3, p's>.100, respectively). A parallel analysis for non graduate data yielded a much larger, significant attrition x MAP relationship (χ^2 = 138.60, df=3, p<.001) and nonsignificant attrition x race and attrition x race x MAP relationships (χ^2=.39 and 1.94, df's = 1 and 3, p's>.100). These analyses showed that while there was a significant relationship between MAP scores and attrition rate, for graduates and for non graduates, there was no significant effect of race on either education subgroup.

Last, attrition x MAP x age analyses were conducted to determine whether the attrition x MAP relationships differed by age group. Analyses were again conducted separately for graduates and non graduates. The analysis for graduates yielded a significant attrition x MAP relationship (χ^2=38.95, df=3, p<.001) and nonsignificant relationships for attrition x age, and for attrition x age x MAP (χ^2's=5.15 and 13.49, df's=3 and 9, p's>.100, respectively). The same relationships were found with the analysis of non graduate data. The attrition x MAP relation was significant (χ^2=163.79, df=3, p<.001) while attrition x age and attrition x age x MAP were not (χ^2's=4.53 and 5.24, df's =3 and 9, p's>.100.) Because sample sizes differed slightly between analyses using the different sets of three variables, slightly different χ^2 were produced. For example, for graduates χ^2=38.95 for the attrition x MAP relationship in the attrition x MAP x age analysis, but χ^2=33.48 for the same attrition x MAP relationship in the attrition x MAP x race analysis.

Attrition/Exclusion Tables

The purpose of this section is to address the potential impact on attrition rates expected with the use of MAP. Previous analyses showed that MAP scores were significantly related to attrition rate and that the attrition levels associated with MAP scores differed significantly for graduates and non graduates, necessitating separate treatment of MAP scores for graduates and non graduates. Neither the race nor the age variable moderated the attrition x MAP relationships for either educational group.

While these findings demonstrated the relationships between MAP and attrition, they provided no direct information about how policymakers could use a MAP score to reduce attrition. Table 1 was prepared to demonstrate the expected attrition and exclusion rates for graduates and non graduates at specified MAP scores. For example, if a cutscore of 58 were set for non graduates, then 21% of these applicants would be excluded. Attrition rates for those who were accepted would be 15%. Seen from a different perspective, if one wished to reduce attrition to 15% for non graduates, then the table would show that a MAP cut score of 58 would be needed, and an exclusion rate of 21% would result. Observation of the values in the table suggests that while marked shifts in attrition rate can result from changes in MAP cut scores for non graduates, there would be little impact for graduates.

DISCUSSION

The analyses presented were designed to address the applicability of MAP to various educational, age, and racial subgroups of the initial MAP validation sample, and to evaluate possible cut scores for those subgroups. Initial analyses illustrated the strong relationship of MAP scores to attrition for the overall group. The second set of analyses showed that MAP scores, and hence attrition, were a function of high school graduation status. Further, MAP x attrition relationships did not differ significantly across age or race when the two education groups were considered separately. Consequently, MAP appeared to be a valid and appropriate predictor of attrition rate for both race groups and all four age groups when graduates and non graduates were evaluated separately. The third set of analyses provided distributions from which to make decisions about MAP cut scores for use in attenuating attrition. A table was provided to indicate the relationship between cut score, exclusion rate, and attrition rate selected. For example, for non graduates, if an attrition rate of 13% of those selected were desired, a cut score of 66 would be required while 46% of those tested would be excluded. If a higher attrition rate of 15% were acceptable, a cut score of 58 would be used, yielding an exclusion rate of 21% in this non graduate sample.

Several cautions are required in interpreting these data. First, these data were collected in a research context. Motivations, perceptions, or procedures may be different in an operational environment, causing unknown shifts in distributions of exclusion and attrition rates. A lower exclusion rate than that shown is known to exist for the 17 year old non graduates for whom MAP is currently in operational use. Second, these data are five to six years old. Today's youth, farther from the Viet Nam era and in a different economic climate, may respond somewhat differently, with different MAP-attrition relationships. Third, today's youth entering the Army are likely to have somewhat different demographic characteristics from enlistees in 1976-77. Fourth, while the overall sample is large, the sizes of the subgroups are not. This could yield variations between those distributions shown and the actual population distributions. Fifth, it was with these data that the operational MAP keys were developed and hence the validity may be somewhat inflated due to chance factors. Any application of this key to a different group is likely to have somewhat reduced validity.

These 1976-77 data show that attrition rates as a function of MAP scores of male Army applicants are quite different for high school graduates and non graduates, but are essentially the same for different race and age groups. Based on these data, the 1979 policy decision to require minimal MAP scores from 17 year-old non graduates was technically sound. Further, the data indicate the requirement may properly be extended to additional ages of the non graduate group. Given the low attrition rate for graduates, use of MAP for these personnel appears to be unwarranted.

Using these data as a basis for a policy decision to extend the operational requirement of MAP to different ages of non graduates should be carefully monitored. Verification based on the 1982 data collection is required. The reports of these data will be available in late 1983.

Table 1

Exclusion and Attrition Rates as a Function of
Military Applicant Profile (MAP) 4B Cut Score
for Male Enlistees, by High School Graduation Status
1976-1977 Sample

	Graduates		Non Graduates	
MAP 4B Cut Score	Excl Rate	Att Rate	Excl Rate	Att Rate
85	98	5	100	0
80	89	4	97	6
78	81	4	93	9
76	71	4	88	10
74	63	4	81	10
72	51	4	73	10
70	42	6	65	11
68	33	6	56	12
66	26	6	46	13
64	20	6	37	14
62	15	7	31	14
60	12	7	25	14
58	8	7	21	15
56	7	7	17	16
54	5	8	13	16
52	4	8	10	17
50	3	8	8	17
45	1	8	4	18
40	1	8	2	19
35	0	8	2	19
30	0	8	1	20
n	1999	169	2279	452

References

Bell, D. B., Bolin, S. F., & Houston, T. J. Development of the Background and Opinion Questionnaire 72. Arlington, VA: Army Research Institute for the Behavioral and Social Sciences, Research Memorandum 74-13, December 1974.

Bell, D. B., & Holz, R. F. Summary of ARI research on military delinquency. Arlington, VA: Army Research Institute for the Behavioral and Social Sciences (ARI Research Report 1185), June, 1975.

Bell, D. B., Kristiansen, D. M. & Seeley, L. C. Initial considerations in the development of the Early Experience Questionnaire (EEQ) Arlington, VA: Army Research Institute, ARI Research Memorandum 74-10, July 1974.

Bell, D. B. & Houston, T. The Vietnam Era Deserter. Alexandria, VA: U.S. Army Research Institute for the Behavioral & Social Sciences, Research Problem Review 76-6, July 1976.

BMDP Statistical Software, 1981 Edition. Los Angeles, CA: University of California Press.

Carleton, F, O., Burke, W. A., Klieger, W. A. & Drucker, A. J. Validation of the Army Personality Inventory against a Military Adjustment Criterion. Washington, DC: Personnel Research Branch, The Attorney General's Office, Department of the Army, PRB Technical Research Note 71, May 1957.

Erwin, F. W. & Herring, J. W. The feasibility of the use of autobiographical information as a predictor of early Army attrition. Alexandria, VA: US Army Research Institute, Technical Report TR-77-A6, August, 1977.

Erwin, F.W. Personal communication, 1982.

Frank, B. A. & Erwin, F. A. The prediction of early Army attrition through the use of autobiographical information questionaires. Alexandria, VA: US Army Research Institute, Technical Report TR-78-A11, July 1978.

Johnson, C. D. & Kotula, L. J. Validation of experimental self-description materials for general and differential classification. Washington, DC: Personnel Research Branch, The Attorney General's Office, Department of the Army, PRB Technical Research Note 95, August 1958.

McConnell, R. D. & McNichols, G. R. Marginal Pipeline Costs of Enlisted Personnel. Falls Church, VA: Management Consulting & Research, Inc. TR-7808-1, Dec. 1979. Prepared for Office of the Assistant Secretary of Defense, Manpower, Reserve Affairs & Logistics.

Plag, J. A. The validity of age, education and GCT scores as predictors of two-year attrition among Naval enlistees. Navy Medical Neuro-psychiatric Research Unit, San Diego, CA: U.S. Navy Personnel Research Center, June 1964.

Seeley, L. C. Early development of the Military Aptitude Predictor, Washington, DC: Army Research Institute, Technical Paper 288, March 1978.

Working Paper RS-WP-83-5

READING ASSESSMENT IN THE ARMY *

REBECCA L. OXFORD-CARPENTER

LINDA J. SCHULTZ

SELECTION AND CLASSIFICATION TECHNICAL AREA

NOVEMBER, 1983

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

* To be published as Working Paper Selection and Classification Technical Area 83-7.

FOREWORD

The following paper was written to serve two purposes. One was to provide a general summary of the available information on the general topic of reading assessment in the Army in response to repeated inquiries from a variety of sponsors, including FORSCOM, DCSPER, and TRADOC. The other was to respond to a specific request from DSCPER relating to use of reading grade level as a potential reenlistment criterion for midterm noncommissioned officers.

READING ASSESSMENT IN THE ARMY

The purpose of this paper is to discuss reading assessment in the U.S. Army drawing upon experience with civilian and military reading tests. The paper is divided into a number of sections: first, the relationship of cognitive development theory to reading ability and reading assessment; second, definitions of aptitude, achievement, and general ability tests; third, types and purposes of reading tests; fourth, characteristics to look for in a test; fifth, special issues of test validity and test bias; sixth, adult reading tests available from commercial sources; seventh, adult reading tests available from military sources; eighth, correlations between specific reading tests and the Armed Services Vocational Aptitude Battery (ASVAB); ninth, use and abuse of reading grade level, more commonly known as grade equivalent, scores; tenth, other potentially useful scores; and eleventh, recommendations for future reading assessment.

How Does Cognitive Development Theory Relate to Reading Ability and Reading Assessment?

A discussion of the role of cognitive development theory is germane to the understanding of the mental processes involved in the acquisition of reading skills. Knowledge of these processes facilitates *constructive* assessment of reading ability. A useful model that has been identified in this area of cognitive development as it relates to reading is that proposed by Sticht, Beck, Hauke, Kleiman, and James (1974). The Sticht et al. model is an eclectic, information-processing approach which incorporates some of the basic tenets held by noted psychologists and cognitive researchers, among them Piaget, Gibson, Blumenthal, Bloom, Bruner, Durkin, and Hochberg.

This developmental model of auding and reading basically attributes the acquisition of literacy skills to an interaction between the environment and the cognitive processes which involve languaging. Sticht et al. (1974, p. 14) define auding as "the process of listening to speech in order to language." Languaging is the process for representing conceptualizations in such a way that communication may occur.

According to this model, the environment is perceived as a sea of energy from which the individual can derive perceptual experience through the senses. Such perceptions constitute the mental constructions which represent the structural information contained in the environment. When energy change patterns involve the auditory nervous system, the sense of hearing is called into play and an auditory percept results; a change involving the visual system concerns the process of seeing and results in a visual percept. A percept, according to the American Heritage Dictionary (1982, p. 920), is defined as, "An impression in the mind of something perceived by the senses, viewed as the basic component in the formation of concepts." Hearing and seeing together comprise the basic adaptive processes which function as precursors to languaging. When the memory processes are combined with the sensory percepts resulting from hearing and seeing, the receptive precursors to languaging, i.e., the processes of looking and listening, are then initiated. Looking and listening are therefore sensorimotor information-processing activities which permit manipulation of the external environment in order to language.

Languaging provides the means of communicating ideas by using properly ordered sequences of signs for the representation of mental conceptualizations, which are derived from the cognitive content and processes. A conceptual base of cognitive content resides in the memory; it is composed of abstract elements and mental constructions which permit conceptualizing. Languaging therefore involves an indirect mapping of conceptualization to language and vice versa.

The memory is viewed as being composed of three fundamental components which include the sensory information storage (SIS), the short-term memory (STM) and the long-term memory (LTM). The SIS information is recoded into language information during processes functioning in the STM. The STM processes are under the control of the individual. The LTM is a virtually permanent memory store composed of cognitive content. This cognitive content contains the conceptual base and the language system as subcontents. The control processes occurring in the STM act to unite structural information accessed from the SIS with information retrieved from LTM. Thus, sensorimotor activity is at the heart of the formation of the conceptual base which is vital for the conception of languaging.

Listening and looking have already been identified as the receptive precursors to languaging. The process of languaging itself involves the oracy and literacy components. Just as literacy entails reading and writing, oracy includes auding and speaking. Literacy cannot be attained without first having mastered oracy, except in the case of the deaf and others who may acquire literacy skills without the benefit of listening and spoken language. Thus, an individual progresses through several stages of cognitive development from usage of the basic adaptive processes through intermediary processes of interacting with the functional environment and the cognitive content, and culminating in the attainment of various degrees of oracy and literacy skills. Because oracy and literacy require languaging, the three linguistic subsystems of phonology, syntax, and semantics must be mastered and used.

This paper is primarily concerned with the reading component of literacy. According to the Sticht et al. model, reading may be viewed as a special case of the more general information-processing activity of looking; it is in effect looking at script in order to language. For most people, reading cannot occur until the oracy competency has been rather well developed, because it necessitates the same language content, i.e., signs and rules. However, reading requires additional special competencies beyond those for oracy competency that are needed for understanding language in written form and that involve the graphic display of language elements. Indeed, it can be said that reading ability is dependent upon two sets of factors: those concerned with languaging and those needed for visually exploring and extracting information from visual displays.

In association with the languaging factor, reading involves a decoding stage in which the printed symbols are converted into language signs which are then used in conceptualizing. The visual factor addresses information processing as a function of the focus of attention and the margin of attention which deal with the processing of information serially and in parallel, respectively. Reading performance may be hindered if the rate of focal attending is limited because the rate at which conceptualization can be formed is decreased. With practice a reader learns to scan visual material in the margin of attention while focal attention is devoted to conceptualizing its content. That is, decoding proceeds automatically and a concept is perceived. When reading decoding skills become automatic, reading then becomes primarily a process of languaging and conceptualizing.

In analyzing reading problems, Sticht et al. have isolated the areas most needing improvement. They (1974, p. 68) assert that "most people in our society learn to read/decode reasonably well--it is the lack of languaging and conceptualizing which are the major factors in functional illiteracy; that is, the inability to perform a given set of reading tasks." To improve reading ability Sticht et al. suggest that one goal of a literacy training program might be to ascertain the disparity between one's auding level and one's reading level and then train that individual to comprehend by reading what he can comprehend by auding. Because of the literacy component of languaging, reading competence is further restricted by the numbers of visual signs (i.e., a low frequency vocabulary) that are contained within the linguistic subcontent of an individual's cognitive memory. Therefore, a limited vocabulary exercises a restraint on the reader's ability to extract information from the SIS, to recode it into meaningful conceptualizations, and to relate it to prior knowledge. A broad range of vocabulary and of conceptual knowledge is thus important in the acquisition of good reading skills.

There are various methods for assessing reading that are used in education and have been described in the literature. For example, the ingredients in the acquisition of good reading skills (i.e., range of vocabulary and extent of prior knowledge or conceptual base that an individual possesses) also provide a means for assessing reading ability and reading comprehension. Langer and Nicolich (1981) found that this kind of knowledge had an impact upon a reader's ability to comprehend a particular text successfully and that level of prior knowledge greatly related to ability to recall a passage. Therefore, measures of vocabulary and concept knowledge provide indices for assessing reading comprehension and ability.

Another popular technique for assessing reading comprehension and ability is the cloze procedure, whereby the reading examinee is asked to either supply by way of free-response or by multiple choice a missing word based on the context. Tables for converting cloze test performance to reading grade level score equivalents have been devised by Bormuth (1975). The availability of these conversion tables combined with the simplicity of the procedure make the cloze method an attractive option for determining word recognition, reading comprehension, and reading ability. Additionally, the base material used for the reading test under the cloze procedure may be work-related, or job-specific, and therefore may be advantageous for literacy screening into various training programs.

Other methodologies for assessing reading include standardized tests of general reading ability, readability analysis, and reading rate in terms of words per minute read. Because of high correlation, tests of general intelligence may even be used to assess reading ability (Jensen, 1981). Now that we have an idea of what constitutes the reading process and of how we can assess it, we shall proceed to a discussion of the general types of mental tests that are available.

What is the Difference Between Aptitude and Achievement Tests?

Because, according to Jensen (1981), both achievement and aptitude tests are measures of general ability, the difference between them is essentially a time factor. Aptitude tests project into the future, i.e., they predict performance of a particular kind. The content domain is relatively specialized and is designed to investigate such issues as the probability of success in a training course or the level of proficiency on the job. Achievement tests, in contrast, "postdict" (Sticht, personal communication, September 20, 1983) achievement, i.e., assess specific knowledge or skill that has been learned, for example, at the conclusion of a particular course of study. A test that assesses achievements attained over many years in broad and varied areas of experience extending beyond the limits of formal schooling is then termed a test of general ability, according to Jensen's definition. The same test may even function as an aptitude, achievement, or an ability test from one administration to another, depending on the purpose it is serving at a particular time. Having knowledge of what general types of mental tests exist and what their functions are, we now need to investigate specifically the types and purposes of reading tests.

What Are the Types and Purposes of Reading Tests?

Reading tests, like tests in other subject matter areas, have different purposes and come in different packages. Table 1 compares characteristics of three general kinds of reading tests--norm-referenced, objective-referenced, and criterion-referenced.

Some reading tests are designed to rank individuals in broad areas such as reading comprehension and vocabulary (Hambleton & Novick, 1973). Such tests usually assess a person's ability or achievement in those areas as compared to the ability or achievement of others in a norm group. Tests of this type are "norm-referenced" (NRT), because they compare or reference an individual's performance to the norms provided by a separate group, upon whom the test was originally standardized. Even though examinee comparison is used primarily as the basis for norm-referenced score interpretation, it should be pointed out that any good test item taps content from a domain carefully specified in the test blueprint. Norm-referenced reading tests are often used for screening, placement, selection, classification, evaluation of individual progress across a broad subject area, and overall program evaluation. Typical norm-referenced scores include grade equivalents (called grade levels in the military), percentiles, stanines, and standard scores. Norm-referenced reading tests include such well-known, useful instruments as the Metropolitan Achievement Test - Reading Subtest, the Nelson-Denny Reading Test, the Test of Adult Basic Education, and the Adult Basic Learning Examination. The armed services have used commercial, norm-referenced reading tests extensively but not exclusively. In one case, the U.S. Armed Forces Institute (USAFI) officially put its own cover on the Metropolitan Reading Subtest and called it the USAFI Reading Test. The ASVAB is the armed services' largest, most influential "home-grown" norm-referenced test. Although the ASVAB does not have a reading subtest per se, several verbal subscales are available, such as Word Knowledge and Paragraph Comprehension. Also, the entire ASVAB is known to be highly reading-related, as we discuss later. Additional information relative to norm-referenced testing and measurement may be found by reading Ebel (1972), Mehrens and Lehmann (1978), and Angoff (1971).

A second kind of instrument is known as the "objective-referenced" test. This kind of test is built by a careful process of identifying specific objectives based on an analysis of subject

TABLE 1

COMPARISON OF TEST CHARACTERISTICS

What Kind of Test?	To What Is Individual Performance Compared?	What Is the Purpose?	How Is the Test Used?	What Kind of Scores Are Available?	What Are Examples?
Norm-referenced	Performance of norm group	Ranking of individuals	Screening, placement, selection, classification, individual's general progress evaluation, general program evaluation	Grade equivalents, percentiles, stanines, standard scores	Adult Basic Learning Examination
Objective-referenced	A set of specific objectives	Determining how an individual performs relative to a set of objectives	Screening, individual's specific progress evaluation	Percentage or number correct on objective or test	Army Job Reading Test*
Criterion-referenced	A prespecified performance criterion for "passing" or "mastery"	Determining whether an individual has met the criterion	Diagnosis, individual's specific progress evaluation, specific *program evaluation*, selection, classification	Number of objectives passed, GO/NO-GO, number of tries before pasing, time to mastery	Many Army training tests

*Can be interpreted in objective-referenced way.

matter, desired skills, and so on. The objective-referenced test approach can be traced to the early work of Gagne (1965), who defined procedures for the analysis and classification of behavioral objectives. Test items are carefully written to measure performance as compared or referenced to each objective, and ordinarily there are at least several items to measure each objective. In order to facilitate usage of objective-referenced instruction and testing, Morreau (1974) has developed a hierarchical analysis and classification schema which encompasses both behavioral and non-behavioral objectives. The underlying assumption of this procedure is that all objectives are or can be functional. Morreau has identified four levels which serve as the basis for the structural analysis and classification of objectives which range from the abstract or global to the very specific or behavioral. These four levels include the conceptual, educational, instructional, and behavioral objective levels. At the conceptual level general goals are presented. At the educational level, goals are set and compared, e.g., on an interdepartmental or intergroup basis. At the instructional level, goals for specific levels of instruction are set and prerequisite skills for course entrance are stated. At the behavioral level, criterion measures are established and evaluation occurs in order to provide for pupil-progress indicators. Morreau's system thus permits the integration of tests into an objective-referenced framework.

Purely objective-referenced tests can be used for screening but generally have limited use otherwise due to lack of norms and of a "mastery" or "passing" criterion. Typical objective-referenced scores are number or percentage of items correct per objective and number or percentage of items correct per test. For example, if the objective concerns topic sentences in a paragraph, a person could be said to have identified the correct topic sentences in three out of four paragraphs, or to have a score of 75% for that objective. Few commercial adult reading tests are purely objective-referenced. However, some military reading tests, such as the Army's early Job Reading Task Tests, discussed later, might be interpreted in an objective-referenced way. In general, while many commercially prepared tests are sold as "criterion-referenced," they may in fact be more appropriately categorized as "objective-referenced."

Unlike objective-referenced tests, "criterion-referenced" tests (CRT) compare a person's performance to a prescribed performance standard or criterion which indicates "passing," "mastery," "acceptable performance," or (in Army terms) "GO." Usually criterion-referenced tests are constructed using specific objectives, as found in objective-referenced tests; however, criterion-referenced tests add the crucial performance criterion. According to Glaser and Nitko (1971, p. 653), "A criterion-referenced test is one that is deliberately constructed so as to yield measurements that are directly interpretable in terms of specified performance standards." In addition, criterion-referenced testing emphasizes a well-specified content domain and the development of procedures for generating appropriate samples of test items (Hambleton & Novick, 1973). To use an example similar to the one presented above, the prescribed performance standard might be that a person must correctly identify the topic sentences in at least three out of four paragraphs in order to master or pass the objective or to receive a "GO." Typical criterion-referenced scores are number of objectives passed, GO vs. NO-GO on the total test, number of tries before passing, and time to mastery. Criterion-referenced tests are frequently used for diagnosis of specific weaknesses and strengths before and after a period of instruction--a purpose that cannot be served as well by the more general, broadly based, norm-referenced surveys. Criterion-referenced tests can also serve a number of the purposes listed for norm-referenced tests, such as screening, individual progress evaluation, and perhaps program evaluation (Hambleton & Eignor, 1978). However, criterion-referenced tests do not provide information about how an individual performs in comparison to his or her peers in a norm group, so purposes such as placement, selection, and classification may not be as well fulfilled by criterion-referenced tests as by norm-referenced tests. Criterion-referenced tests are becoming increasingly popular for measuring preadult reading, but few if any commercial adult reading tests are criterion-referenced. The armed services often use criterion-referenced tests in their training installations. The Army Job Reading Test, described later, could be used in a criterion-referenced way if performance standards for passing or mastery were employed. For more in-depth discussion of criterion-referenced testing, see Popham (1978), Popham and Husek (1969), Hambleton and Gorth (1971), Ebel (1975), and Kriewall (1972).

Some other less frequently encountered types of reading measures are "domain-referenced" and "adaptive" or "tailored" tests. It is important to know about these kinds of tests, because they may represent future trends. However, it is not crucial to discuss them in great depth here.

Domain-referenced tests (DRT) are tests whose items are selected by means of a particular sampling strategy from a carefully defined knowledge domain. The intent is to be able to generalize from the sampled items to the entire domain with a certain degree of confidence; that is, to be able to say that if the person performs in this fashion on the tested items, he or she would in all likelihood perform similarly on the whole domain if it were possible to test the person on the whole domain (which it is not). Some experts say that criterion-referenced tests are a subset of domain-referenced tests; other experts say that domain-referenced tests are a subset of criterion-referenced tests. According to Hively (1974), domain-referenced testing is rooted in learning theory, and items for a domain-referenced test may be generated in an empirical way or via a logical approach. Furthermore, items may cross categories, i.e., the same item may be a member of several different sub-domains based upon the characteristics used to classify it, and items are not limited to those related to paper-and-pencil tests. There are no acceptable, trusted domain-referenced tests in the area of adult reading at this time. For further discussion of domain-referenced tests and related issues see Baker (1974), Millman (1974), and Sension and Rabehl (1974).

Adaptive or tailored tests are those in which items that are the most appropriate to the individual in terms of difficulty level are selected by means of an iterative process of asking easier and then harder questions until the right level for the person is found. This process allows the individual to take only the items that are the most useful in discriminating his or her performance from the performance of others. Thus, administration time is shortened (McBride, 1979). Adaptive or tailored tests are a breed of norm-referenced tests and therefore have the capability of ranking individuals against norms. Lord has conducted work in the area of adaptive/tailored tests. His self-scoring flexilevel test (1971) was an early branching test designed to address the appropriate ability level of an examinee. Additional sources which provide insight into the theory and development of adaptive or tailored testing include Lord and Novick (1968), Weiss (1975), Wood (1973), Urry (1977), and other works by Lord (1971, 1977). The joint services are developing and testing a computerized-adaptive verion of the ASVAB at this very moment in installations throughout the U.S. However, there are no well-known adaptive or tailored tests of adult reading ability or achievement.

What Characteristics Should Be Looked for in a Reading Test?

In selecting a reading test, as in choosing any other type of cognitive achievement test, we can look for certain important characteristics. First, does the test have adequate validity, that is, does it measure what it purports to measure? Several types of validity may be relevant, such as content validity, concurrent validity, predictive validity, and construct validity. Any useful test must have demonstrated validity of one of these kinds.

Content validity pertains to how well the test samples the domain of subject matter representative of skills from which conclusions are to be drawn; it is particularly crucial when using achievement and proficiency measures. The test content and the language used must be up-to-date and relevant to the group with whom we want to use the test. We must beware of childish or antiquated content and language in selecting tests for adults. Other aspects of content validity concern availability of appropriate subtests, proper item types, and good item quality. There is no specific numerical expression used to indicate content validity, rather it is derived through a thorough inspection of test items.

Concurrent and predictive validity represent two forms of criterion-related validity which look at the relationship between test scores and some independent external measures, i.e., criteria. Concurrent and predictive validity vary according to the time at which the criterion data are gathered. Data relating to predictive validity are gathered later. That is, these data concern the correlation between test scores and some measurement of future performance on a criterion of interest. Concurrent validity addresses criterion data that are collected at approximately the same time as the test data. Concurrent and predictive validity are usually expressed statistically as a correlation coefficient, such as .65, which means the test correlates with another measure at a (positive) level of .65 out of a possible 1.00.

Construct validity concerns the use of tests for scientific inquiry and for the assessment and measurement of human traits, i.e., constructs. Construct validation addresses theory confirmation and relates to the degree to which specific human traits, e.g., intellectual, interest, or personality characteristics, account for performance on a test. The operational definition of constructs and the specification of the relationship between constructs are required in theory development. The construct validation process is used to clarify these constructs and to test the basic theory whose traits are embedded within the test under investigation. For more detailed discussion regarding the concepts of test validity, see Cook and Campbell, 1979; Jensen, 1981; Lemke and Wiersma, 1976; and Mehrens and Lehmann, 1973.

To be valid, a test must first be reliable, a topic which is discussed under the second question in this section. It should be pointed out that the relationship between reliability and validity is such that test validity cannot exceed the square root of the reliability; hence, extremely high validity coefficients are not generally found in the literature, while very high reliability coefficients often are. We now turn to the second question, which regards reliability.

Second, does the test have adequate reliability, that is, does it measure in a consistent fashion? Reliability measures often used for norm-referenced tests include internal consistency, test-retest, and alternate-forms reliability. Reliability is usually easier to document than validity, hence, adequate reliability is usually in the .80s or .90s for norm-referenced tests. Methods for calculating the reliability of criterion-referenced tests have been investigated by measurement specialists with somewhat inconclusive and conflicting results regarding the interpretation of criterion-referenced reliability. Livingston (1972) has proposed a procedure for calculating a reliability coefficient based upon the squared deviations of scores from the performance standard (or cut-off score).

Livingston's coefficient is purportedly analogous to the reliability coefficient for norm-referenced tests in classical test theory which is based upon the squared deviations of scores from the mean. However, several noted psychometricians, including Harris (1972), Shavelson, Block, and Ravitch (1972), and Hambleton and Novick (1973), have not attributed much merit to the Livingston technique, going so far as to assert that Livingston's reliability statistic should more appropriately be labeled under some other name because it deviates so much from the conventional concept of reliability. Because of the lack of consensus regarding one appropriate model, it may be concluded that more research is needed to advance the cause of calculating reliability as applied to criterion-referenced testing.

Third, does the main intended purpose of the test match the use to which we want to put the test? If the test is designed to determine the overall education level of soldiers, but we want to use the test for specific diagnostic purposes for designing a remedial program, then the test's purpose does not match our use.

Fourth, does the intended test population (and the norm group on which the test is based) match our own population? This is especially important in the Army setting, where the target group is adult soldiers. Reading tests used by the Army must be intended for adult use and must have adult norms. Better yet, the norms should include Army personnel of the type with whom we are concerned for our particular purpose. If we are concerned with non-commissioned officers (NCOs), norms should include NCOs. If we are concerned with West Point cadets, norms should include cadets. Otherwise, there is a high potential for test bias, as discussed in a later section. If norms are not appropriate, it is possible to develop relevant norms through our own research. This takes time and effort but may be worth it.

Fifth, are the appropriate scores--percentiles, stanines, reading grade levels, standard scores--available? We must determine how we want to express the test results and determine what tests provide those necessary scores. A detailed discussion of score types is offered in sections nine and ten of this report.

Sixth, is the test easy to administer in terms of group vs. individual testing, hand vs. machine scoring, and timing? We need to know our requirements in these areas. Cost, of course, is a factor, but it is often not as great an issue as test administration.

Seventh, are there alternate forms of the test? Do these forms correlate highly enough with each other so that they are interchangeable? This is important for test security reasons and for a number of other reasons, such as pretesting and posttesting for the purpose of assessing achievement gain.

How Does Test Validity Relate to Test Bias?

The misuse of tests resulting from test bias is essentially a test validity issue. In testing circles, test bias is frequently referred to as differential validity. Test validity can be threatened and bias against particular subgroups thereby can be introduced when, for example, that subgroup is not represented in the group upon which the test was normed. Likewise, when people are being tested upon items in a subject matter area in which certain subgroups have been systematically unexposed, bias enters. Differences in familiarity with language can bias test results. When used properly, valid and reliable tests provide evidence of achievement independent of subjective judgement (Schultz & Fortune, 1981).

One of the important measures of the quality of a test is its power to discriminate. When bias causes a test to discriminate across subgroup membership rather than across knowledge of the material being tested, social injustice occurs. It must be recalled that the word "discriminate" possess two distinct connotations by definition as cited by Mehrens and Lehmann (1973, p. 668):

(1) "to make a distinction; to use good judgement;" and

(2) "to make a difference in treatment or favor on a basis other than individual merit."

It is the usage of tests in association with the second definition that gives rise to bias in testing, particularly as related to cultural fairness, and causes grave concern to test users and test takers alike.

Several court cases have demonstrated the significance of using valid tests. Decisions in the cases of Hobson v. Hanson, 1967; Diana v. California Board of Education, 1970; and Griggs v. Duke Power Co., 1971 carried with them the message that use of construct, content and predictive/criterion valid tests must be the rule in order to guard against test bias and discriminatory practices (Schultz & Fortune, 1981).

Regarding the issue of content validity and potential bias of a test, Ebel (1975, p. 86) stated:

> If the test items sample representatively the areas of knowledge and skill that constitute competence in the ability being tested, then the test possesses content validity for all examinees regardless of cultural, ethnic, or racial origins. Language or experience differences that handicap the minority examinee in his attempt to demonstrate knowledge and skill on the test are likely to handicap him also in his attempt to

utilize that knowledge and skill in other situations. When this happens, the test cannot be said to be biased against a member of a minority. Rather, it reflects quite accurately the usable competence he possesses.

Concerning the argument that educational tests are biased against minorities because the tests reflect middle-class values, Ebel comments in the same article, that because tests involve essentially cognitive tasks, items are not culturally specific, but rather reflect what is valued by the common culture. As such, Ebel attributes the test bias problem to a more profound and serious underlying issue asserting that, "The bias which accounts for poor test performance by some minority persons is not in the tests so much as it is in the culture, and thus is another problem altogether" (p. 87).

Jensen (1981, p. 137) defines a biased test as one which "yields scores that mean something different for persons of one group than for persons of another group, even when two persons from different groups have identical scores on the test." Regarding predictive bias, which involves the scores that are used as a predictive index based on the test's validity coefficient for the criterion of concern, Jensen (p. 141) states that:

> Prediction is biased if persons from different populations (e.g., blacks and whites) who obtained the same test score do not, on the average, perform the same on the criterion. In other words, bias exists if one and the same test score actually predicts different levels of criterion performance, depending on the person's group membership.

That is, there will be consistent under- or overestimation of predictions for subgroup population members.

Thus far the focus has been concentrated on areas of potential bias relative to tests in general. However, discussion of the biasing effects of criterion-referenced tests versus norm-referenced tests is pertinent. Many test and measurement specialists view CRTs more favorably than NRTs with respect to the issue of bias, particularly when associated with culture-fair assessment, because it is the nature of NRTs to pit individual against individual or group against group, whereas CRTs compare an individual or group to a standard (criterion) which ostensibly decreases the probability of making inequitable comparisons (see Womer, 1976; Barnes, n.d.; and Drew, 1973 as cited in Oxford, pp. 10-11). However, CRTs cannot be regarded as the panacea for countering the biasing aspects of NRTs. For example, the cultural fairness of CRTs depends upon who determines the performance standards, what those standards include, how test items are phrased, and what the content of the items is (Oxford, 1976). Thus, it is important to obtain input from members of all subgroups representing the prospective examinee population concerning objectives and content development during a consensus-reaching process in order to reduce potential bias in CRTs.

It may therefore be concluded that tests _per se_, irrespective of design, are not inappropriate educational tools; indeed, they provide a means of objective assessment. However, because tests are used in such important decision-making activities, e.g., diagnostic and selection processes, it is imperative that testing purposes be clearly identified and that well-defined guidelines or standards of utilization serve to promote the enlightened and ethical construction and use of tests among the user population in order to alleviate misuse and instances of bias. The state-of-the-art has progressed to the point that the technology to permit this is within grasp.

What Adult Reading Tests are Available from Commercial Sources?

A quick survey of adult reading and general achievement tests from commercial sources indicates that the quality of such tests is quite spotty, and many of them have questionable relevance to Army reading assessment. Table 2 provides a review (based on Buros, 1972, 1975 and many other sources) of a number of tests that are either (a) designed specifically for adult use or (b) though designed for children have been used for adult testing by the military services. Most of these tests are norm-referenced.

From a purely technical standpoint, the Adult Basic Learning Examination (ABLE) appears to be the best commercially available reading test for Army use. The ABLE has well documented reliability and validity, as indicated in Table 2, although Army-related predictive validity would be helpful. Unlike most commercially available reading tests, the ABLE offers adult norms based on Job Corps personnel and adult basic education students. The ABLE is well accepted in the testing community. However, as noted by Raines (1983), the ABLE was not popular for use in the Army's Basic Skills Education Program (BSEP) program for a number of reasons unrelated to technical quality. Some Army personnel felt that ABLE results were not specific enough for diagnosis of educational weaknesses of individual soldiers--which is understandable, since it is a norm-referenced test designed to assess general educational level and not specific strengths and weaknesses. Furthermore, the ABLE sometimes failed to detect soldiers needing remedial instruction, was considered by some Army instructors to be inconvenient to administer due to an audiotaped portion, and does not offer grade equivalent scores for its highest test level (Raines, 1983). However, it should be noted that the reading subtests of the ABLE can be ad-

TABLE 2

COMMERCIAL ADULT READING TEST INFORMATION

TEST SERIES	DEVELOPER OR PUBLISHER	LATEST COPYRIGHT (1)	MAIN INTENDED PURPOSE(S)	INTENDED TEST POPULATION	SCORES	VALIDITY	RELIABILITY	NORMS	ADMINISTRATION: GP vs. INDIV	ADMINISTRATION: TIMING	ADMINISTRATION: SCORING	FORMS	SUBTESTS OR SCORE AREAS	ITEM QUALITY	ARMY RELEVANCE
Adult Basic Learning Exam (ABLE)[2] AND	Psychological Corporation		To determine general education level of adults; to evaluate adult education programs	Adults with achievement levels grades 1-4, 5-8, 9-12 adults who have not completed formal 8th grade education.	Grade equivalents, based on equating ABLE to Stanford Achievement Test (SAT) using scores of students in grades 2-7; authors encourage use of local norms (percentiles & stanines).	Concurrent validity low to moderate based on: (1) administration of ABLE & SAT to elementary & junior high students (.60-.76); (2) administration of ABLE & SAT paragraph meaning to job corps group (.38-.72).	Split-half reliability good to excellent for school group (.87-.95), job corps group (.85-.96), & adult basic education students (.91-.94). No information on test-retest or other reliability.	Based on 6,000 elementary & junior high students, 800 job corps members, & 450 adult basic education students.	Group	Level 1: 145 min. Level 2: 145 min. Level 3: 207 min.[2]	Hand or Machine	A & B	Vocabulary, reading, spelling arithmetic.	Items appear technically well constructed. Most items relevant to adults. Reading items are cloze types & do not measure all facets of reading-just reading comprehension.	Useful for Army. Was used in BSEP screening; superceded by TABE.
SelectABLE	Psychological Corporation	1974	To screen for determination of which ABLE level to administer.	Same as for ABLE.	Raw score, percent correct	INA[1]	KR-20 .80-.87; split-half .86-.91.	152 military recruits	Group	15 min.	Hand	One Form	Total	Adequate	Was used as preliminary screener for ABLE in Army.
Carver-Darby Chunked Reading Test	Revrac Publications (developed by American Institutes for Research)	1972	To measure information stored during reading.	Grades 9-16 and adults.	Efficiency, accuracy, and rate scores; no standardized scores.	Validity best viewed as pilot study.	Reliability best viewed as pilot study-not high (.43-.81)	Totally unsatisfactory; means and standard deviations available on 41 "college student volunteers."	Group	25 min optional	Machine	A & B	Efficiency, accuracy, rate.	Divides sentences into chunks but without a consistent theoretical basis related to the reading act.	Chunking may not be relevant to skills needed in most Army jobs.
Gates-MacGinitie Reading Tests	Houghton Mifflin	1978	To assess three areas of reading.	Grades 1-12	Raw score, extended standard score, percentiles.	Content validity suitable; may have high verbal-IQ component (has high correlation with large Thorndike verbal IQ).	Alternate form reliability (.72-.90)	Norming procedures good.	Group	Level A, B-40 min Level C 50 min Level CS 7 min Level D 40 min Level E 40 min	Hand or Machine	1,2,3	Speed/accuracy, vocabulary, comprehension.	Adequate	Has been used in the Navy, though not intended for adults. Not relevant to Army as an adult test.
Gray Oral Reading Test	Bobbs-Merrill	1967	To assess oral reading skill.	Grades 1-16 and adult.	Grade equivalents (total score only)	INA[1]	Standard errors of measurement adequate.	Norms based on very small samples (N<10 per grade).	INA[1]	INA[1]	INA[1]	A,B,C,D	Comprehension, total	Items satisfactory to good; oral reading only.	Oral reading not often relevant to Army jobs.
Metropolitan Achievement Test[2]	Psychological Corporation	1978	To assess achievement in a number of skill areas.	Grades 2.5-9.5 for reading subtest.	Scaled score, percentiles, stanines, grade equivalents.	Content validity good, no predictive or other validity mentioned in reviews; test authors suggest local curriculum validity be checked.	Internal consistency reliability .90+, standard errors of measurement excellent.	Norm group carefully selected.	Group	Level P-2: 160-170 mins; Level E: 225-238 mins; Level I: 270 mins; Level A: 265 min[3]	Hand or Machine	J & K	Word knowledge, reading, total word analysis (or language, depending on test level), spelling, mathematics (computation, concepts problem solving, total).	Items written for children not adults. Content problems because content remote from adult life.	An earlier form of this test has been used as USAFI Reading Test, although the content was designed for children not adults.

TABLE 2

COMMERICAL ADULT READING TEST INFORMATION (Continued)

TEST SERIES	DEVELOPER OR PUBLISHER	LATEST COPYRIGHT	MAIN INTENDED PURPOSE(S)	INTENDED TEST POPULATION	SCORES	VALIDITY	RELIABILITY	NORMS	ADMINISTRATION GP vs. INDIV	ADMINISTRATION TIMING	ADMINISTRATION SCORING	FORMS	SUBTESTS OR SCORE AREAS	ITEM QUALITY	ARMY RELEVANCE
Nelson-Denny Reading Test	Houghton-Mifflin	1973	To assess reading for college placement and adult reading classes.	Grades 9-16 and adults.	Percentiles, grade equivalents.	Little data on concurrent and predictive validity; nothing on content or construct validity.	High alternate forms reliability for most subtests; based on small samples, however.	Standardization OK; adult norms based on cut time administration only.	Group	30-35 min.	Hand or Machine	A & B	Vocabulary, comprehension, total, rate.	Mostly well constructed items.	Has been used in the Navy; potentially usable for the Army.
RBH Basic Reading and Word Test	Richardson, Bellows, Henry & Company	1969	To assess reading comprehension & vocabulary.	Disadvantaged adults.	Percentiles, standard scores.	Lacks convincing validity data; no statistics.	Lacks convincing reliability data (.86 probably an overestimate)	INA[1]	Group	25-30 min	Hand	One form	Total	Items cover mostly vocabulary. Choice of vocabulary tested is questionable.	Vocabulary tested probably not relevant to Army jobs.
RBH Test of Reading Comprehension	Richardson, Bellows, Henry & Company	1963	To test reading comprehension using six articles related to business & industry.	Business and industry.	Percentiles, standard scores.	Lacks convincing validity data; low validity coefficients (.17-.46)	Reliability data not adequate; no means & standard deviations given.	Test probably not too easy for norm groups; norms not adequately standardized.	Group	20-25 min.	Hand	One form	Total	INA[1]	Test quality not adequate for Army use.
Tests of Adult Basic Education (TABE)	CTB/McGraw-Hill	1976	To measure reading and arithmetic levels of adults using an adapted version of a children's test.	Adults reading at levels of children in grades 2-4 (Level E), 4-6 (Level M), 7-9 (Level D).	Grade equivalents based on California Achievement Tests (CAT), 1970.	Content validity based only on item selection procedures from CAT. Questionable validity. Concurrent validity .56 with GED tests.	KR-20 for reading test section .86-.94; test-retest for reading .79-.85.	No adult norms; norms based on students in grades 2-9.	Group	Level E: 127 min. Level M: 209 min. Level D: 191 min.[3]	Hand	3 & 4	Reading (vocabulary, comprehension, total). Arithmetic (reasoning, fundamentals, total). Language (mechanics, spelling, total) - for top two levels. Total (Locator test also available).	Items were written for children but have been revised to omit purely childish references.	More validity information and adult norms would make this test more relevant and useful for the Army.
Wide Range Achievement Test Level II[2]	Jastak Assessment Systems	1978	To assess quickly three discrete areas of achievement.	12 years to adult.	Standard scores, grade equivalents, percentiles based on age.	Content validity questionable.	High reported reliabilities (.90) are suspect.	No national norming sample. Norms developed from continuous age data. Identity and nature of norm group not clear.	Part Individual.	20-30 min.	Hand	One form	Spelling, arithmetic, reading (based on "clinical factor analysis").	Questionable item development and quality.	This test has been used in the Navy. Questionable test quality makes use of test in Army dubious.

1 Information not available
2 Review concerns reading-related subtests only (e.g., reading, vocabulary, spelling)
3 Timing for whole test

ministered separately from the other ABLE subtests without harming validity and reliability--which are assessed for the subtests as well as the whole test.

The Test of Adult Basic Education (TABE) replaced the ABLE for Army BSEP screening and student progress evaluation in 1979. The TABE was chosen on practical, not technical, grounds after an intensive examination of a number of available tests, although at the time the Army did not have documentation on the reliability or the validity of the TABE. The TABE is actually a new version of the California Achievement Test (CAT), which was designed for children. To develop the TABE, patently childish references were simply deleted from the CAT or replaced by adult references, so the change was more cosmetic than content-based. The CAT itself has adequately documented reliability and validity. However, as the TABE development project leader, Dr. Pat Lister, stated (personal communication, September 8, 1983), the idea of "inherited" reliability and validity from an earlier test--such as the CAT--to a later test which has a different purpose and a different target group--such as the TABE--is fallacious. The TABE publisher later provided studies which demonstrated a high degree of reliability for the TABE and moderate though not strong evidence of validity, as shown in Table 2. The TABE lacks adult norms of any kind, so its use is limited to those Army programs that find it useful to compare soldiers' scores to children's norms. Clearly, key career decisions for soldiers should not be made on the basis of children's norms, as pointed out by statistician Dr. George Burket (personal communication, September 8, 1983), who works for the TABE publisher. For key career decisions such as reenlistment, a test like the TABE must be supplemented by adult norms, preferably Army-based ones, and should have a greater degree of demonstrated validity. Army research could produce relevant norms and Army-related validity data to make the TABE, or other tests of like characteristics, more useful. Despite Bachem's (1982) strong criticisms of the TABE due to lack of adult norms, use of grade equivalents for adults, and alleged "middle-class elementary school" orientation, Army BSEP managers claim that it is easier to administer than the ABLE and provides better diagnostic information as well (Raines, 1983; also Raines, personal communication, September 6, 1983).

Many commercial reading tests and achievement batteries were developed for children but have been used by the military for adults. These include the Gates-McGinitie and the Metropolitan Achievement Test, among others. The use of a child-based test for military personnel can be questioned because of inappropriate norms and potentially inappropriate content.

The advantage of many commercial tests is that they offer the kind of subtests and scores that the Army believes it wants. For example, many commercial tests provide reading comprehension and vocabulary subtests and grade equivalent or grade level scores. Persons who select Army reading tests should consider whether a given commercial test really offers what the Army needs, according to the characteristics outlined in an earlier section. We must look beneath the glossy veneer of commercial reading tests in order to determine which ones best serve Army requirements.

Moreover, for any test, commercial or not, to be used for military career decisions such as reenlistment, it is important that the test be shown to be related to job performance. Decisions affecting a person's career should not be made based on a test that has not been shown in some fashion to be related to how the person performs on the job. Reading ability would logically seem to relate to how well the individual performs in his or her military occupational specialty (MOS), and indeed informal observations appear to support this contention. However, it is important to verify such assumptions and informal observations, preferably by correlating reading scores with job performance evaluations or ratings for an adequate sample of soldiers. If the correlation is high, then the Army could consider itself to be in a good position to make personnel decisions based on reading scores. In this way, the Army would have some evidence that the test was valid for use as a criterion for such decisions.

It may turn out that for some specific Army purposes noncommercial, militarily developed reading tests are more useful than commercially developed reading tests. The next section discusses reading tests available from military sources.

What Adult Reading Tests are Available from Military Sources?

From time to time the armed services have entered the business of reading test development, as witnessed by Table 3. Throughout the seventies and into the early eighties the Army worked on a series of job-related reading tests (see Claudy & Caylor, 1982; Sticht, 1975, 1982; Sticht, Hooke, & Caylor, 1982; Sticht, Caylor, & James, 1978; Sticht, Caylor, Kern, & Fox, 1971).

The earlier versions of Army reading tests, known as Job Reading Task Tests (JRTT), circa 1971-1973, had moderate validity and reliability and were available in three forms--for cooks, vehicle repairmen, and supply clerks. They covered such reading tasks as tables, standards and specifications, identification and description, procedural directions, procedural check points, and functional description and were clearly aimed at Army enlisted personnel. The tests appear to be more objective-referenced than

TABLE 3

MILITARY READING TEST INFORMATION

TEST SERIES	DEVELOPER OR PUBLISHER	LATEST COPYRIGHT	MAIN INTENDED PURPOSE(S)	INTENDED TEST POPULATION	SCORES	VALIDITY	RELIABILITY	NORMS	ADMINISTRATION: GP vs. INDIV	ADMINISTRATION: TIMING	ADMINISTRATION: SCORING	FORM[3]	SUBTESTS OR SCORE AREAS	ITEM QUALITY	ARMY RELEVANCE
Air Force Reading Ability Test (AFRAT)[1]	U.S. Air Force	INA[2] 1 Dec 1981	To assess reading ability among fair to excellent readers in the Air Force.	5th grade through college level; not intended for poor readers.	GE, percentile.	Concurrent validity is .72 with California Achvt. Test, .75 with Nelson-Denny. Predictive validity low to moderate (.40 in predicting Officer Training School grades, -.13-.61 in predicting technical training grades). Comprehension is better predictor than vocabulary.	Internal consistency reliability approx. .90 for whole test.	Adult male & female Air Force & Army enlistees.	Group	50 min. total test.	Hand or Machine	A,B	Comprehension, vocabulary, total.	Vocabulary and comprehension items above average in quality.	Relevance for Army officers not known. Shows little relevance for Army enlisted personnel.
Job Reading Task Tests[4]	U.S. Army	circa 1971-1973	To assess performance on Army job reading tasks with items based on 3 military occupational specialties (MOS); for research purposes only.	Army enlisted personnel	Percent correct, reading grade level.	Moderate (.65-.80) concurrent validity with standardized achvt. test. Lower concurrent validity with AFQT. Moderate correlations with course grades and job knowledge tests.	Test-retest reliability .74-.80. No other reliability information.	INA[2]	Group	INA[2]	INA[2]	Three forms: cook, supply clerk, vehicle repairman.	Tables, standards & specifications, identification & description, procedural directions, check points, functional description.	INA[2]	Highly relevant for Army use but needs more reliability data; may need MOS content updated.

1-Information from Mathews (personal communication, September 7, 1983) and Valentine (n.d.)
2-Information not available
3-One form assumed unless more than one is specified
4-Information from Sticht (1975)
5-Information from Claudy & Caylor (1982)

TABLE 3

MILITARY READING TEST INFORMATION (Continued)

TEST SERIES	DEVELOPER OR PUBLISHER	LATEST COPYRIGHT	MAIN INTENDED PURPOSE(S)	INTENDED TEST POPULATION	SCORES	VALIDITY	RELIABILITY	NORMS	ADMINISTRATION: GP vs. INDIV	ADMINISTRATION: TIMING	ADMINISTRATION: SCORING	FORMS	SUBTESTS OR SCORE AREAS	ITEM QUALITY	ARMY RELEVANCE
Job Reading Task Tests[4]	U.S. Army	circa 1975	To assess performance on Army job reading tasks with items based on 6 MPS.	Army enlisted personnel.	GE, percentiles	Content validity good; based on [illegible] job reading tasks; concurrent validity [illegible] with USAFI RGL criterion	KR-20: .93-.96; alternate forms: .80-.76; SE of meas: .02-.06.	750 young adult male Army recruits.	Group	1 hr.+	Hand	A,B,C	Using tables to locate information, extracting information from tables and narrative prose, following procedural directions	Free-response, fill-in-blank items cause slow scoring.	Highly relevant for Army use but awkward for administration and scoring.
Job Reading Tests[5]	U.S. Army	circa 1982	To assess performance on Army job reading tasks with items based on 6 MOS.	Army enlisted personnel.	Percentile	Validity data not convincing. Item validity appears adequate. Only information on test validity is that the content is drawn carefully from 6 MOS.	Not available	Norming appears adequate.	Group	30-40 min.	Machine	A,B,C	Locating job information in tables, index, graphs, & narrative; forms completion.	Item statistics appear adequate. Multiple choice items.	Highly relevant for Army use but needs more reliability data; may need MOS content updated.
U.S. Armed Forces Institute (USAFI) Reading Test (same as Metropolitan, see Table 2)			←					SEE TABLE 2							→

1-Information from Mathews (personal communication, September 7, 1983) and Valentine (n.d.)
2-Information not available
3-One form assumed unless more than one is specified
4-Information from Sticht (1975)
5-Information from Claudy & Caylor (1982)

norm- or criterion-referenced. More reliability data are needed on the JRTT, and current use would demand updated content to match changing job requirements.

More recently, the Job Reading Test (JRT) was developed by the Army around 1982 to assess reading performance with items drawn from material in six high-density MOS. The JRT is clearly norm-referenced. While norming appears adequate and machine scoring makes the test more appealing than earlier Army job-related reading tests, the JRT lacks convincing validity and reliability data.

The Air Force developed its own Air Force Reading Ability Test (AFRAT) (Valentine, n.d.) to assess reading ability among fair to excellent readers. It covers fifth grade through college level and is not intended for poor readers. Validities are quite variable (-.13 to .75), while reliability is high (.90). The AFRAT might be relevant to Army officers, although this is not certain. It is likely to be irrelevant to Army enlisted personnel because of the relatively high reading level of the intended test population.

As mentioned earlier, the U.S. Armed Forces Institute (USAFI) has used a commercial reading test as its own military reading test. Form D of the Metropolitan Achievement Test is also known as the USAFI Reading Test. Only very minor changes were made in the "green-suit" transformation of this test from civilian to military.

The Navy has also developed its own reading tests, which are similar to some developed by the other services. Information on Navy reading tests is available through Dr. Frederick Chang and Dr. Thomas Sticht at the Navy Personnel Research Laboratory.

Sticht (personal communication, August 15, 1983), the developer of many of the Army's job-based reading tests, feels that adult reading tests created by the armed forces have never been accepted because of the lack of advocacy within the armed forces themselves. In other words, military or contract psychometricians and military trainers were involved in test development, but there was no one around to champion the use of the new tests after the tests were developed.

It is very clear that some military reading tests are more relevant to the Army than some commercial reading tests that are or have been used by the Army. The General Accounting Office (GAO) has recently recommended that Army training, particularly BSEP, be functionalized so that it relates closely to Army jobs. This is now being done through the Job Skills Education Program (JSEP), which will also have job-related testing. The potential of job-related testing has never been fully tapped in the Army, despite the availability of at least rudimentary versions of useful job-based reading tests. More information needs to be gathered about the psychometric quality of these tests, but their availability for at least some MOS should not be overlooked.

One problem with the use of job-based reading tests is that they currently cover only a small number of MOS. Another problem is that they may not provide the type of scores that some Army personnel feel are necessary, such as reading grade level, i.e., grade equivalent in reading. The strengths and weaknesses of this type of score are discussed in a later section. Suffice it to say here that the utility of grade equivalents is debatable in the military and that interpretation problems exist in using grade equivalents with adults. Before turning to the subject of grade equivalents, we will briefly discuss the degree to which reading tests correlate with the ASVAB, the military's most prominent screening, selection, and placement test.

How Do Specific Reading Tests Correlate with the ASVAB?

The ASVAB has reliabilities in the high .80s and predictive validity (for predicting scores on Skill Qualification Tests) in the .50-.60 range (Hanser, personal communication, September 29, 1983). The Armed Forces Qualifying Test (AFQT), which covers four key ASVAB composites, has been called by Jensen (1981) a test of general ability, and the Army considers it a measure of "trainability." Jensen points out that among high school students who have all had much the same schooling, a standard test of reading comprehension can serve as a good measure of general ability except for the few students with a specific reading disability, also known as dyslexia. Furthermore, a test of general ability can serve as a proxy measure of reading comprehension, although it cannot be said to be a reading test per se. Many tests of general ability or "trainability" are reading-dependent, that is, the examinee must be able to read in order to take the test. Reading skill is necessary but not sufficient to perform well on the test. A person's ability or "trainability" is not separable from his or her reading skill on such a test, unless special statistical techniques are used. Such is the case with the ASVAB.

There have been many studies indicating the degree to which the ASVAB and various reading tests relate to each other. For example, Sticht (1975, p. 36) shows as .65 correlation between an unspecified reading achievement test and the Armed Forces Qualifying Test (AFQT), which covers four key ASVAB composites. This is a moderate but not high correlation. The Job Reading Task Tests (JRTT) have a moderate concurrent validity with the AFQT (Sticht, 1975, p. 43). Although there is no published information on the correlation between the Air Force test (AFRAT) and either the AFQT or the ASVAB, Valentine (n.d., p. 7) reported that the AFRAT contributed significantly to the AFQT in predicting Air Force training grades. Fischl (1981) found that the USAFI Reading Test, which as we have seen is the same as the Metropolitan, was correlated moderately to highly (.80-.95) with various composites of the ASVAB and with the total ASVAB for a sample of 600 soldiers. In a recent investigation involving 2,385 Army and Marine recruits Grafton (personal communication, August 16, 1983) discovered a .85 correlation between the ABLE and the General Technical composite of the ASVAB. Among several groups of limited English proficient soldiers headed for Army English-as-a-second-language training, Oxford-Carpenter (1982) found very low correlations--in the teens and twenties--between the ABLE and a number of ASVAB composites. This result was undoubtedly due to the language problem. After ESL instruction the correlation between ABLE and ASVAB scores rose to a higher level--the forties and fifties.

In summary, reading tests are usually moderately to highly related to the ASVAB as long as the language problems (for limited English proficient soldiers and others) do not intervene. There is clearly a large verbal component in the ASVAB, as noted by Valentine (n.d., p. 2), and some have conjectured that the ASVAB already indirectly measures reading ability.

The question arose in 1977 as to whether the ASVAB could be used to determine the reading ability of applicants for the armed forces (see Valentine, p. 2). The Principal Deputy Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics) wondered whether the ASVAB could screen out applicants with low reading skills and whether a reading grade index could be derived from the ASVAB for remedial instruction purposes. Several years later Frances Grafton of the Army Research Institute developed something like this index through equating of the ASVAB-GT with a composite of three "literacy" subtests (reading, vocabulary, and arithmetic reasoning) in the ABLE.

A 1977 study reported by Valentine found that while there is a relatively high correlation between reading grade level and enlistment qualification on the ASVAB, enlistment standards do not screen out all of the poor readers. Slightly over 7% of the "qualified" applicants were poor readers. However, as noted by Grafton (personal communication, September 2, 1983), this result is explainable by the fact that the ASVAB was misnormed for several years before the error was discovered in 1980. Poor readers were allowed to enter the Army under the incorrect ASVAB norms, but in 1981 that problem was largely alleviated by correction of the norms. Valentine found that different reading grade levels were found for applicants using two different tests, the Nelson-Denny and the Gates-McGinitie. "For any given reading grade level in their region of overlap, one would predict a slightly higher AFQT score from the Gates-McGinitie than from the Nelson-Denny" (Valentine, p. 4). The 7% who were poor readers under old, incorrect norms are now in mid-career and have caused some concern about reenlistment.

The possibility of using a commercial reading test as a reenlistment criterion was thoroughly considered by the Office of the Deputy Chief of Staff for Personnel (ODCSPER) in 1983. Main tests considered included the TABE and the ABLE. After much consideration regarding the merits of these tests, ODCSPER decided to use ASVAB-GT along with other non-reading measures for reenlistment. As noted, a type of general literacy index including reading ability as a component has been calculated for the ASVAB-GT through equating it with the ABLE subtests.

Now it is time to move from discussions of ASVAB and reading test correlations to the general question of reading test scores. We will now turn to the most prevalent score in the military: the grade equivalent or grade level score.

What are the Uses and Abuses of Grade Equivalent Scores?

Grade equivalent or GE scores, popularly known in the military as reading grade levels when applied to reading tests, are the most used and abused of all score types. They are naturally norm-referenced, because they compare the individual's performance to the typical performance of a norm group. Everyone thinks he or she knows what a grade equivalent is, because the score appears to be so simple. Most of us are deceived by the apparent simplicity.

In this discussion, we will use the acronym GE score to designate grade equivalents in any subject matter area. A GE score is defined as the average or median score a norm group attains when tested at a particular grade and month. GE scores are expressed in two numbers. The first number represents the grade in school, and the second number represents the month of participation within the grade. For example, a GE score of 3.5 represents the average score that a norm group attained on a given set of test items when tested in the third grade, fifth month of school. Since test publishers assume that little or no academic learning occurs during the summer, the decimal number for months ranges from .0 to .9. If a second-grader attains a GE of 3.5 on a reading test designed for grades 1-2, it could be said that he or she can do second grade reading as well as the average fifth-month third grader can do second grade reading.

Several cautions must be noted in using GE scores with children of school age. First, it should never be assumed that a student is ready to do work at the level indicated by his or her GE score. Therefore, placement decisions are very shaky when made using GE scores. Second, GE scores do not represent equal intervals of growth across grade levels and across subject areas within a given grade. Since GE scores are supposedly but not actually based on equal intervals, caution must be exercised in averaging GE scores to provide meaningful interpretation of an average GE for an entire group. Likewise, it is also dangerous to subtract GE scores to obtain growth indices. Third, GE scores are useful for comparing a set of students with a norm group, but there are limitations on its usefulness for evaluating the impact of an instructional program. Fourth, GE scores do not represent desirable performance; they reflect only typical performance of a particular norm group at a particular time. GE scores ought not be interpreted as performance standards that should be met.

What about use of GE scores for assessing adult performance? All the cautions above are relevant, but additional caveats must be added when GE scores are used with adults. As noted by well-respected psychometricians Ysseldyke and Marston (1982), GE scores are not applicable to high school or adult levels because they have no meaning past the earlier years of constant growth; and furthermore, GE scores fail to compare adults with the most appropriate reference group, their peers. Instead, in using GE scores with adults, comparisons are made between adults and children. Bachem (1982, p. 49), in commenting upon the inappropriate use of GE scores in the military issued a strong statement, "The use of elementary school grade levels to categorize adult combat soldiers seems little short of an insult, no matter how desperate their need for remedial work may be." Sticht, Caylor, and James (1978) found empirically that grade equivalents mean different things when applied to adults than when applied to children. What does it mean to say that an Army soldier has a GE or reading grade level of 6.2 on a test intended for grades 6-9? It means that the soldier performed on the test as well as the typical second-month sixth grader would have performed on the same test, keeping in mind that the content of the test is based on children's experience and covers content in the 6-9 grade range. Few tests that offer GE scores have adult test content. A notable exception is the ABLE. Even with a test such as the ABLE which has adult test content, it is difficult to know how to interpret the meaning of the GE score. Angoff (1971), Ebel (1972), and Lemke and Wiersma (1976) may be consulted regarding additional in-depth treatment of the GE score topic.

The GAO in 1977 recommended to the Secretary of Defense that the Department of Defense develop a policy to have the services determine the reading grade level required for each military occupation and establish an overall minimum level required for enlistment (Valentine, n.d., p. 1). This only served to emphasize and promote the use of reading grade levels, although the GAO recommendation has not been implemented. Instead, basic skills programs were used to aid in remediating weaknesses in reading, mathematics, and language. Official regulations list specific reading grade levels as criteria for entry into the Army basic skills program.

Reading grade levels or GE scores are so much a part of the military system that at this point it might be foolhardy to throw them out entirely, even though they are badly misunderstood and overinterpreted by most military and civilian personnel. Therefore, we should consider other types of scores to use along with, not in place of, GE scores in order to provide greater meaning to soldiers' tested reading performance. Perhaps GE scores could be slowly phased out over time by wise military leadership, but for now the best we can do is to supplement them with more meaningful scores.

What Other Scores Are Potentially Useful?

Scores other than GE or reading grade levels are abundant, and many of them are potentially useful for reading assessment in the Army and the other services. We will discuss here a number of possible score types.

A percentile rank, commonly though somewhat inaccurately called a percentile, indicates the percentage of persons in a particular norm group who obtain lower scores than a particular individual. If PFC Adams earns a percentage of 70 on a test, it means that she scored better than or as well as 70% of the individuals in the norm group. If PFC Smith has a percentile of 65, he has scored better than or as well as 65% of the people in the norm group. The percentile distribution is by definition rectangular. It is not a bell-shaped, normal curve. Percentile distributions are obtained by dividing the bell-shaped, raw-score distribution into one hundred groups of equal frequency. Persons near the middle of the raw-score group are spread apart, while persons at the ends are squeezed together. Often, a large percentile difference near the median represents a small raw-score difference in performance. Since different percentile intervals do not represent equal increments of performance, the percentile ranks of a group of individuals cannot be averaged meaningfully to provide an average percentile rank for that group. Likewise, the utility of subtracting one percentile from another (to obtain growth scores, for example) is questionable. However, percentiles are immediately understandable by most people and hence represent an improvement over grade equivalent scores in terms of interpretability.

A second type of score that holds promise for reading assessment in the Army is the standard score. A standard score indicates how many standard deviations a particular raw score is above or below the mean. IQ scores and College Board scores are special versions of standard scores. Some standard scores are called T-scores or z-scores depending on the metric used.

Standard scores are frequently used because they provide comparability of observations obtained by different procedures or by different tests. For instance, it would be possible to compare directly a person's standard score on the ASVAB with his or her standard score on the Nelson-Denny Reading Test or any other test in any subject area. An additional feature is that standard scores can be averaged, added, or subtracted without the interpretive problems of other types of scores.

Stanines are a third type of potentially useful score. Stanines were in fact developed by the armed forces and have gained widespread use in the civilian sector. Stanine scores are normalized standard scores with a range of 1 to 9 and an average value of 5. Like percentile ranks, they are status scores within a particular norm group. Percentile ranks and standard scores roughly correspond as follows:

Percentile	Stanine
96 and over	9
89-95	8
77-88	7
60-76	6
40-59	5
23-39	4
11-22	3
4-10	2
Below 4	1

One advantage of stanines is that if a person is in stanine 9, we immediately understand that the person has performed in a superior fashion. A low stanine would indicate the opposite. The interpretability factor strongly favors stanines, although it is not meaningful to average them. For further discussion pertaining to scaling, standard scores, and transformations see Angoff (1971), Ebel (1972), and Lemke and Wiersma (1976).

In addition to norm-referenced scores such as percentiles, standard scores, and stanines, criterion-referenced scores may have many potential uses for reading assessment in the Army. The Army training community has already discovered the utility of GO/NO-GO testing, which is clearly criterion-referenced. Other criterion-referenced scores that are useful include number of objectives passed or mastered, trials to mastery, and time to mastery. While these criterion-referenced scores do not have the sophisticated psychometric properties of some norm-referenced scores, they are highly useful for skill and knowledge assessment in many Army settings. Moreover, progress is being made currently regarding the identification of more sophisticated procedures specifically suited to elevating the technical quality of criterion-referenced measures, especially with respect to reliability, and regarding the establishment of guidelines for properly assessing criterion-referenced tests (see, for example, Hambleton & Eignor, 1978; Hambleton & Novick, 1973; Hambleton & Gorth, 1971; and Livingston, 1972). These endeavors have contributed to the enhancement of the state-of-the-art and the utility of criterion-referenced tests in general.

Another type of score which has potential is the gain score, which addresses measurement of change. With the advent of the many federally funded programs designed to institute educational change and social reform, accountability became a serious issue. Therefore, an appropriate methodology for determining those programs which demonstrated the most impact--in other words, did the most good per investment of tax dollar--had to be identified or developed. It was this concern regarding program effects on participants that primarily prompted investigation into the merits and applicability of gain score techniques as a measure of change.

Reasons for measuring change may be categorized into three classifications. Gain scores thus may be helpful when we want to:

1. identify individuals who made specific levels of change;

2. identify correlates of change;

3. compare amount of change across groups (Fortune, 1981).

A gain score may be grossly defined as the pretest score subtracted from the posttest score, i.e., a difference attributed to change in educational achievement. Such a simple measure based upon the discrepancy between the beginning and ending status of an individual is not desirable, however, because of the presence of measurement error which contributes to low reliability.

Several approaches have been offered by various psychometricians as a means of overcoming this problem. These approaches include residual or adjusted gain-score models (DuBois, 1957; Manning & DuBois, 1962) and the true score adjustment model (Lord, 1956, 1958, 1960, 1967, 1969). Residual gain-score models employ the difference between an actual posttest score and an estimated posttest score where the estimate has been derived from the pretest scores using regression techniques. Lord's true score adjustment model permits the correction of data by estimating true score by partitioning the score into the regression coefficient multipled by the individual's deviation from his or her own group mean and adding the overall group or total population mean. This procedure serves to equalize the error between pretest and posttest measurements.

Perhaps one of the most ingenious solutions to the gain score problem is suggested by Campbell (Cook & Campbell, 1979) in the regression-discontinuity model, which recognizes the need for estimation of change when addressing the comparison of non-equivalent control group designs. This technique views the solution to the gain score problem from a different perspective--the perspective of research design as opposed to the perspective of measurement. This model adjusts for either group or individual differences in order to generate equal comparisons across groups. It focuses on the identification of clusters around different regression lines for the pretest and posttest regression of the treatment and no-treatment groups. If two distinct regression lines emerge, this indicates the presence of a difference in relationships and the treatment has therefore had an effect.

In addition to those models just cited there are many other gain score techniques proposed in the literature, e.g., Rubin, 1977; Porter, 1973; and Linn and Slinde, 1977. An in-depth treatment of the topic of gain scores relative to the measurement of change is not within the scope of this discussion and the sources referenced may serve as a beginning point for the interested reader. Rather, it is the purpose of this discussion to expose the reader to the nature, complexity, and controversy regarding gain scores and to advise that their use should be approached with great care and after extensive consideration.

Not all tests have all types of scores available. In fact, one way to select a test is to determine what types of scores are relevant and then to use that information as one of the important criteria--along with test content, intended test population, reliability, validity, norms, and other factors--for choosing the test.

What Recommendations Can Be Made for Future Army Reading Assessment?

Several clear recommendations can be made for future reading assessment in the Army. First, the Army should recognize the difficulties associated with the easily misunderstood reading grade level and consider the attractiveness of other types of scores to use along with reading grade level. Second, the Army's reading test selection process should give the same or more attention to technical quality as to administrative ease. Third, the Army should take a good look at its many attempts to develop job-based reading tests and consider whether these tests can be revived and/or revised for use in the future. Perhaps something valuable can be gained from all the money, time, and effort poured into the reading test development process by the Army over the years. Fourth, the Army might consider collaborating more closely with the other services in terms of reading assessment. The Army, the Air Force, and the Navy have independently developed their own reading tests in the last few decades. The armed services have important ideas to share with each other on the topic of reading assessment, and perhaps these ideas can be exchanged and consolidated to the benefit of the entire military establishment.

References

American Heritage Dictionary. (1982). Boston, MA: Houghton Mifflin.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement*, 2nd ed. (pp. 508-600). Washington, DC: American Council on Education.

Bachem, M. B. (1982). Test of adult basic education (TABE or not TABE). In *Proceedings, 24th Annual Conference of the Military Testing Association* (pp. 48-51). Alexandria, VA: Defense Technical Information Center.

Baker, E. L. (1974, June). Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. *Educational Technology*, pp. 10-16.

Barnes, cited in Oxford (1976).

Bormuth, J. (1975). Literacy in the classroom. In W. Page (Ed.), *Help for the reading teacher: New directions in research.* Urbana, IL: National Council of Teachers of English.

Buros, O. K. *Seventh Mental Measurements Yearbook* (1972). Highland Park, NJ: Gryphon Press.

Buros, O. K. *Eighth Mental Measurements Yearbook* (1975). Highland Park, NJ: Gryphon Press.

Caylor, J. S., Sticht, T. G., & Fox, L. C. (1975). *Manual - Job reading task test, Forms A, B, C.* Carmel, CA: Human Resources Research Organization.

Claudy, J. G., & Caylor, J. S. (1982). *Development of job-based reading tests* (ARI Technical Report 479). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation - Design and analysis issues for field settings.* Chicago, IL: Rand McNally.

Drew, cited in Oxford (1976).

DuBois, P. H. (1957). *Multivariate correlational analysis.* New York: Harper.

Ebel, R. L. (1975). Educational tests: Valid? biased? useful? *Phi Delta Kappan*, *57*, 83-88.

Ebel, R. L. (1972). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hall.

Fischl, M. A. (1981). Soldier reading ability: The advocacy point of view. In Proceedings, 23rd Annual Conference of the Military Testing Association (pp. 1447-1450). Arlington, VA.

Fisher, A. H., Jr. (1981). Army "new standards" personnel: Relationships between literacy level and indices of military performance (AFHRL-TR-71-12). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Fortune, J. C. (Speaker)(1981, March). The measurement and detection of educational change. Speech presented at the National Institute of Mental Health, Washington, DC.

Gagne, R. M. (1965). The conditions of learning. New York: Holt, Rinehart, & Winston.

Glaser, R., & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.), Educational measurement, 2nd ed. (pp. 625-670). Washington, DC: American Council on Education.

Hambleton, R. K., & Eignor, D. R. (1978a, October). Competency test development, validation, and standard-setting. Paper presented at the AERA Minimum Competency Testing Conference, Washington, DC.

Hambleton, R. K., & Eignor, D. R. (1978b). Guidelines for evaluating criterion-referenced tests and test manuals. Journal of Educational Measurement, 15, 321-327.

Hambleton, R. K., & Gorth, W. P. (1971). Criterion-referenced testing: Issues and applications (Report No. 13). Amherst, MA: University of Massachusetts Center for Educational Research.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 10, 159-170.

Hively, W. (1974, June). Introduction to domain-referenced testing. Educational Technology, pp. 5-10.

Jensen, A. R. (1981). Straight talk about mental tests. New York: Free Press.

Kriewall, T. E. (1972, April). Aspects and applications of criterion-referenced tests. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Langer, J., & Nicolich, M. (1981). Prior knowledge and its relationship to comprehension. Journal of Reading Behavior, 13, 373-379.

Lemke, L., & Wiersma, W. (1976). Principles of psychological measurement. Chicago, IL: Rand McNally.

Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. Review of Educational Research, 47, 121-150.

Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 9, 13-26.

Lord, F. M. (1956). The measurement of growth. Educational and Psychological Measurement, 16, 421-437. (See also Errata, ibid. 1957, 17, 452).

Lord, F. M. (1958). Further problems in the measurement of growth. Educational and Psychological Measurement, 18, 437-54.

Lord, F. M. (1960). Large-samples covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 55, 307-21.

Lord, F. M. (1967a). Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change (pp. 21-38). Madison, WI: University of Wisconsin Press.

Lord, F. M. (1967b). A paradox in the interpretation of group comparison. Psychological Bulletin, 68, 304-5.

Lord, F. M. (1969). Statistical adjustments when comparing preexisting groups. Psychological Bulletin, 72, 336-7.

Lord, F. M. (1971a). The self-scoring flexilevel test. Journal of Educational Measurement, 8, 147-151.

Lord, F. M. (1971b). A theoretical study of two-stage testing. Psychometrika, 36, 227-242.

Lord, F. M. (1977). A broad range tailored test of verbal ability. Applied Psychological Measurement, 1, 95-100.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Manning, W. H., & DuBois, P. H. (1962). Correlational methods in research on human subjects. Perceptual Motor Skills, 15, 283-321.

Mathews, J. J., & Roach, B. W. (1983). Reading abilities tests: Development and norming for Air Force use (Report No. AFHRL-TR-82-26). Brooks Air Force Base, TX: Air Force Human Resource Laboratory.

McBride, J. R. (1979). Adaptive mental testing (Technical Report No. 423). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mehrens, W. A., & Lehmann, I. J. (1973). Measurement and evaluation in education and psychology. New York: Holt, Rinehart & Winston.

Mehrens, W. A., & Lehmann, I. J. (1978). Standardized tests in education. New York: Holt, Rinehart & Winston.

Millman, J. (1974, June). Sampling plans for domain-referenced tests. Educational Technology, pp. 17-21.

Morreau, L. E. (1974, March). The structured analysis and classification of objectives. Educational Technology, pp. 46-48.

Oxford, R. L. (1976). State of the art in criterion-referenced testing and issues concerning use of criterion-referenced tests in evaluating the national ESEA Title I migrant program (USOE Contract No. 300-76-0095). Research Triangle Park, NC: Research Triangle Institute.

Oxford-Carpenter, R. (1982). Data tables for English-as-a-second language. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Popham, W. J. (1978). Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. Journal of Educational Measurement, 6, 1-9.

Porter, A. C. (1973, February). Analysis strategies for some common evaluation paradigms. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Raines, J. (1983). Information paper: Literacy testing for mid-term reenlistment. Alexandria, VA: Education Directorate, Adjutant General's Office, Department of the Army.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. Journal of Educational Statistics, 2, 1-26.

Schultz, L. J., & Fortune, J. C. (1981). The three i's: Sources of test bias. Education, 102, 117-123.

Sension, D. B., & Rabehl, G. J. (1974, June). Test-item domains and instructional accountability. Educational Technology, pp. 22-28.

Sticht, T. G. (Ed.). (1975). Reading for working: A functional literacy anthology. Alexandria, VA: Human Resources Research Organization.

Sticht, T. G. (1982). Basic skills in defense. (HumRRO-PP-3-82). Alexandria, VA: Human Resources Research Organization.

Sticht, R. G., Beck, L. J., Hauke, R. N., Kleiman, G. M., & James, J. H. (1974). Auding and reading: A developmental model. Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., Caylor, J. S., & James, J. H. (1978). Job-related reading tasks: Teaching marginally literate adults to read (HumRRO-PP-10-78). Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., Caylor, J. S., Kern, R. P., & Fox, L. C. (1971). Literacy skill requirements in four military occupational specialties (HumRRO-TR-71-23). Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., Hooke, L. R., & Caylor, J. S. (1982). Literacy, oracy, and vocational aptitude as predictors of attrition and promotion in the armed services (HumRRO-PP-2-82). Alexandria, VA: Human Resources Research Organization.

Urry, V. W. (1977). Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 14, 181-196.

Valentine, L. D., Jr. (n.d.). Development of the Air Force reading ability test. Unpublished manuscript, Air Force Human Resources Laboratory, Brooks Air Force Base, TX.

Valentine, L. D., Jr. (1980). Reading measurement research in the Air Force. In W. S. Sellman (Chair), Innovations in basic skills education for military personnel. Symposium conducted at the 88th annual convention of the American Psychological Association.

Wood, R. (1973). Response-contingent testing. Review of Educational Research, 43, 529-544.

Weiss, D. J. (1975, November). Computerized adaptive ability measurement. Naval Research Reviews, pp. 1-18.

Womer, cited in Oxford (1976).

Ysseldyke, J. E., & Marston, D. (1982). A critical analysis of standardized reading tests. School Psychology Review, 11 (3), 257-266.

V. CRITERION DEVELOPMENT

John P. Campbell (HumRRO)
Michael G. Rumsey (ARI)

A large proportion of our efforts during the first two years of Project A is being devoted to the development of criterion measures for assessing training and job performance. Estimating the prediction parameters for a selection and classification system that must place so many people in such a variety of jobs within such tight time constraints demands the most complete and precise information that we can gather. Consequently, the validation of predictors must be based on reliable, meaningful, and comprehensive criteria. To the extent that the criteria on which our statistical estimates are based lack relevance and are unreliable or deficient, the effectiveness of the classification system will suffer.

Rather than simply pick whatever traditional criterion measures happen to be available we have elected to conceptualize the criterion problem in construct validity terms. The strategy is to begin with a conceptual model of the entire work performance environment, incorporating external and organizational influences on performance, the person and job components of behavior, as well as their interaction, organizational controls which impact on performance measurement, and performance outcomes. A paper describing this model appears at the end of this section. This broad integrative model will provide a context in which a description of the criterion space that we believe will account for a large proportion of the major facets of soldier performance and effectiveness can be developed. We will proceed from this specification of criterion constructs to the development of an integrated set of criterion measures that reflect individual task proficiency, contributions to general organization goals, minimization of human-resource-related costs, and the relative utility of performance across jobs.

Ideally, if our criterion development efforts were to proceed according to the latest doctrine in conceptualizing a domain of latent variables or constructs, we would most likely adopt a structural model as the guiding heuristic (e.g., James, Muliak, & Brett, 1982). That is, the initial focus would be on trying to specify, however imperfectly, the latent variables or constructs that comprise the criterion space, as well as the nature and degree of their interrelationships. The next step would attempt to specify the manifest, or measureable, variables that represent each latent variable and to predict how the manifest variables are interrelated. The relevant issues then become:

(1) How good is our current theory and knowledge about each latent variable and about how they should interrelate (causally or otherwise)? Unfortunately, applied psychology in general knows a lot more about the latent structure of the predictor side than the criterion side.

(2) Are all the relevant latent variables measured by one or more manifest variables? Is there redundancy? Are some constructs unmeasured?

(3) How much do we know about the validity of the manifest variables as a measure of the latent constructs? How much more do we need to find out?

(4) Should two manifest variables be related to one another? If so, is it because they are measures of the same construct(s), different constructs that covary, or different constructs that stand in causal relation to one another?

(5) Can we specify other factors that will determine the relationship between two manifest variables besides their relationship to the latent structure (e.g., unreliability, common method variance, "halo," the implicit theories of performance held by raters, etc.)?

These are not easy questions to answer but they are relevant to building an understanding of criteria and their interrelationships. In the best of all possible worlds, the explication of such a structural model will be an iterative process over the course of such a large project. Although it may never get to a statistically "testable" form in the confirmatory analysis sense, it would be refined on the basis of each new increment of research data, and it would also guide data collection and analysis. Consequently, it is very much a bootstrapping process. Its aim is to maximize our understanding of the criterion space as well as to provide a basis for developing a composite criterion for validation purposes.

The project began with the basic premise that there are three major components to the total criterion (job effectiveness) space: (a) the individual's performance and effectiveness during training, (b) performance on the specific job tasks for which the individual is responsible, and (c) aspects of performance and effectiveness that are not MOS specific but that are a major part of the effectiveness of every enlisted person. Within each of these major domains the task of the project is to explicate the constructs that define it, develop operational measures of these constructs, and combine operational measures into criterion composites that are maximally useful for developing the selection and classification system.

Relative to the above framework, work on each of the three major criterion domains began in earnest during the past 8 months. These efforts are described in the following sections.

Associated Reports

As noted in the Introduction, a relevant report follows.

A SYSTEMIC MODEL*
OF
WORK PERFORMANCE

Leon I. Wetrogan
Darlene M. Olson
Helen M. Sperling

U.S. Army Research Institute
for the Behavioral and Social Sciences
Alexandria, Virginia 22333

The Army has initiated a large-scale, multi-year Soldier Selection, Classification and Utilization Project which has three major components: 1) the development of an improved set of predictor measures, 2) the development of an improved set of performance measures and 3) an examination of the linkage between the two sets of measures. Our focus today is on that portion of the project dedicated to the development of improved performance measures. At the outset, it was recognized that if the project were to be truly successful in terms of developing measures which comprehensively and accurately encompass the performance domain, considerable effort would be required to define the major parameters of this domain. From this recognition came the concept of a performance model, which we shall describe today.

In order to define the approach to be followed in the examination of work performance, a different recognition was required; the recognition that performance does not occur in a vacuum, but is instead an integral component of an organization's human resource management system. Key to the Army's project is the intent that the improved performance measurement system being developed will be the core of an improved personnel management system. Thus, the purpose of this paper is twofold. First, to identify and discuss the components involved in human work performance. Second, to discuss how these components are related to specific human resource management functions.

Paper presented at the 91st Annual Convention of the American Psychological Association in Anaheim, California, August 1983.
All statements expressed in this paper are those of the authors and do not represent the official opinions or policies of the U. S. Army Research Institute or the Department of the Army.

*Paper presented at the 91st Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

Human Resource Management

In order to put the work performance model into perspective, it is necessary to briefly discuss some components of human resource management. These components are recruitment, selection, training, and performance evaluation.

Recruitment

Stated very simply, recruitment involves acquiring the personnel needed to operate an organization. Not only is recruitment one of the first steps in developing personnel policies, but it is one of the most critical for the establishment and growth of an organization. In order to establish a recruitment policy, three questions need answering (Strauss and Sayles, 1960):

1. Whom should we look for, what type of personnel and in what measures?

2. Where should we look? Inside or outside the organization? Or both?

3. What methods should we use to encourage the "right" people to come to the organization seeking employment?

Selection

This function involves picking the right person for the job. Selection results in a prediction that among applicants being considered, the person selected has a higher probability of success on the job than do other candidates. According to Tiffin and McCormick (1965), the selection process involves four steps: a) determining the qualifications or characteristics desirable in candidates for the job, b) obtaining information about candidates related to the desirable characteristics or qualifications, c) comparing the information obtained on candidates to the desirable characteristics and, d) making a prediction about the likelihood of success.

Training

Training is a planned organized effort that is specifically designed to help individuals develop increasing capacities. Tiffin and McCormick (1965) point out that training can serve any of three purposes. First, training can develop knowledge and skills that will be useful to employees in performing their present job or possible future jobs. Second, training might serve to communicate general information about the organization rather than specific job information. Third, training might be intended to change employee attitudes in some way. Whatever the purpose, however, training should be based on sound principles and practices that are conducive to human learning.

Performance Evaluation

Performance evaluation is typically a systematic method of assessing personnel in terms of their work performance. In the present context performance evaluation is not limited to ratings, but includes all information used to assess performance, i.e., productivity, attendance, percent defects, and so on.

Performance evaluation information is generally used for either administrative or self-improvement purposes. As an administrative tool, performance evaluation information is used in decisions related to promotions, transfers, layoffs and discharges. It is also useful for wage and salary administration, as well as training. When used for self-improvement, performance evaluation information serves to communicate to the employee: a) how well he is doing in the eyes of management, and b) his strengths and weaknesses so that improvement can take place.

WORK PERFORMANCE MODEL

In our work with the Army, work performance has been conceptualized as a complex multidimensional process which is determined by a diverse group of individual difference, job, environmental, and organizational factors. Hence, a comprehensive and fluid systems framework has been developed, which defines relevant performance variables and examines methodologies that could be utilized to assess performance dimensions in relation to various designated criteria.

We have expended considerable effort in the development of a model of work performance with the expectation that such a model will serve the following purposes:

1. The model will assist in understanding which patterns of individual differences and contextual/situational factors contribute to observed differences in individual and unit performance.

2. A broad conceptual work performance model will be a foundation from which to direct, plan, and execute research on the measurement of work performance.

3. The model can function as a post-hoc tool to explain serendipitous research findings.

4. Applied research based on the model may assist in the development of better performance criteria to validate selection and classification instruments and training measures.

The actual development of this conceptual model of work performance has focused on the consolidation and integration of empirical research from industrial-organizational psychology, development of research questions which address gaps in the existing psychological knowledge base, and the application of this information from the generic conceptual model to research problems in the military context.

In accordance with a systems approach to work performance, this model building endeavor has identified the following sub-systems and their relationships: 1) external environment, 2) organization, 3) person, 4) job, 5) work performance, 6) performance evaluation system and 7) the decision making process. Figure 1 depicts all of the component sub-systems of the work performance model.

Each of the sub-systems and their components is important in developing a comprehensive understanding of work performance. The external environment profoundly affects the organization. The organization, viewed in terms of its internal macro-environment, its plans and objectives, and its personnel objectives and strategies, defines the context in which performance takes place. The organization defines the job that needs to be performed and selects and trains the people who are to perform it. The person, the job, and their interaction constitute the micro-environment, of which work performance is a product. The performance evaluation system assesses the match between observed performance and defined standards of performance and provides input into the decision making process. The decision making process then impacts upon the organization, the person, the job and ultimately upon performance so that the entire system is viewed as dynamic and self-correcting.

The first sub-system to be discussed is the environment. The environment can be subdivided into the external environment - those factors outside of the organization's direct control which affect the organization's functioning, and the internal environment - those factors within the organization which are directly controllable and which affect its functioning.

External Environment

No organization operates in isolation. It is constantly meeting the challenges posed by the environment within which it exists. The environment places demands on the organization in terms of the goods and services it needs as well as the laws and regulations which protect the public interest. The organization, on the other hand, is dependent on the external environment for satisfying many of its needs such as human, financial and physical.

Figure 2 shows the four categories of variables that are important for understanding the interaction between the organization and the external environment. These variables are: 1) competitors, 2) contractors or suppliers, 3) laws and regulations and 4) the community.

Competitors within the external environment take two forms. The first involves those organizations competing for the limited available resources in the environment. This takes the form of available raw material (such as oil, steel, coal) or human resources, people needed to produce goods and deliver services demanded by society. For the military, the resources most actively sought are men and women who can be trained to help insure the readiness of the Army. The second group of competitors are those organizations offering the same goods and services for use by the external environment. For example, there are over two hundred companies producing microcomputers, each competing for a share of the market. In terms of the Defense community, each of the services needs to enlist talented men and women, and each offers an opportunity for service in defense of the country.

The second major factor in the external environment consists of contractors or suppliers. Both military and nonmilitary organizations are dependent to varying degrees on contractors and suppliers for goods and services needed to accomplish their objectives. The automobile industry is dependent upon the steel industry, the glass industry and the tire industry. The Army is dependent upon contractors for designing and producing weapon systems; air, land and sea vehicles; as well as many other goods and services. To the extent that these external resources are available, the attainment of organizational objectives will be accomplished without major modifications to the internal organization.

The third major factor in the external environment is laws and regulations. Almost all organizational functions are to some extent controlled by laws and regulations. Most laws and regulations work to restrict certain activities of the organization in order to protect society. Examples of such governmental regulations include the child labor law and the Fair Labor Standards Act. Many personnel functions are influenced by such agencies as the Equal Employment Opportunity Commission (EEOC) and the Office of Federal Contract Compliance (OFCC). Laws and regulations are also related to the health and safety of the work place. Many of these regulations are enforced by the Department of Labor and the Occupational Safety and Health Administration (OSHA). Products and services available from private and public organizations are also subject to laws and regulations. Many are administered by agencies such as the Federal Trade Commission (FTC), and Food and Drug Administration (FDA).

The fourth factor in the external environment which impacts on the organization is the community . The community can be viewed as multidimensional. It consists of: a) cultural norms and values, b) economic climate, c) demographic composition of the population and d) international relationships. Cultural norms and values affect the way community members respond to the organization as well as the way the organization responds to the community. For example, opening a burlesque theatre in a very religious part of town may be a fatal mistake for the owner since it is contrary to the values of the community. Within the military, attitudes toward the Army play an important role. Such attitudes can be directed toward the Army as either an "employer" or as a "constructive force". These attitudes help determine whether citizens are attracted to the Army as an employer, or whether the community supports the Army in terms of appropriations.

The second component within the community is the general economic condition. When economic conditions are poor, unemployment is usually high and it is easier for an employer to attract better quality employees for lower wages. On the other hand, poor economic conditions may jeopardize the survival of organizations that are not deemed essential by the community. Economic conditions affect the military in a number of ways. First, these conditions help determine if the military will achieve their recruiting and retention goals. It has been observed that during times of high unemployment enlistment and retention rates increase for the military. Second, general economic conditions impact on the size of the military appropriation. When economic conditions are poor, money may be taken out of government and used to stimulate economic growth in the private sector.

The third component in the community relates to the demographic composition of the population. The composition of the population in terms of such factors as income, education, age, and sex affects the organization in at least two ways. First, it helps define goods and services that the organization should provide in order to be responsive to community needs. Second, it determines whether the community provides an adequate applicant base for satisfying the organization's personnel objectives. For example, a knowledge of the composition of different communities in terms of age and sex should be very important for deciding where to set up Army recruiting stations.

A fourth factor in the community which impacts on an organization is international relationships. One important aspect of international relations concerns whether it is wartime or peacetime. During wartime many organizations (including the Army) which are involved in the efforts are enhanced, while those involved exclusively in peacetime efforts either change their plans and objectives or terminate their existence. The military during wartime obviously receives its maximum political and financial support. As the country reverts back to a peacetime situation, a shift in the economy occurs. Organizations supporting the war effort now must change their plans and objectives to respond to the changing needs of the community.

The external environment impacts significantly on the various human resource management functions. First, it has a major effect on recruiting. The establishment of the recruitment policy is affected by the general attitude of the community toward the organization. If the attitude is negative, the organization may have to try to change its image or try to recruit from a different segment of the population. The external environment impacts on selection. If selection standards are imposed which are not attainable by the available applicant population, the organization will be unable to fill its vacancies. In this case, the organization has at least three choices: a) recruit from a different segment of the population, b) lower selection standards and be satisfied with a reduced level of job performance, or c) lower selection standards, but compensate by increased training so that no reduction in job performance is obtained.

The impact of the external environment on training is two-fold. First, training must include knowledges and skills which are necessary for those recruited. Second, training can be used to change attitudes which are prevalent among those recruits.

Of all the components in the model, the external environment is the component least directly linked to performance. Yet it is not for this reason unimportant. Through its impact on the organization, it exerts tremendous influence on the way in which duties are defined, in which performers are selected, in which performance objectives are set, and in which performance standards are defined and implemented. It is perhaps that component of the model which has historically been most often neglected, but such neglect leads inevitably to an incomplete understanding of performance.

Organization

Internal Macro-Environment. The internal environment of an organization can be described on either a macro-level or a micro-level. On the macro-level, the variables important for consideration impact organization-wide. The relevant variables at the micro-level impact directly on the job and represent a source of variance between jobs within an organization.

Three variables in the internal macro-environment are important for consideration in the performance model. They are financial, human and physical resources. The financial resources of the organization consist of funds in the bank, funds invested and funds projected from future business. Within the military context, these funds are limited to the military appropriations. The size of these appropriations is controlled to some extent by factors in the external environment such as economic conditions and international relationships. Obviously, during wartime the appropriation would be much larger than during peacetime. Also, during poor economic conditions, the government may reduce the military appropriation in order to direct more money into the civilian economy.

The human resources within the organization can be described along a number of dimensions. The relevance of the various dimensions is dictated by the personnel objectives of the organization. Potentially relevant information includes the number of people by specialty, level in the organization, sex, and race. Within the Army, variables of interest might include the number of soldiers by: a) MOS, b) Combat, Combat Support or Combat Service Support, c) knowledges, skills, abilities and other individual characteristics (KSAO's) and d) rank. Knowledge of the manpower mix in terms of these relevant dimensions is necessary for formulating personnel objectives and strategies .

The third variable of the internal macro-environment is physical resources. Physical resources include plants, offices, equipment and materials. In terms of the Army, physical resources might include Army bases and office buildings, weapon systems, vehicles, computers, and other military inventory.

The internal macro-environment impacts on all human resource management functions. First, financial resources affect the recruiting strategy by imposing limits on the type of people that can be attracted to the organization. Second, it impacts on selection by establishing limits on the KSA's that can be required in the applicant population. Third, it affects training in two ways: a) financial resources relate to the quality of personnel selected and the quality of personnel selected impacts on training needs, and b) financial resources limit the amount of funds available for training. Financial resources affect performance evaluation to the extent that pay is related to performance. This might be demonstrated on such criteria as productivity.

Available human resources, as well as physical resources, also impact on human resource management functions. Knowledge of available human resources in the organization helps determine where needs still exist and impacts on the type of people that the organization needs to attract. In addition, knowledge of human resources determines the KSA's that applicants need to be selected on in order for the organization to operate effectively. Knowledge of physical resources relates to recruitment and selection, because physical resources such as equipment impose requirements on the type of people needed to effectively interact with the physical environment.

Human and physical resources in the organization also impact on training. First, developmental training for promotion or transfer is based on the knowledges, skills and abilities of current employees. Second, changes in the physical environment may suggest new training needs.

Human resources and physical resources have definite effects on performance evaluation. Human resources directly impact on the level of performance which is an integral component of the evaluation process Physical resources, on the other hand, may have a more indirect effect. Physical resources may impose limitations on an employee's productivity. This is typically called criterion contamination. If this is not taken into account in the evaluation process, validity of the evaluations may be sacrificed. Second, physical resources help establish the criteria along which employees may be evaluated.

Organizational Plans and Objectives. The external environment and the internal macro-environment are the inputs for defining organizational plans and objectives. Objectives can be viewed as targets which the organization sets out to achieve. Objectives can be either long-term or short-term. The specification of these objectives is based on the demands placed on the organization from the external environment, as well as the available resources within the internal environment. The organizational plans are the specific steps taken by the organization to accomplish its objectives. In the Army, a major long term organizational objective is readiness. Short term objectives may include recruitment quotas, percentage reductions in attrition and increased technical competence. One Army plan addressing its long term objective, "readiness", is known as Air-Land Battle 2000, which is aimed at both personnel and equipment readiness in the year 2000. Plans for achieving short-term goals might include: a) increasing the number of recruiters and improving the quality of recruiters to meet recruitment quotas, and b) enhancing job relevance of training and improving performance evaluations to increase the technical competence of the force.

<u>Personnel Objectives and Strategies</u>. Coleman (1979) points out that organizational plans and objectives are the bases for establishing personnel objectives and strategies. Generally, the number of personnel needed and the composition of the personnel in terms of the mix of critical KSAO's is an outgrowth of the specification of organizational plans and objectives. The personnel strategies involve specifying the steps that will be taken to attain the personnel objectives. Such strategies or action plans might involve changes in procedures related to recruitment, selection, training and development. Specification of the personnel strategies requires: a) knowledge of the KSAO mix of the human resources already in the organization, b) knowledge of the KSAO mix found in the relevant applicant population and c) knowledge of the KSAO mix critical for achieving the organizational objectives. Much of the information can be obtained from adequate job analyses. In the Army, for example, there might be an organizational objective to modernize a particular weapon system. Part of the personnel objective might be to have a certain number of personnel operate the new weapon system. In order that these personnel are available to use the system, personnel strategies have to be developed. These strategies might consist of: a) conducting a job analysis to determine the KSAO's critical for successful performance on the new weapon system, b) assessing current employees to identify those possessing the critical KSAO's, c) defining the number of people in the relevant applicant population and d) examining institutional recruiting, selection, training and development processes necessary for adequately manning the system.

Organizational plans and objectives and personnel objectives and strategies are critical components for the recruitment, selection and training human resource management functions. They serve to guide the establishment of recruitment policy. They serve to establish selection criteria as well as to specify methods for making selections. And, they are important for determining training content, since one of the major personnel strategies is training.

Internal Micro-Environment

The Organizational Plans and Objectives and the Personnel Objectives and Strategies serve as input to the internal micro-environment. The internal micro-environment consists of the "person" and the "job". The job can be further divided into its Formal Structure, Task Structure and Social Structure. According to Kane (1975), the person and job variables have a dual nature, "one consisting of those properties which are objectively measurable; and the other consisting of those properties which are products of the phenomenology of the organizations' members." Those properties which are objectively measurable are minimally influenced by those doing the measuring. In objective measures, variance attributable to the measurer is considered to be error. On the other hand, phenomenological properties are those which treat measurer variance as true variance. Phenomenological properties are thought of as a product of the interaction between the person and the objective properties of the job.

The Person. Psychological research (Tyler, 1968; Anastasi, 1958, 1983; Dunnette, 1976) has shown that individuals differ in their capacities to exhibit and/or acquire with training specific knowledges, skills, abilities, and other behaviors (KSAO's), such as those required for certain jobs and activities. In addition, measurement of individual differences in human abilities has been frequently utilized to predict job performance, training proficiency and other relevant work criteria (Ghiselli, 1966). Hence, in order to better specify what variables influence work performance and how performance should be assessed, it is crucial that the domain of human attributes be comprehensively described and patterns of abilities linked to task/job requirements, levels of effective performance, and the goals of the individual and work organization.

In this conceptual framework the person is viewed as a composite of those KSAO's brought to the job or acquired after the individual has been hired. Those KSAO's brought to the job are seen as either innate or acquired through prior training or experience. Those KSAO's acquired after employment are learned through formal training or through experience on the new job. The objective nature of the person has traditionally been assessed through measures such as I.Q. tests, psychomotor performance tests or personality inventories. The phenomenological nature of the person is typically determined based on subjective, self-report assessments on the domain of KSAO's.

The job performance of an individual within this broad systems approach is directed by a complex group of individual difference factors. Specifically, individual difference variables determine which KSAO's a person brings to the job, how rapidly a person learns new tasks, and how responsive the person will be to training strategies and organizational incentives.

In the person sub-system which is shown in Figure 3, the behavior of an individual can be examined in relation to the following four general attribute domains: cognitive, psychomotor/ physical proficiencies, affective, and vocational preferences/ interests (Dunnette, 1976). The cognitive domain, which emphasizes abilities that involve sensory processing and cognitive manipulation of information, has been taxonomized through factor analytic techniques by researchers at the Educational Testing Service (ETS) (Ekstrom, 1973; Ekstrom, French, and Harman, 1976). These abilities are exemplified by such factors as Fluency, Inductive Reasoning, Verbal Comprehension, and Spatial Orientation and Visualization. Table 1 provides definitions of the various cognitive factors.

The psychomotor and physical proficiencies domain, researched by Fleishman and his colleagues (Fleishman, 1964; 1972; 1982; Theologus, Romashko, and Fleishman, 1973), consists of 19 nearly orthogonal dimensions. These dimensions, which include motor skills such as Reaction Time and Manual Dexterity, and physical proficiencies such as Strength and Stamina, have been examined with respect to changes in task requirements and practice effects. In Table 2 the constructs which comprise this domain are more thoroughly explained.

The affective domain, which focuses on personality related variables, has been conceptualized from trait, situational and interactionist perspectives. Browne and Howarth (1977) developed a trait-oriented taxonomy of personality which posits that personality attributes such as Sociability and Dominance are relatively stable. The taxonomy of variables which are proposed as representative of the affective domain is presented in Table 3. Although various trait measures have moderately high convergent validity, this perspective does not adequately consider the influence of environmental factors on behavior. Perhaps, a reciprocal interaction approach (Bandura, 1978), which considers both the impact of persons' perceptions/feelings on their responses to situational contingencies and the controlling effects of environments, can better account for the complexity of human personality and facilitate its accurate assessment.

The fourth attribute domain is comprised of vocational preferences and interests. Research (Strong, 1955; Holland, 1976) indicates that various intra-individual interest dimensions tend to be highly stable for long periods of time, and have consistently moderate correlations with dimensions in other domains which predict job performance. Variables in this domain may assist in the linkage of classes of human attributes to groups of tasks and behavior dimensions. Table 4 summarizes the interest domain as developed by Holland (1976).

In conjunction with the basic attributes discussed under the four domains, performance on the job is influenced by such relatively immutable characteristics as the person's sex, race, and age. The manner in which patterns of attributes/characteristics are exhibited on the job can be conceptualized as an out-growth of the interaction between the person's general life and job experiences and his/her specialized, job-relevant knowledges and skills, which tend to reflect achievement or opportunities for achievement.

In order to account for variance in observed job performance, an understanding of the organization of attributes/abilities and fluctuations in performance capabilities is crucial. For example, different patterns of individual attributes can often lead to comparable levels of effective job performance. Also, fluctuations in performance capabilities can occur as a result of such factors as stress, fatigue, training opportunities, and changes in technological-systems requirements.

Although the four separate attribute domains and the moderating influence of experience and environmental demands provide a conceptual structure for the person model, a complex set of cognitive-control functions may help explain variability in performance effectiveness. These cognitive control functions, which include such information processing operations (Carroll, 1976; Sternberg, 1977) as attention, transformational capabilities and storage and retrieval of information, are thought to exist on a continuum for all persons and possibly account for subtle differences among persons with respect to the efficiency/effectiveness of their performance on a wide array of tasks.

Although currently cognitive processing research is in its infancy, examination of the information processing components and requirements of various job relevant tasks might provide valuable information for clustering/sequencing tasks, or developing interventions to facilitate performance in both training and actual job contexts. Table 5 briefly describes the cognitive processing operations that can impact on the accuracy and efficiency of task/job performance.

The Objective Job. The objective job consists of the: a) Objective Formal Structure, b) Objective Task Structure and c) Objective Social Structure. These objective job categories are identified in Figure 4 . Factor analytic work by Pugh et al. (1968) led to the identification of four factors related to the objective formal structure of the organization. The factors include: a) Concentration of Authority, b) Structuring of Activities, c) Line Control of Workflow, and d) Relative Size of Supportive Components. To this list Kane (1975) has added a fifth factor, Formal Reward System. Definitions of the factors are given in Table 6.

The second category of the Objective Job involves the Objective Task Structure and includes the work by Perrow (1967) and Lynch (1974) on the identification of task and technology dimensions. Those factors most appropriate to the Army are a) Coordination and Interdependence of Work, b) Control, c) Routineness, d) Search Process when exceptions occur, and e) Variability of Equipment and Information. Definitions of the five dimensions are presented in Table 7 .

Golembiewski (1965) discusses six factors of small group relations which help define the Objective Social Structure of the job. Four of these fit into the domain of objective measures and are appropriate for inclusion in the Army context. These are: a) Leadership, b) Role Style of the Leader, c) Personality and d) Norms and Sanctions. Kane (1975) has added two factors to the list for the Objective Social Structure: Presence of Co-workers and Communication Feasibility. Table 8 provides definitions for the six Objective Social Structure factors.

The Phenomenological Job. The phenomenological nature of the job characterizes the perceptions or affective meaning of the objective job as subjectively experienced by members of the organization. The implication is that the model views the phenomenological job as an interaction between the person and the objective job. Although the components of the phenomenological job are identical to those of the objective job, the variables which comprise these components are different.

Table 9 presents the factors and their definitions for the Phenomenological Formal Structure. While little research has been directed toward this dimension, some work on control has been conducted by Tannenbaum (1968). Within this dimension there are four factors a) Perceived Structure, b) Perceived degree of Control, c) Perceived Distribution of Support Functions, and d) Equity. Equity, for example, is the phenomenological counterpart of the Formal Reward System. To the extent that the person perceives the formal rewards as consistent with the amount of effort expended, the person will perceive the reward system as equitable.

Inspection of Table 10 indicates that six factors have been identified as comprising the Phenomenological Task Structure. Work by Hackman and Lawler (1971); and Lawler, Hackman and Kaufman (1975) lead the research in this area. The dimensions include: a) Subjective Expected Value of Extrinsic Outcomes, b) Autonomy, c) Task Identity, d) Variety, e) Feedback, and f) Challenge.

The last dimension in the phenomenological job is the Phenomenological Social Structure. This dimension, like the Phenomenological Task Structure, consists of six factors. Factors included in this dimension have been identified and studied mainly by social psychologists interested in small group processes. Included in this dimension are: a) Leadership, b) Role Style of Leader, c) Compatibility, d) Social Justice; Equity, e) Cohesiveness and f) Isolatedness. Table 11 provides definitions for these six factors.

Since the person, the objective job and the phenomenological job are the inputs to defining work behavior, these components are also critical to the human resource management functions. Typically, information on these components is gained through job analyses. Job analysis can be thought of a systematic method for collecting job information. Many methods exist for performing job analysis. However, two broad categories of job analysis exist: worker-oriented and job-oriented. Worker-oriented job analysis focuses on the individual within the job rather than the job independently of the person. Methods of job analysis in this category include the critical incident technique, which focuses on work behaviors, and job elements, which involves a direct identification of KSAO's. Job-oriented analysis, on the other hand, analyzes the job independently of the person in the position. Methods such as task analysis and functional job analysis fall into this category.

Since a major function of job analysis is to specify behaviors which are critical to successful job performance, this information is crucial for all human resource management functions. First, job analysis specifies the human attributes that need to be considered in establishing recruitment, selection, and training programs. Second, job analysis information is the key to developing performance evaluation criteria.

Work Performance

Work Behavior. The behavior of an individual is seen as a composite of three components. Different combinations of these three components determine whether the person has the ability to engage in the appropriate behavior and whether the person is motivated or willing to expend the effort to engage in appropriate behavior. The first component is the person. The KSAO's that the person brings to the job or that are acquired through training have a main effect on the behavior. Research on validity generalizaion supports the finding that validities of many individual difference measures generalize across situations. The second component, the job, is also seen as having a main effect on the variance in behavior. That is, given a fixed level across people on the KSAO's, a significant portion of the variance in behavior can be accounted for on the basis of variability on the dimensions that define the job. In the field of social psychology, research has indicated that situational factors have a strong influence on both individual and group functioning. The third component which influences behavior, is an interaction between differences among people on the KSAO's and variability on the dimensions defining the job. For example, in order to achieve a constant level of effective behavior, individual differences on the KSAO's would be associated with differences on the job dimensions.

Motivation. There is general agreement that work performance is a result of some multiplicativity of ability and motivation (Lawler, 1973). However, there is little agreement as to the relative salience of each component. Opinions on the subject range from Dunnette's (1973) conclusion that ability differences are "the most important determiners of differences in job performance", to the Naylor, Pritchard, and Ilgen (1980) assertion that the ability component is not particularly relevant in most cases. There is agreement that in the case of lower-level jobs, where ability requirements are minimized, the significance of this factor is reduced, thereby increasing the importance of motivation as a determinant of work performance (Steers & Porter, 1979; Lawler, 1973). In addition, Atkinson & Birch (1978) consider that an individual's true level of ability will only be completely manifested in his level of performance when the individual is optimally motivated.

According to Steers & Porter (1979) and Campbell & Pritchard (1976), there is currently no completely unifying theory of motivation although most of the theoretical approaches to motivation are complementary rather than contradictory. The major approaches to motivation - Need theory, Equity theory, Reinforcement theory and Expectancy/Valence theory all provide important perspectives from which to view motivation (Steers & Porter, 1979). Need theory stresses individual characteristics, but not the job and work environment. Equity theory focuses on individual characteristics and feelings regarding the equity/inequity of the organizational reward system. It fails, however, to consider many other sets of variables which affect motivation. Reinforcement theory emphasizes work environment variables but ignores individual differences, needs, and attitudes. Expectancy/Valence theory recognizes the role of individual differences, variations in needs, values attached to potential outcomes, perceptions of equity/inequity of rewards, and expectancies regarding work performance leading to various outcomes. However, Expectancy theory does not appear to be a powerful predictor of motivation to perform (Kress, 1981).

The ideal theory of motivation according to Steers & Porter (1979) should consider the three categories of variables relating to the individual, the job, and the work environment as well as their interactive effects. In accordance with this theoretical perspective the job performance model postulates that motivation is an interaction among the attributes of the worker, characteristics of the job and the organizational opportunities/incentives.

Performance Evaluation System

Organizational Control Systems. The performance evaluation function of human resource management is viewed in the current model as an organizational control system. The results of a series of behaviors are a set of outcomes. While many outcomes may occur, only a subset may be relevant in a specific organizational context. The importance of these outcomes is a function of the priorities established by the organization as reflected in its control system.

In order to monitor its progress an organization must have a control system. Such a control system must include performance indicators, performance standards and feedback mechanisms. Performance indicators are those measurement methodologies that assess the movement toward organizational objectives. Since organizational objectives tend to be multidimensional (e.g., productivity, goodwill, absenteeism) more than one performance indicator may be necessary. Along with performance indicators there are typically performance standards. These standards specify levels on the performance indicators which represent degrees of effective performance. For example, how many target hits for a tank crew is the standard of acceptable performance? How many recycles in a particular MOS suggest that there might be a problem in the selection system and/or training of new recruits?

Individual versus Organizational Controls. Controls can be viewed as either organizational or individual. Organizational controls include those performance indicators, standards and feedback mechanisms which assess and modify an organization's progress toward its goals and objectives.

Individual controls consist of those performance indicators, standards, and feedback mechanisms which assess and help modify the extent to which individual performance contributes to the accomplishment of organization goals. If one short term organizational objective is to recruit 12,000 active Army enlistees a month and there are 6000 recruiters then each recruiter must enlist two persons per month. Using as a performance indicator an auditing procedure it would be possible to assess whether each recruiter achieved the minimum performance standard of two enlistees. As can be seen from this example, an individual control should only be defined within the context of organizational objectives. To the extent that individual controls correlate with organizational objectives, the individual performance indicators are relevant. To the extent that the organizational objectives contain variance not included in the individual controls, the performance indicators are deficient. And, to the extent that the individual performance indicators contain variance not included in organizational objectives, the individual controls are biased.

Table 12 presents a taxonomy of performance indicators. These indicators fall into three broad categories: a) Direct Soldier Performance Data, b) Soldier Personnel Data and c) Judgmental Data. The table further specifies performance indicators by the situation within which the evaluation is conducted. Two evaluative situations are included- Training and Actual Job. For example, if the performance indicator of interest in the military is rate of advancement of a soldier, the taxonomy specifies the ways it can be measured in each evaluative situation. During training we can administer weekly knowledge tests on radio repair and use the change score as an indicator of advancement. In terms of the actual job, we can review personnel records and determine how long it took each soldier to advance from an E-2 to an E-3 or from an E-3 to E-4.

The measurement of individual and organizational outcomes through the use of individual and organizational performance indicators generates the basic information for decision-making. The individual decision-making process involves those decisions made by the organization which affect only one person. This could take the form of termination, promotion, training development, or some other result. These decisions feed back into the Personnel Objectives and Strategies component of the model. For example if a person is promoted or terminated, this change may require that new personnel objectives and strategies be defined.

The organizational controls, on the other hand, take two forms. The first is in terms of personnel. The personnel aspect of organizational controls represents a judgment based on an analysis of individual performance indicator data across all persons in a relevant unit of analysis such as job, unit, team, or organization. Such an analysis might result in a determination that all soldiers in a particular MOS are deficient in driving a certain type of important vehicle. This would necessitate the development of a special training program for all soldiers in that MOS. The second aspect of organizational controls relates to equipment and facilities. This involves a judgment as to the adequacy of equipment and facilities for the achievement of organizational objectives. Within the Army, an assessment of equipment readiness based on a compilation of information on each type of equipment serves as the input for determining whether the equipment is at an acceptable standard for achieving organizational objectives.

The Decision Making Process

Information on both components of organizational performance indicators, personnel, and equipment and facilities, is used as input into the organizational decision-making process. Information can be used in any of three basic decision models. These models have been named (Cronbach & Gleser, 1965): a) compensatory, b) conjunctive and c) disjunctive. Assuming that organizational controls are multidimensional, in a compensatory model high scores on one dimension can compensate for low scores on another dimension. Multiple regression analysis is typically used with this approach. The conjunctive model, on the other hand, is more commonly referred to as a multiple hurdles approach. In this case, organizational performance on each indicator must meet the specified standard in order to be defined as satisfactory. The disjunctive model is different in that only one performance indicator out of the total array defining an organization's performance must be at or above standard for an organization to conclude that it is minimally successful.

The organizational decisions information must be fed back into the Organizational Plans and Objectives and the Personnel Objectives and Strategies. In terms of organizational objectives, the decision might involve increasing or decreasing production of new weapon systems, renovating old weapon systems or designing new weapon systems. Decisions related to Organizational Plans might include contracting out part of the work or funding Research and Development projects for new weapon systems.

The impact of organizational decisions on Personnel Objectives and Strategies are both direct and indirect. Organizational decisions impacting directly on Personnel Objectives are based on a compilation of individual performance indicators across the relevant employee subgroup. For example, all computer programmers working in the scientific area may have to acquire a working knowledge of the Statistical Analysis System (SAS). The Personnel Plan in this case, might be either for the organization to develop and conduct a SAS training course or to contract with the SAS Institute to conduct such a course. Another possible Personnel Plan might involve the recruitment and selection of people with a knowledge of SAS. Each alternative plan might be evaluated in terms of costs and benefits to the organization. The major difference between the impact of organizational decisions and individual decisions on the planning process is that in the former case Personnel Objectives and Strategies are defined for an entire subgroup of employees due to a common deficiency or need while in the later case personnel decisions are made for a specific individual.

Organizational decisions also affect the Personnel Objectives and Strategies by defining new personnel needs based on changes in Organizational Objectives. For instance, one organizational decision might be to improve clerical efficiency of soldiers in the Clerk-Typist MOS since the average cost to type one letter is excessively high. The plan for achieving this objective might include the purchase of state-of-the-art word processing equipment. Instituting this Organizational Plan and Objective dictates modifications to the Personnel Objectives and Strategies. In the present case, this may include an objective of having a certain number of soldiers proficient in the use of word processing equipment. The Personnel Strategy needed to accomplish the personnel objective might include the recruitment and selection of already trained people or the training and development of current soldiers in that MOS. Which approach the organization takes must be based on many factors including the availability of trained human resources in the external environment, the competition for these resources and the financial resources available to the organization.

Implications for Research

The model clearly demonstrates that work behavior involves an extremely complex set of variables. While the attributes that the person brings to the job are important, many other variables help determine whether the behavior exhibited will lead to effective or ineffective job performance. The situation obviously has a strong impact on work behavior. The objective aspects of the job interact with individual difference factors that the person brings to the job, and has a main effect of setting limits on actual job behavior. The interactions between individual KSAO's and the objective job are very complex. First, different combinations of KSAO's can interact with objective job factors to lead to very different behaviors. Second, the objective job interacts with KSAO's resulting in the phenomenological job which also affects behavior. The model quite clearly points out that one's perceptions of work requirements/ demands has an effect on work behavior.

Many areas of research are needed to better understand this relationship between the person, the job and work behavior:

1. Investigations which compare different job analysis methodologies to determine the unique contribution that each makes in understanding job performance.

2. The evaluation of the role of affective, experiential and interpersonal dimensions as requirements for effective performance in various MOS. For example, research addressing the role of affective/experiential dimensions in distinguishing levels of effective individual and unit performance is suggested by the model.

3. The development of a taxonomy of job dimensions that addresses both the objective job and the phenomenological job.

4. Research to better understand how KSAO's interact with the objective job to produce the phenomenological job.

5. Selection studies which go beyond individual difference measures as predictors, but include the objective as well as the phenomenological variables. The expectations, interests, and preferences of soldiers for particular jobs and/or types of military work could be examined. In particular, discrepancies between individual expectations for a job (phenomenological job) and such objective job dimensions as task structure, supervisory span of control and routineness might be studied.

6. Comparison of different methods of inferring KSAO's from job analysis methodologies in terms of psychometric adequacy.

7. Investigations involving the clustering of jobs not only in terms of KSAO'S but also in terms of objective and phenomenological job variables.

Since job performance goes well beyond the direct effects of individual KSAO's, the question arises as to how the effects of other factors are accounted for in the performance evaluation process. Much research is needed in this area, for example:

1. Research investigating the question: What are the sources of criterion contamination that are part of the objective job and how do they affect the evaluation of performance?

2. Investigations of the effects of rater behavior on the evaluation of subordinates. For example, how do such variables as rater/ratee demographics, rater information

processing parameters, raters' own performance/standards and type of rater training impact on the performance evaluation process?

3. A comparison of different evaluation methods in terms of such properties as reliability, validity, and acceptability.

Once we have established the psychometric adequacy of the performance evaluation measures, it is necessary to focus on how this information is used in the decision process. Research needed in this area includes:

1. A comparison of clinical vs statistical prediction.

2. An investigation of the effects of man's information processing capabilities on the quality of his appraisals.

3. Policy analysis comparing good versus poor evaluators.

To this point, we have been talking exclusively at the level of the job and the individual. However, it was demonstrated in the model that a macro view can help in understanding how the organization operates. Policies and procedures established at the organizational level have a significant impact both on the understanding of work behavior and on overall organizational effectiveness. If one defines effectiveness only in terms of individual performance, he can never completely understand the impact of organizational policies and procedures on the total organization. For example, a policy decision to increase all employees' pay by 10% may have a positive effect on employee performance but a negative effect in terms of overall organizational effectiveness. In order to better understand the impact of the macro- environment it would be worthwhile to study:

1. The dimensions that define organizational effectiveness in terms of organizational goals and objectives and the relative importance of each dimension.

2. The impact of existing organizational policy on the dimensions of organizational effectiveness.

3. The effect of changes in organizational policy on organizational effectiveness.

Summary

Job performance in military organizations is a multidimensional process that is dependent on the interactions among the attributes/characteristics of the individual, requirements of the job, and the objectives and policies of the organization. The Army, as most organizations, has tended toward selection/classification based on cognitive abilities, related to measurement emphasizing training performance. In order to provide a more complete basis for evaluating the individual for enlistment/reenlistment, providing appropriate training, more complete measurement of the individual's total Army performance, and overall improvement of the productivity of the individual and the Army, a large scale, multi-year Soldier Selection, Classification, and Utilization Project has begun. It will address many of the variables discussed in the systemic model of job performance proposed in this paper. In general, the job performance model defines important performance variables, examines interactions among the person-job-organization components, and identifies broad areas for research.

When used as a conceptual tool to direct and guide the development of research projects, the job performance model should be of benefit in terms of 1) optimizing the job/person match, 2) improving individual performance both in training and on the job, and 3) improving organizational effectiveness.

Table 1

Cognitive Domain

Factor	Definition
a. Fluency:	
1. Associational Fluency	The ability to produce words from a restricted area of meaning.
2. Expressional Fluency	The ability to supply the proper verbal expression for ideas already stated or describe a suitable expression that would fit a given semantic frame of reference.
3. Ideational Fluency	The ability to quickly produce ideas, and exemplars of an idea about a stated condition or object.
4. Word Fluency	The ability to produce isolated words that contain one or more structural, essentially phonetic, restrictions without reference to the meaning of the words.
b. Inductive Reasoning:	The ability to form and test hypotheses directed at finding a relationship among elements and applying this principle to the identification of an element which fits the relationship.
c. Associative (Rote) Memory:	The ability to remember bits of unrelated information or material.

Table 1 (continued)

Factor	Definition
d. Span Memory:	The ability to recall perfectly for immediate production a series of items after only one presentation of the series.
e. Number Facility:	The ability to rapidly manipulate numbers in arithmetical operations, performance of elementary arithmetical operations under speeded conditions.
f. Syllogistic (Deductive) Reasoning:	The ability to reason from stated premises by ruling out nonpermissible combinations and thus to arrive at the necessary conclusions.
g. Verbal Comprehension:	The ability to apply a knowledge of words and their meaning in understanding connected discourse.
h. Verbal Closure:	The ability to solve problems which require the identification of visually presented words when some of the letters are missing, disarranged or mixed with other letters.
i. Flexibility and Speed of Closure	The ability to mentally maintain a visual percept and find it embedded in distracting stimuli.
j. Perceptual Speed:	The ability involves speed in finding figures and symbols, making comparisons, and executing other relatively simple tasks which involve visual perception.

Table 1 (continued)

Factor	Definition
k. Spatial Orientation and Visualization	The ability to perceive spatial patterns or to maintain orientation with respect to objects in space, and the capacity to manipulate or transform the image of spatial patterns into other vi[illegible]ual arrangements
l. Figural Fluency:	The ability [illegible] produce a response qui[illegible]y by drawing a number of [illegible]mples, elaborations [illegible] restructurings when presented with a given visual or descriptive stimulus.
m. Visual Memory:	The ability to remember the configuration, location, and orientation of figural material.

Table 2

Psychomotor and Physical Proficiencies Domain

Factor	Definition
a. Control Precision	The ability to execute controlled muscular movements, which are necessary to adjust or position a machine or equipment control mechanism.
b. Multi-limb Coordination:	The ability to coordinate the movements of two or more limbs (e.g., two legs, two hands or one leg and one hand). This ability does not apply to tasks in which trunk movements must be integrated with limb movements.
c. Response Orientation:	The ability involves the selection of the correct movement in relation to the correct stimulus, when speed is a factor.
d. Reaction Time:	The ability involves the speed with which an individual is able to make a single response to the onset of a stimulus condition.
e. Speed of Limb Movement:	The ability concerns the speed with which an individual can execute a gross, discrete arm or leg movement where accuracy does not apply.
f. Rate Control:	The ability to make timed anticipatory motor adjustments, relative to changes in the speed and/or direction of a continuously moving object.

Table 2 (continued)

Factor	Definition
g. Manual Dexterity:	The ability to make skillful, coordinated movements of a hand or a hand in conjunction with its arm; requires coordination within a limb.
h. Finger Dexterity:	The ability to make skillful, coordinated movements of the fingers, where manipulations of tiny objects may or may not be involved.
i. Arm-hand Steadiness:	The ability to make precise, steady arm-hand positioning movements, where both strength and speed are minimized.
j. Wrist-finger Speed:	The ability emphasizes the speed with which discrete movements of the fingers, hands and wrists can be made.
k. Choice Reaction Time:	The ability to select and initiate the appropriate response relative to a given stimulus in the situation where two or more stimuli are conceivable and where the correct response is chosen from two or more alternatives; speed with which response is initiated is a factor.

Table 2 (continued)

Factor	Definition
l. Static Strength:	The ability consists of the degree of muscular force exerted against a fairly immovable or heavy external object in order to lift, push, or pull that object; general ability common to different muscle groups.
m. Dynamic Strength:	The ability involves the power of arm and trunk muscles to repeatedly or continuously support/move the body's own weight. The amount of performance decrement when muscles are repeatedly stressed is a focal concern.
n. Explosive Strength:	The ability to expend energy in one or a series of explosive muscular acts; ability requires a mobilization of energy for a burst of muscular energy.
o. Stamina:	The ability involves the capacity to maintain physical activity over prolonged periods of time.
p. Extent Flexibility:	The ability to extend, flex or stretch the trunk and back muscle groups.
q. Dynamic Flexibility:	The ability to make repeated trunk and/or limb movements where both speed and flexibility of movement are involved.

Table 2 (continued)

Factor	Definition
r. Gross Body Equilibrium:	The ability involves maintaining the body in an upright position (body balance) or regaining body balance in cases where equilibrium is threatened or temporarily lost.
s. Gross Body Coordination:	The ability to coordinate movements of the trunk and limbs when the whole body is in motion or being propelled.

Table 3

Affective Domain

Factor	Definition
a. Social Shyness:	This dimension is described as a lack of initiative in developing new friendships; hesitancy to engage in interpersonal social exchanges.
b. Sociability:	This dimension involves an active involvement and enjoyment of interpersonal group social activities.
c. Mood-swing Readjustment:	This dimension is reflected in strong emotional mood swings which can shift from feelings of elatedness/happiness to listlessness/depression.
d. Adjustment-Emotionality:	This dimension is manifest as nervousness and noticeable worry over potential problems.
e. Impulsiveness:	This dimension involves responding to situations, people, and events with little forethought as to behavioral consequences.
f. Persistence:	This construct involves attending to and expending effort until a task/job is completed.
g. Hypochondriac-Medical:	This dimension involves self-disclosure of complaints about physical health symptoms.

Table 3 (continued)

Factor	Definition
h. Dominance I:	This factor describes individuals who exhibit submissive behavior and do not stand up for their rights/concerns in an argument.
i. Dominance II:	This factor is indicative of the ability to take charge and exert leadership behavior.
j. General Activity:	This dimension reflects a preference for and enjoyment of a diverse group of sports activities; action-oriented.
k. Trust versus Suspicion:	This factor combines a carefree attitude about life with a tendency toward sensitivity.
l. Superego:	This dimension involves the expression of behaviors which are related to a strong "social conscience".
m. Social Conversation:	This construct is indicative of an individual who is gregarious or very talkative.
n. Inferiority:	This dimension involves feelings of being unsuccessful or inability to make a positive impression with other people.
o. Cooperativeness - Considerateness:	This dimension is reflective of self-sacrificing behaviors or playing a good Samaritan role; lending help to solve the problems of others.

Table 4

Vocational Preference/Interest Domain

Realistic:

Rugged, robust. practical, good motor coordination and skills; lack interpersonal and verbal skills; persistent.

Conventional:

Prefer well ordered environment and systematic verbal and numerical activities; usually conform and prefer subordinate roles; view self as conscientious, efficient, calm, orderly and practical.

Investigative:

Scientifically oriented, introspective, social; enjoy ambiguous tasks and prefer to work independently; possess unconventional attitudes and values; confident of intellectual abilities, but not leadership or persuasive abilities.

Enterprising:

Possess verbal skills suitable to selling, dominating and leading; have a strong drive to reach organizational goals or economic gains; concern for power status, and leadership. See self as aggressive, self-confident, cheerful, and sociable.

Artistic:

Impulsive, creative, expressive, original, intuitive, introspective, non-conformist and independent; value aesthetic qualities; dislike highly structured problems or use of gross physical skills; perform well on standard creativity measures.

Social:

Sociable, responsible, humanistic, religious; like working in groups; prefer to solve problems through feelings and interpersonal manipulations. See self as understanding, responsible, idealistic and helpful.

1) Adjacent interest types are complimentary.
2) Interest types on the diagonals conflict.

Table 5

Information Processing Operations

Cognitive Processes	Definition
a. Encoding	Operation by which information is input to the system; process by which different types of information about the stimulus are extracted.
b. Synthesis/Integration:	Operation by which new information is organized and generated from information already encoded.
c. Transforming:	Operation by which some stored rules are applied to the information structure already present.
d. Storing	Operation by which new information is incorporated into existing information structures, while the entire content is retained.
e. Retrieving:	Operation by which previously stored information is made available to the processing system.
f. Searching	Operation by which an information structure is examined for the presence or absence of one or more properties.
g. Comparing:	Operation by which two information structures (either internal or external to the processing system) are judged to be same or different.
h. Responding:	Operation by which the appropriate (motor) action is selected and executed.

Table 6

Objective Formal Structure

Factor	Definition
a) Concentration of Authority	The degree to which authority for decisions in an organization is centralized at the higher hierarchial levels within it.
b) Structuring of Activities	The degree of standardization of organizational routines, degree of role specificity, and the degree of formalization of written procedures.
c) Soldier Control Workflow	The degree to which the control of work is in the hands of the soldier as compared to control by impersonal procedures.
d) Formal Reward System	The degree to which formal extrinsic rewards are differentiated by job.

Table 7

Objective Task Structure

Factor	Definition
a) Coordination and Interdependence	Extent to which a soldier's task is of work coordinated by feedback. The extent to which tasks performed by one soldier is dependent on the completion of other tasks by other soldiers.
b) Control	The degree of discretion a soldier or unit has in carrying out tasks in order to arrive at predetermined outcomes.
c) Routines	The degree to which the stimuli being operated on all fall into a fixed number of categories such that they are perceived as familiar.
d) Search Process When Exceptions Occur	When exceptions to the normal work situation occurs, are there logical, systematic and analytical processes for solving the problem or is the problem solved on an intuitive basis?
e) Variability of Equipment and Information	The extent to which the stimuli can be treated in a standardization fashion or has to be adjusted to continuously.

Table 8

Objective Social Structure

Factor	Definition
a) Leadership	Extent to which rankings of soldiers in a unit are congruent on three dimensions of status: individual prominence, sociability and aiding group attainment.
b) Role Style of Leader	The style of behavior of the formal or informal leader of a unit especially along the continuum from authoritarian to democratic.
c) Personality	The personality mix of the unit as characterized on any set of traits exhibited by the soldier in the unit.
d) Norms and Sanctions	The existence and nature of rules governing the behavior of soldiers in a unit and of social pressure for enforcing such rules.
e) Presence of Co-Workers	The extent to which other soldiers are physically present during work.
f) Communication Feasibility	The extent to which communication with other soldiers is feasible during work.

Table 9

Phenomenological Formal Structure

Factor	Definition
a) Perceived Structure	The degree to which individual soldier behavior is perceived to be controlled by impersonal rules and procedures.
b) Perceived Degree of Control	The degree to which soldiers perceive that higher organizational levels control decisions related to basic objectives as well as procedures and methods for performing the job.
c) Perceived Distribution of Support Functions	The degree to which support functions not related to work flow are perceived to be performed by soldiers.
d) Equity	The degree to which soldiers perceive that the formal extrinsic reward system reflects their relative worth to the organization.

Table 10

Phenomenological Task Structure

Factor	Definition
a) Subjective Expected Value of Extrinsic Outcomes	The degree to which extrinsic outcomes are perceived to vary in direct relation to the effort a soldier exerts in his tasks and the quality of the soldier's job performance and the value of those outcomes to the soldier.
b) Task Identity	The degree to which the job is perceived to provide a) a distinct sense of a beginning and end to a transformation process, b) high visibility of transformation to the worker, c) high visibility of transformation in finished product and d) transformation of considerable magnitude.
c) Autonomy	The degree to which a soldier feels personally responsible for what he accomplishes in his work.
d) Variety	The degree to which the job is perceived to require the use of a variety of valued skills and abilities.
e) Challenge	The degree to which a soldier perceives that effective task performance requires the maximum exertion of his skills and abilities.
f) Feedback	The degree to which the job is perceived to provide feedback to the job-holder as to how effectively he is performing.

Table 11

Phenomenological Social Structure

Factor	Definition
a) Leadership	The degree to which the average rankings of each soldier in a unit by soldiers in the unit are congruent on the three dimensions of status.
b) Role Style of Leader	The style of leadership of a unit leader as perceived by group numbers, authoritarian, democratic, etc.
c) Compatibility	The degree to which soldiers in a unit perceive harmony in their relations with each other.
d) Social Justice; Equity	The degree to which soldiers in a unit perceive the norms governing their behavior and the sanctions for violating those norms as proper and fair.
e) Cohesiveness	The degree to which soldiers feel attracted to the unit of which they are members.
f) Isolatedness	The degree to which soldiers feel isolated from social contact and communication with co-workers.

Table 12

Taxonomy of Performance Indicators

	CRITERION	EVALUATIVE SITUATION: Training Performance	EVALUATIVE SITUATION: Actual Job Performance
DIRECT SOLDIER PERFORMANCE DATA	QUANTITY OF WORK	TIME TO LEARN	AMOUNT PER UNIT TIME
	QUALITY OF WORK	RATINGS/MECHANISTIC	RATINGS/MECHANISTIC
SOLDIER PERSONNEL DATA	RATE OF ADVANCEMENT	IMPROVEMENT DURING TRAINING	SALARY OR PROMOTION HISTORY
	ABSENTEEISM (AWOL)	NUMBER OF DAYS	NUMBER OF DAYS
	JOB TENURE (REENLISTMENT)	TIME TO LEARN	LENGTH OF TIME
	JOB KNOWLEDGE	RATINGS ON TESTS	RATINGS ON TESTS
	ACCIDENTS AND BREAKAGE (MALFUNCTIONS)	ACCIDENT RATES	ACCIDENT RATES
JUDGMENTAL DATA	SUPERIOR JUDGMENT	RATINGS	RATINGS
	PEER JUDGMENT	RATINGS	RATINGS
	SELF JUDGMENT	RATINGS	RATINGS

Figure 1

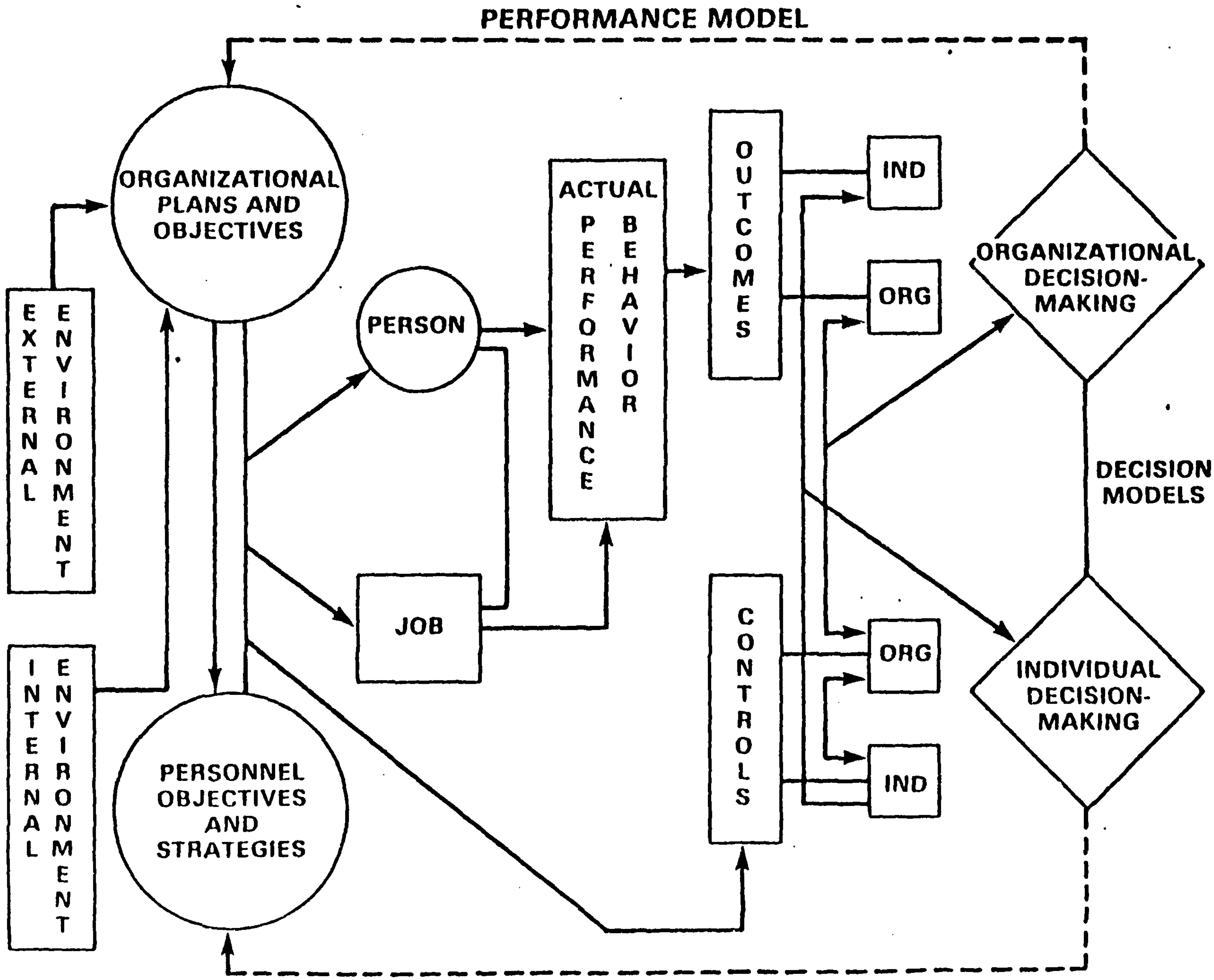

PERFORMANCE MODEL
EXTERNAL ENVIRONMENT
INTERNAL ENVIRONMENT
ORGANIZATIONAL PLANS AND OBJECTIVES
PERSONNEL OBJECTIVES AND STRATEGIES
PERSON
JOB
ACTUAL PERFORMANCE BEHAVIOR
OUTCOMES
CONTROLS
IND
ORG
ORG
IND
ORGANIZATIONAL DECISION-MAKING
INDIVIDUAL DECISION-MAKING
DECISION MODELS

Figure 2

INFLUENCES ON PERFORMANCE

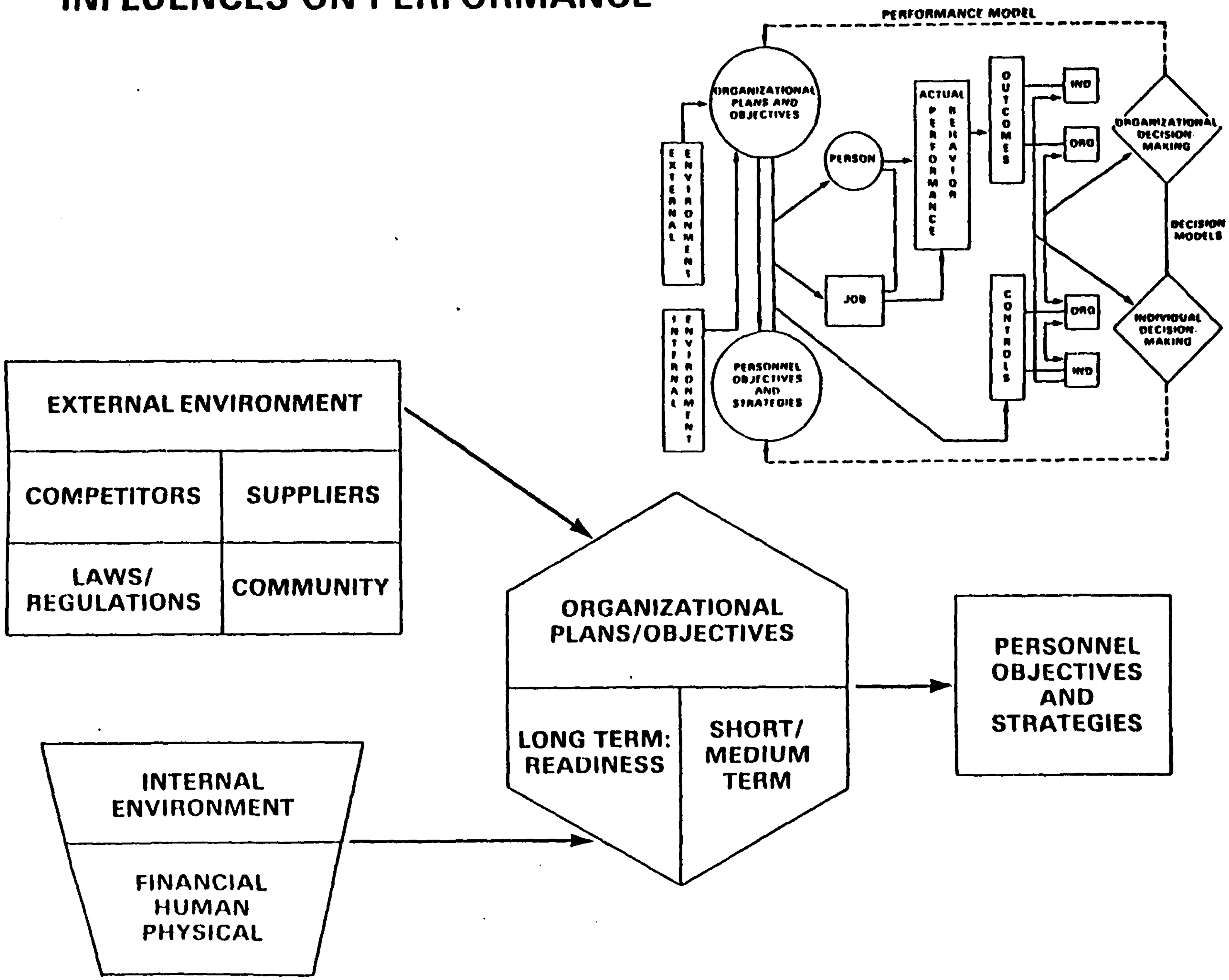
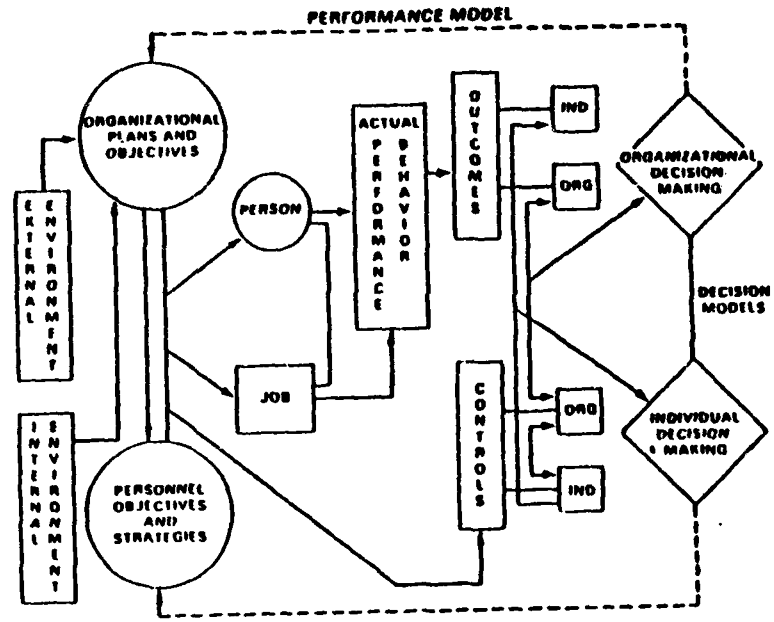


Figure 3

PERSON COMPONENT

COGNITIVE CONTROL FUNCTIONS



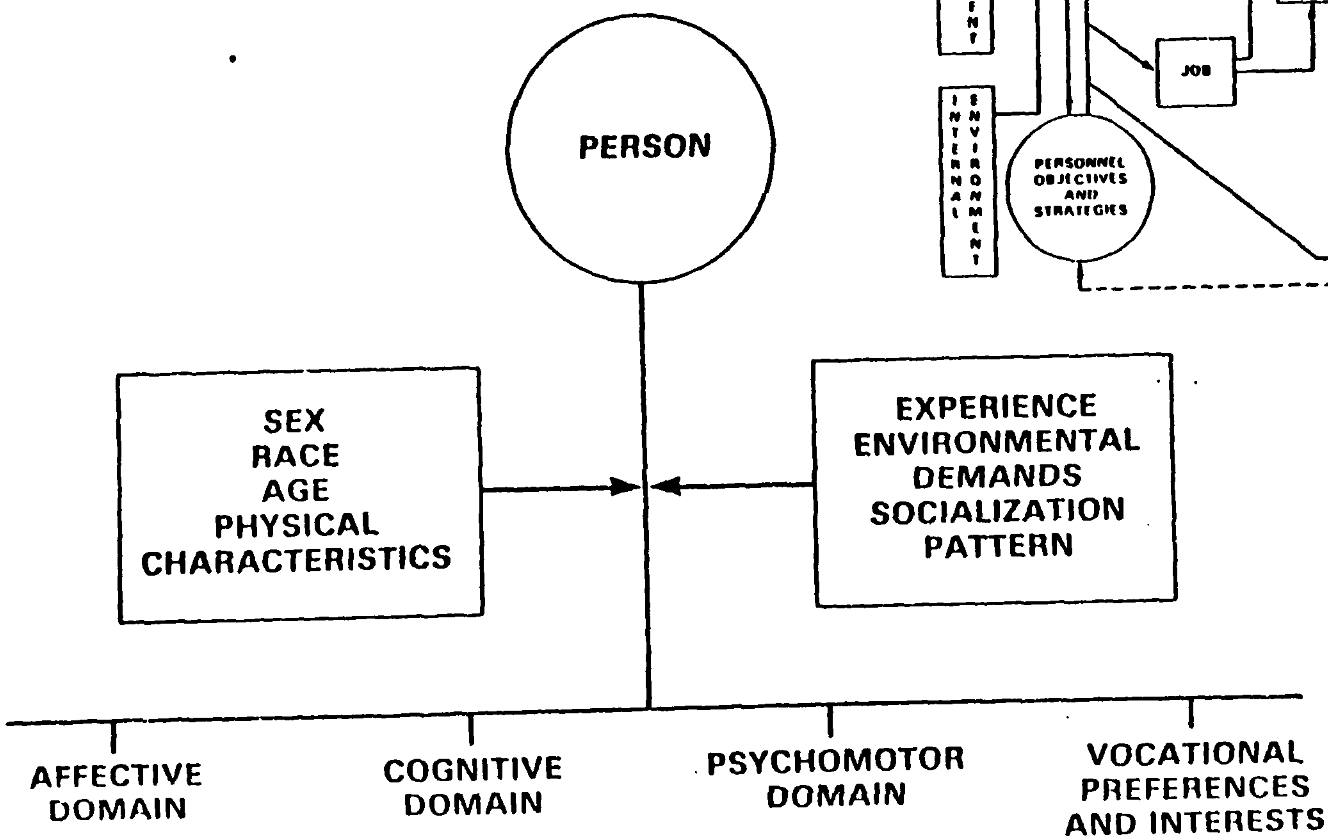
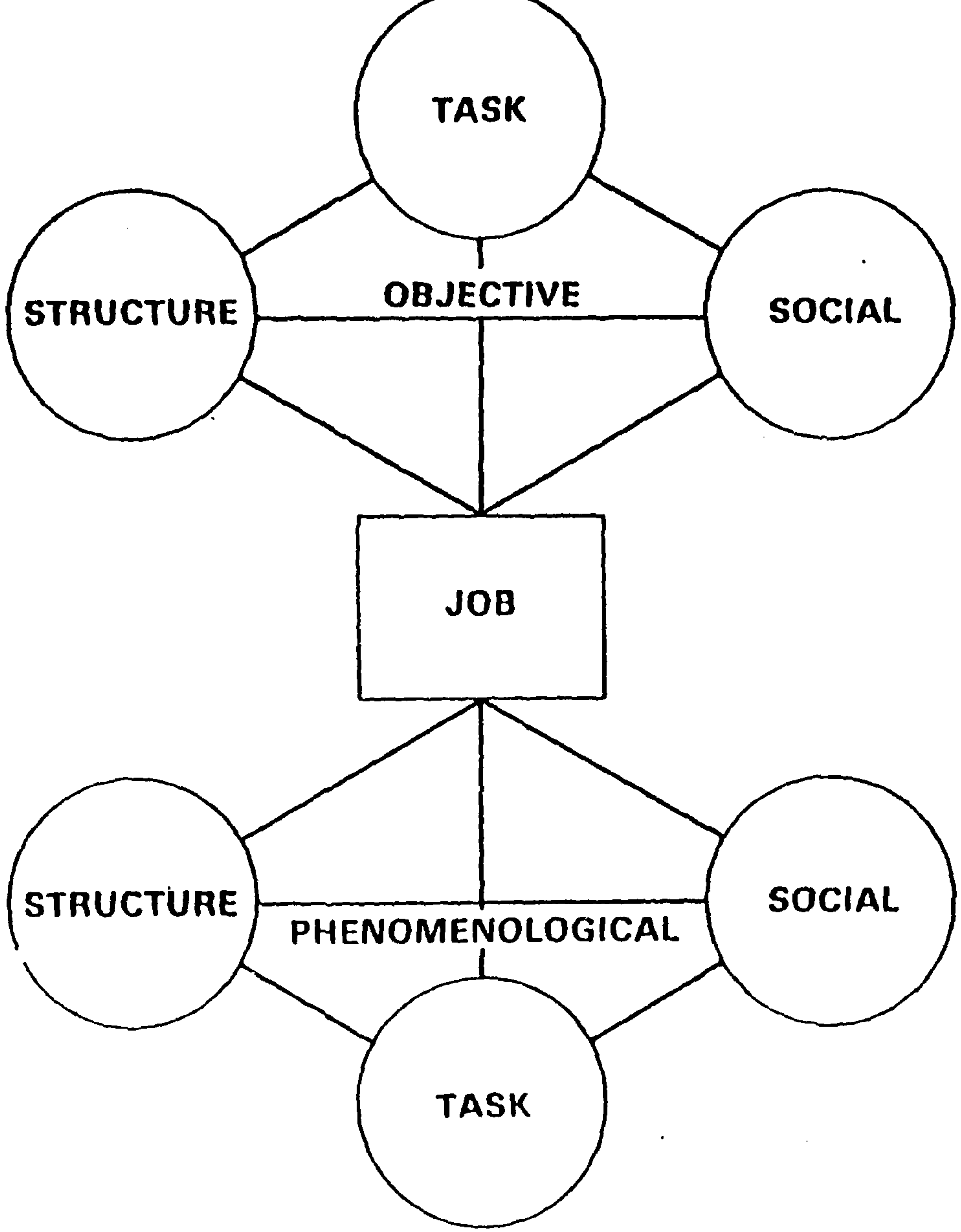
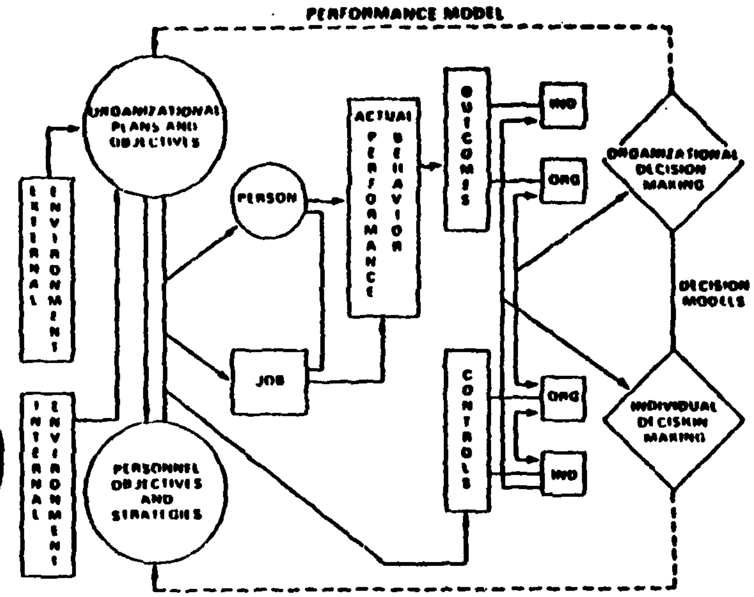


Figure 4


TASK
STRUCTURE
OBJECTIVE
SOCIAL
JOB
STRUCTURE
PHENOMENOLOGICAL
SOCIAL
TASK


PERFORMANCE MODEL
ORGANIZATIONAL PLANS AND OBJECTIVES
EXTERNAL ENVIRONMENT
INTERNAL ENVIRONMENT
PERSONNEL OBJECTIVES AND STRATEGIES
PERSON
JOB
ACTUAL PERFORMANCE BEHAVIOR
OUTCOMES
CONTROLS
IND
ORG
ORG
IND
ORGANIZATIONAL DECISION MAKING
DECISION MODELS
INDIVIDUAL DECISION MAKING

REFERENCES

Anastasi, A. Differential psychology (3rd ed.) New York: MacMillan, 1958.

Anastasi, A. Evolving trait concepts. American Psychologist, 1983, 38 (2), 175-184.

Atkinson, J. W. & Birch, D. Introduction to Motivation. New York: D. Van Nostrand Company, 1978.

Bandura, A. The self system in reciprocal determinism. American Psychologist, 1978, 33, 344-358.

Browne, J. A., & Howarth, E. A comprehensive factor analysis of personality questionnaire items: a test of 20 positive factor hypotheses. Multivariate Behavioral Research, 1977, 12, 399-427.

Campbell, J. P., Dunnette, M. D., Lawler, E. E. III & Weick, K. E. Jr. Managerial behavior performance, and effectiveness. New York: McGraw-Hill, 1970.

Campbell, J. P., & Pritchard, R. D. Motivation theory in industrial and organizational psychology. In M. D. Dunnette, Ed. Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Carrol, J. B. Psychometric tests as cognitive tasks: a new structure of intellect. In L. B. Resnick (Ed.), The nature of intelligence. New Jersey: Lawrence Erlbaum Associates, 1976.

Coleman, C. J. Personnel: an open system approach. Mass.: Wintrop Publishers, Inc., 1979.

Cronbach, L. J. & Gleser, G. C. Psychological tests and personnel decisions. Urbana, Illinois: University of Illinois Press, 1965.

Dunnette, M. D. Aptitudes, abilities, and skills. In M. D. Dunnette (Ed.) handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Dunnette, M. D. Basic attributes of individuals in relation to behavior in organizations. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Dunnette, M. D. Performance equals ability and what? (Technical Report 4009) Office of Naval Research March, 1973.

Ekstrom, R. B. Cognitive factors: some recent literature (Technical Report No. 2, ONR Contract N00014-71-C-0117, NR 150-329). Princeton, N. J.: Educational Testing Service, 1973.

Ekstrom, R. B., French, J. W., Harman, H. H. Manual for kit of factor-referenced cognitive tests. Princeton, N. J.: Educational Testing Service, 1976.

Fleishman, E. A. The structure and measurement of physical fitness. Englewood Cliffs, N. J.: Prentice Hall, 1964

Fleishman, E. A. On the relation between abilities, learning, and human performance. American Psychologist, 1972, 27, 1017-1032.

Fleishman, E. A. Structure and measurement of psychomotor abilities. In R. N. Singer (Ed.), Psychomotor domain: Movement behavior. Philadelphia: Lea & Febiger, 1972.

Fleishman, E. A. Systems for describing human tasks. American Psychologist, 1982, 37 (7), 821-834.

Ghiselli, E. E. The validity of occupational aptitude tests. New York: Wiley, 1966.

Golembiewski, K. T. Small groups and large organizations. In J. G. March (Ed.), Handbook of organizations. Chicago: Rand McNally, 1965.

Hackman, J. R. and Lawler, E. E., III Employee reactions to job characteristics. Journal of Applied Psychology, 1971, 55, 259-286.

Holland, J. L. Vocational preferences. In M. D. Dunnete (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Kane, J. S. The impact of organizational properties on performance appraisal: a review of the research. Doctoral Dissertation, University of Michigan, 1975.

Kress, G., Sharon, B., & Basan, D. Work motivation: theory and practice. (Technical Report 553) Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, September 1981.

Lawler, E. E. III Motivation in work organizations. Monterey, California: Brooks/Cole Publishing Company, 1973.

Lawler, E. E. III, Hackman, J. R. and Kaufmann, S. Effects of job redesigns: a field experiment. Journal of Applied Social Psychology, 1969, 53, 467-471.

Lynch, B. P. An empirical assessment of Perrow's technology constructs. ASQ, 1974, 19, 338-356.

Naylor, J. D., Pritchard, R. D., Ilgen, D. R. A theory of behavior in organizations. New York: Academic Press, 1980.

Perrow, C. A framework for the comparative analysis of organizations. American Sociological Review, 1967, 32, 194-208.

Pugh, A. S., Hickson, D. J., Hinings, C. R. and Turner, C. Dimensions of organization. ASQ, 1968, 13, 65-105.

Steers, R. M., & Porter, L. M. Motivation and work behavior. New York: McGraw-Hill Book Company, 1979.

Sternberg, R. J. Intelligence, information processing and analogical reasoning: The componential analysis of human abilities Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Strauss, G. & Sayles, L. R. Personnel: The human problems of management. Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Strong, E. K., Jr. Vocational interests 18 years after college. Minneapolis: University of Minnesota Press, 1955.

Tannenbaum, A. S. Controling organizations. New York: NcGraw-Hill, 1968.

Theologus, G. C., Romashko, T., & Fleishman, E. A. Development of taxonomy of human performance: A feasibility study of ability dimensions for classifying human tasks. JSAS Catalog of Selected Documents in Psychology, 1973, 3, 25-26.

Tiffin, J. & McCormick, E. J. Industrial psychology. (5th ed.). Englewood Cliffs, N. J.: Prentice-Hall, 1965.

Tyler, L. E. The psychology of human differences. (Third ed.) New York: Appleton-Century-Crofts, 1965.

VI. MOS TASK DESCRIPTIONS

Patrick Ford (HumRRO)
John P. Campbell (HumRRO)

Several initial phases of Project A criterion development activities are dependent on both: (a) the generation of a set of task descriptions for each focal MOS and (b) the identification of job behaviors which are not specific to a particular MOS but which are critical for effective performance in general. That is, we need job descriptions that are comprehensive in terms of covering performance factors that are both common to all MOS and MOS specific.

Two different methods are being used to provide this pool of information. One makes principal use of the Army's existing sources of task descriptions for each MOS. The second uses the critical incident method to generate a large pool of critical performance behaviors for specific MOS and that cut across MOS. The two courses are being used for somewhat different purposes and each is discussed briefly below.

MOS-Specific Task Descriptions

The task descriptor "item banks" are being used in the following ways:

(1) To construct knowledge tests for the assessment of training achievement a stratified sample of task descriptors was drawn for each MOS. These tasks were then judged in terms of their importance for the MOS, their relative tendency for error, and the degree to which success on the task is dependent on problem solving vs. application of specific procedures. The task descriptions and their associated judgments will then serve as one basis for the generation of knowledge test items.

(2) The task descriptor item bank is also the principal starting point for the construction of hands-on performance measures to be described in a later section of this report. Several procedural steps are being used to select tasks from this pool so that the resulting sample of tasks is the most appropriate for standardization as performance test items.

(3) A computerized content analysis procedure has been applied to the task item banks as a check against whether any important performance factors are being missed as the various steps in our criterion development efforts proceed.

(4) Later in the project, the task descriptions will be used to develop a common set of descriptors that can be used to describe all MOS and cluster them empirically into homogeneous job families.

Generation of the Task Lists

Procedures for systematic job and task analyses were developed in the 1950s, largely as the result of research conducted by the U.S. Air Force (e.g., Miller, 1953; Van Cott, Berkun & Purifoy, 1955; Christal, 1969, 1974). These procedures have been widely used by the Army in support of the systems engineering of individual training, as articulated in the Instructional Systems Development (ISD) approach (e.g., TRADOC Pam 350-30, 1975).

The Army's use of job analysis procedures has tended to be training oriented. That is, the information provided has been used largely to help make decisions about the need for and content of training given in AIT and other specialized courses. Although other activities have also reflected and benefited from the knowledge gained from task analytic work (e.g., designing job aids, developing skills qualification tests, constructing selection batteries, preparing job-related handbooks and manuals), the primary thrust for task analytic information has come from the various proponent schools. The training emphasis of task analytic work has important implications for the way in which we use these materials in Project A.

The task data collection procedure most favored by the Army and other services is the job inventory, a standardized and self-administered checklist. It is the method of choice because observation, interviews, technical conferences, and open-ended questionnaires have too many limitations to be useful in any large-scale data collection program (Rupe, 1956). It is the approach currently lying at the heart of the Army Occupational Survey Program (AOSP).

The checklist contains items describing a variety of duties and tasks related to a given MOS. These items are drawn from information already known about the job, primarily from existing documentation and from SME. (Guidelines for constructing a job inventory have been described by March and Archer [1967] and are included in TRADOC Pam 351-4.) Soldiers who are incumbents of the target MOS are instructed to check the duties and tasks that they perform, and to rate them on one or more dimensions such as frequency of performance and the related amount of time that they require to perform. Typically, several hundred people in each MOS are surveyed.

A quantitative assessment of job activities can be obtained from a statistical analysis of the checklist responses using the Comprehensive

Occupational Data Analysis Program (CODAP). CODAP can be used to rank-order duties and tasks in accordance with the percentage of soldiers who perform them and the relative time spent on each. This information, when combined with a number of other factors, is often used to select the critical tasks that will be the focus of training and evaluation activities. Survey Reports, prepared by the U.S. Army Soldier Support Center, summarize the results of the surveys for each MOS. These reports include valuable information on the structure and nature of the MOS by skill level. A major use that is made of these analyses is to determine which of the tasks comprising an MOS should be taught in a formal school setting (e.g., in AIT rather than on-the-job). Once these tasks have been selected, they are subjected to more detailed scrutiny to determine (1) how they are best taught, and (2) the nature of the requirements they impose on the trainee.

For Project A purposes two sources of job analytic information were consolidated for each selected MOS at Skill Level (SL) 1. The first source was the Soldier's Manual (SM); the second was CODAP frequency data from the AOSP.

The SM specifies the tasks that by doctrine are critical to the soldier's job performance at a given Skill Level. These critical tasks represent a subset of the tasks a soldier could perform. The immediate purpose of the SM is to guide training on the critical subset. Although the procedure to identify the critical subset for a particular MOS varies by proponent (i.e., the unit responsible for training in a particular MOS), task selection for the SM is typically a high visibility activity that involves the highest levels of a proponent's command. In addition to the tasks in the MOS SM, 71 common tasks have been designated as critical to the job performance of all SL1 soldiers. These tasks are listed in the Soldier's Manual of Common Tasks (FM21-2).

As just noted, CODAP is a description of job activities based on a checklist questionnaire survey of job incumbents. The checklist contains items describing a variety of duties and tasks related to the MOS. The items have been drawn from job analysis materials and subject matter experts. Although the items are intended to reflect job content, by virtue of the way they are generated the items reflect the intended content of the MOS as well as the actual content.

The consolidation of SM and CODAP serves three purposes for selecting task descriptions to serve as a basis for criterion development. As follows, it:

(1) Describes domain of the soldier's job,

(2) Determines frequency for critical tasks,

(3) Confirms completeness of SM.

Describes domain of the soldier's job. A soldier's job consists of tasks and activities he or she is trained to perform (doctrine) at a particular skill level, and tasks and activities that he or she actually performs on the job. Although there really is a major overlap, differences exist between the two sets. For example, in field environments there is seldom the sharp distinction between different skill levels that exists doctrinally. Job doctrine is best reflected in SM while CODAP generally provides a fuller picture of field requirements. Integrating the two documents gives a more complete view of the domain that will be experienced by the SL1 soldier.

Determines frequency for critical tasks. The Soldier's Manual does not provide data on which tasks are most widely performed within a skill level; however, that information is available in CODAP. Having that information protects against randomly selecting, for example, a 13B SM task like "Operate the intercommunications systems," which is apparently performed by only 3 percent of cannoneers.

Confirm completeness of SM. Checking the job descriptions from CODAP against job descriptions from the SM also insured that potentially critical tasks were not lost. For example, during the transition to centralized common task management, the task "Engage targets with an M16" was not included in any of the MOS SM or in the Soldier's Manual of Common Tasks. There also may have been shifts in analysts' assumptions about the scope of tasks which, in conjunction with changes in the MOS, have caused potential critical tasks to be overlooked. For example, the 13B activity "Clean cannon tube and chamber" may have at one time been assumed to be part of preventive maintenance checks and services (PMCS). It does not now appear to be part of PMCS or any other SM task even though 74 percent of cannoneers report doing it.

Next Steps

At this point there was available a pool of task descriptor items for each MOS. How this item bank was subsequently used was different depending on whether the objective was to construct MOS-specific performance measures, to develop MOS-specific training measures, or to develop a general taxonomy of task performance factors that can be used to describe all MOS. These alternative uses of the task item bank will be described in subsequent sections of this report.

Critical Incident Descriptions

The job behavior descriptions being generated by the critical incident method serve the following purposes:

(1) They are a major source of information for the identification and explication of the factors that define job performance and effectiveness, both general and MOS-specific.

(2) They are the primary means by which rating scale measures of general and MOS-specific performance factors will be constructed.

The critical incident procedure involves the following general steps:

(1) Workshops comprised of 10-20 supervisors (NCO and/or officers) are asked to generate specific examples of job behaviors for enlisted personnel that are "critical" in terms of reflecting positive or negative aspects of performance.

(2) The specific incident descriptions are then categorized by a panel of judges into categories that seem to reflect the major underlying performance factors.

(3) Another group of judges then "retranslates" the specific incidents by assigning them to the performance category in which they best fit. To the extent that

this retranslation can be done reliably the category system is a meaningful one.

(4) At the same time the incidents are retranslated they are also judged, or scaled, in terms of the level of effective or ineffective performance they represent.

Next Steps

To date, a number of workshops for both MOS-specific performance and general effectiveness have been carried out. While neither effort is scheduled for completion as of this date, the preliminary results wil be reported in the following sections that are appropriate to MOS-specific and general effectiveness criterion development.

VII. DEVELOPMENT OF TRAINING MEASURES

John N. Joyner (HumRRO)
Rebecca L. Oxford-Carpenter (ARI)

General Purpose

The general purpose of our research on training criteria is to generate information about training performance which can be used in the validation of initial predictors and in the prediction of first-tour and second-tour performance in the Army. To accomplish this purpose, existing measures of training performance are being analyzed and evaluated, new measures are being developed where needed, and composite sets of predictor and criterion measures will be assembled. As job performance surrogates, training measures can serve to reduce the time required for predictor validations from years to months. When used to predict subsequent performance, training measures have the potential to increase the accuracy of classification into MOS over that obtained by the use of preinduction predictors alone. Both the extent to which training measures can be used as surrogates for ultimate job performance criteria and the degree of incremental validity obtained by including training success as a predictor itself will be assessed during the course of the project.

General Procedure

The steps to be followed for training criterion development over the entire course of the project are summarized below. A more detailed discussion of specific accomplishments during the first year will follow.

(1) Literature review. We have reviewed the literature on the issues and methods of evaluating student achievement.

(2) Evaluation of existing measures. Existing criterion measures are being examined based on studies of information available in school records for the FY81/82 and FY83/84 cohorts and already entered in the LRDB by ARI. Additional procedures are being set up to obtain an additional sample of training records from the 83/84 cohort and to enter them in the LRDB.

(3) Analysis of Army training and evaluation procedures. The training school site visits, corresponding to the 19 focal MOS, are being used to collect information on training objectives, training content, and evaluation methods. The tests currently used in Army schools are being examined in discussions with SME to determine the relation of their content to training requirements and will be examined statistically to determine the adequacy of their measurement characteristics.

(4) Revision/construction of new comprehensive knowledge tests. To provide improved measures to serve both as criteria of school performance and as predictors of job performance, new comprehensive knowledge tests will be developed in each of the 19 focal MOS. Knowledge will be sampled based on commonality across the MOS tasks, on estimates of frequency of error in performance, and on representation of two classes of task components: those requiring the application of procedures and those requiring the generalization of information.

(5) Development of prototype measures. To represent components of training performance not represented by existing measures or newly developed job knowledg tests, prototypes of new indexes will be derived from existing measures.

(6) Identification of training-relevant and job-relevant test content. To provide a basis for interpreting predictive relationships between the new knowledge tests and subsequent job performance, the relevance of the knowledge test items to training and job content will be determined in two ways. First, training relevance will be determined empirically by comparing the performance of entering trainees and graduating trainees; job relevance will be determined by comparing the performance of graduating trainees and job incumbents. Second, job relevance and training relevance will be determined judgmentally by trainers at Army schools.

(7) Develop predictor and criterion composites of school measures. School measures determined to have adequate reliable variance and content validity will be assembled into integrated sets to serve as criteria for validating initial predictors and as predictors of MOS-specific and Army-wide performance.

First Year Activities

The project activities during the first year of training criterion development have concentrated on: (a) a review of the literature, (b) the analysis of current school measures, (c) the documentation and analysis of training objectives and training content, and (d) the first steps in the development of comprehensive job knowledge tests.

The development of training measures is being accomplished for 19 selected MOS. These MOS and the installations that have been, or will be, visited are indicated in Table 12. The initial round of site visits to training installations is currently about half complete.

Analysis of Current School Measures

Preliminary identification of current school measures as criteria of training performance and as predictors of subsequent job performance is being accomplished through analyses of test score distributions, a review of the test construction and test scoring process, and a comparison of the measures' coverage of training and job content. The analyses will begin once all site visits are completed.

Visits have been completed to all of the Army schools in Table 12 to arrange for transmittal of training performance data to the Project A data base, for trainees during the period 1 Aug 83 - 31 Jul 84. Arrangements

Table 12

Development of Training Measures - MOS & Training Installations

MOS	Title	Installation
05C	Radio TT Operator	*USA Signal School, Ft. Gordon, GA
11B	Infantryman	USA Infantry Center, Ft. Benning, GA
12B	Combat Engineer	USATC, Ft. Leonard Wood, MO
13B	Cannon Crewman	USA Field Artillery Center, Ft. Sill, OK
16S	MANPADS Crewman	*USA Air Defense Center & School Ft. Bliss, TX
19 E/K	Tank Crewman	*USA Armor Center, Ft. Knox, KY
27E	Tow/Dragon Repairer	USA Missile & Munitions Center & School Redstone Arsenal, AL
51B	Carpentry/Masonry Spec.	USATC, Ft. Leonard, MO
54B	Chemical Operations Spec.	USA Chemical Center & School Redstone Arsenal, AL
55B	Ammunition Spec.	USA Missile & Munitions Center & School Redstone, Arsenal, AL
63B	Vehicle & Generator Mech.	*USATC, Ft. Dix, NJ *USATC, Ft. Jackson, SC *USATC, Ft. Leonard Wood, MO
64C	Motor Transport Opr.	*USATC, Ft. Dix, NJ
67N	Utility Helecopter Rpr.	USA Aviation Center & School, Ft. Rucker, AL
71L	Admin. Spec.	*USATC, Ft. Jackson, SC
76W	Petroleum Supply Spec.	USA Quartermaster School, Ft. Lee, VA
76Y	Unit Supply Spec.	USA Quartermaster School, Ft. Lee, VA
91B	Medical Care Spec.	Academy of Health Sciences, Ft. Sam Houston, TX
94B	Food Service Spec.	USA Quartermaster School, Ft. Lee, VA *USATC, Ft. Dix, NJ *USATC, Ft. Jackson, SC
95B	Military Police	USA Military Police School, Ft. McClellan, AL

*Visits to these installations and courses were made during July and August, 1983, for initiation of training measures development.

have also been made to obtain training performance data for (a) the 400 most recent trainees in the FY82 cohort in MOS 05C, 54E, 63B and 71L, and (b) several MOS not included in the 19 MOS selected for development of new training measures (MOS 11C, 11H, 11BC2, 21L, 24K, 27F, 27G, 31J, 31M, 36C, 36K, 54C, 55G, 67V, 71P, 72E, 76P, 76X, 91D, 92B, 93H, and 93J) for the period 1 Aug 83 - 31 Jul 84. The types of training performance information, e.g., test format, number of scores, available for each MOS are given in Table 13.

The analyses of training coverage in existing measures is being accomplished in two steps. Training/test content overlap is identified by comparing test elements and training outline elements in a training/test matrix. Subsequently, apparent gaps in measurement coverage are discussed in interviews about training performance measurement with subject matter experts (SME) at Army schools (see Appendix D for a copy of the Training Measurement Interview Guide).

Table 14 is a condensed version of the matrix derived from the program of instruction (POI) for the 05C Radio Teletypewriter Operator Course. The POI is divided into Task Cluster Annexes, each of which is subdivided into a number of lessons, one or more of which is a test. The table shows which test covers the material in each of the lessons, what TRADOC designation the school has assigned to the test (E1- hardware-oriented performance test, or E2- non-hardware-oriented performance test) and whether it was made available (+) for examination. The comments indicate why certain lessons are not formally tested (NT) and why certain of the tests are not available (-).

Table 13

Training Measures to be Entered in LRDB
13 Posts
36 MOS

		Hands-On Tests					
Post	MOS	Numerical Scoring: Number of Tests	Pass-Fail Scoring: Number of Tests/ Number of Items	Remediation and Retest Information	Written Tests with Numerical Scores	Course Pass/Fail Only	Other
Ft. Benning	11B10		36/290				
	11C10	1	36/290				
Ft. Bliss	16S10	1	-/46	x			8[a]
Ft. Dix	63B10				9		
	64C10					x	
	94B10				5		
Ft. Gordon	05C10		33/99	x			
Ft. Jackson	63B10		89/-	x			
	71L10		32/-	x			
	76Y10		3/-	x			
	94B10	20	1/-		5		2[b]
Ft. Knox	19D10		3/69	x			
	19E10		3/95	x			
	19K10		3/86	x			
Ft. Lee	76P10	16			1		
	76W10	11					
	76X10	17					
	76Y10					x	
	94B10					x	
Ft. Leonard Wood	12B10		11/14	x			
	51B10		22/22				
	63B10		12/117	x			117[c]

Table 13 (Continued)

Hands-On Tests

Post	MOS	Numerical Scoring: Number of Tests	Pass-Fail Scoring: Number of Tests/ Number of Items	Remediation and Retest Information	Written Tests with Numerical Scores	Course Pass/Fail Only	Other
Ft. McClellan	54E10	2		x	9		
	54C10	5	36/290	x	7		
	95C10		6/-		7		
Redstone Arsenal	27E10		12/-	x	13		
	27F10				9		15[d]
	27G10					x	16[d]
	21L10				5		23[d]
	24H10						12[d]
	55B10		4/53	x	1		
	55G10						15[d]
Ft. Rucker	67N10	1	10/-	x	7		
	67V10	1	10/-		7		
	71P10	1	5/-		8		
	93H10	1	2/-		3		1[d]
	93J10	1	2/-		3		1[d]
Ft. Sam Houston	91B10	9	6/6	x			
	91D10	5			5		
	92B10	16			1		7[d]
Ft. Sill	13B10	3	42/42	x			

[a]time to complete instructional modules
[b]rating of lab performance
[c]time and trials to learn individual tasks
[d]module numerical scores derived from written & performance tests

Table 14

Condensed Matrix: Training and Tests
MOS 05C – Radio Teletype Operator – Ft. Gordon, GA

Task Cluster Annex/Lesson	Tested in Lesson	Test Type	Test Availability	Comment
A. General Subjects				
A01 – Orientation to School and Course				
A02 – Alphanumerics	NT			The alphanumeric phonetic equivalents are used intensive & continuously throughout course
B. Keyboard Techniques				
B01 – Using Keyboard Techniques (typing)	B02	E1	+	
B02 – Test			–	Touch-type on teletypewriter keyboard
C. Radiotelephone Procedures				
C01 – Opening & Closing RT Using AN/PRC-77	C07	E1	+	
C02 – Preparing RT Messages for Transmission	C07	E1	+	
C03 – Transmitting Messages in RT Net	C07	E2	+	
C04 – Maintaining Circuit Log & Operations	NT			Skill evaluated during lesson and used thereafter in course
C05 – Install & Operate AN/VRC-12 Series	C07	E1	+	
C06 – Electronic Counter-Countermeasures	C07	E2	+	
C07 – Test				Install and operate FM radio equipment
D. Teletypewriter Communications Procedure				
D01 – Establishing RTT Communications	D04	E2	+	
D02 – Preparing RTT Messages	D04	E2	+	
D03 – Transmitting Requests & Responses	D04	E2	+	
D04 – Test				Teletypewriter communications procedures
E. Communications Terminal AN/UGC-74				
E01 – Introduction				
E02 – Prepare UGC-74 for Normal Operation	E10	E1	+	

(Continued)

Table 14 (Continued)

Condensed Matrix: Training and Tests
MOS 05C — Radio Teletype Operator — Ft. Gordon, GA

Task Cluster Annex/Lesson	Tested in Lesson	Test Type	Test Availability	Comment
E03 — Transmitting & Receiving Messages in Receive Only (RO) & Keyboard Send/Receive (KSR) States	E10	E2	+	
E04 — Intelligent Communications Terminal (ICT)	E10	E1, E2	+	
E05 — Setting Parameters for UGC-74 in ICT State	E10	E1	+	
E06 — Use Edit Subcommands to Compare & Edit Messages	E10	E1	+	
E07 — System Commands Used in ICT Operational State	E10	E1	+	
E08 — Transmitting & Receiving Messages in ICT State	E10	E1	+	
E09 — PMC&S on UGC-74	E10	E1	+	
E10 — Test				Install, operate, & perform PMC&S on UGC-74
F. Single Sideband RTT Equipment				
F01 — Preparing Equipment for Operation	F03	E1	+	
F02 — Operating & Trouble-shooting RTT Set AN/GRC-142/122	F03	E1	+	
F03 — Test				Install, operate, & troubleshoot AN/GRC-142/122
G. Cryptographic Devices				
G01 — Regulations & Procedures for Safeguarding Crypto Information	NT			Regulations & procedures are followed throughout annex
G02 — Preparing TSEC/KW7 for operation	G04	E1	–	Classified
G03 — Installing Speech Security Equipment	G04	E1	–	Classified
G04 — Test				Install & operate TSEC/KY8 & TSEC/KY38 (classified)
G05 — Installation & Operation of Vinson	G07	E1	–	Classified
G06 — Installation & Operation of Net Control Device (Vinson)	G08	E1	–	Classified
G07 — Test				Install & operate Vinson
G08 — Test				Install & operate net control device (Vinson)

(Continued)

Table 14 (Continued)

Condensed Matrix: Training and Tests
MOS 05C — Radio Teletype Operator — Ft. Gordon, GA

Task Cluster Annex/Lesson	Tested in Lesson	Test Type	Test Availability	Comment
H. Ancillary Equipment				
H01 — Install Ground Plane Antenna RC-292	H03	E1	+	
H02 — Install Antenna Group AN/GRA 50 with Mast Base AB-155	H03	E1	+	
H03 — Test				Install tactical FM & RATT antennas
H04 — Generator Set 5KW/AC	H06	E1	+	
H05 — PM & Operator Trouble-shooting on Generator Set 5KW/AC	H06	E1	+	
H06 — Test				Install, operate, & perform operator PM & troubleshooting on generator set 5KW/AC
I. Tactical Evaluation & Skills Training				
I01 — Skills Training	NT			Field exercises combining all of above lessons
I02 — End of Course Comprehensive Test		E1, E2	—	Test covers all lessons (classified)

Results

To date (9/15), instructors and supervisors in the following 11 courses have been interviewed:

05C - Radio Teletype Operator	Ft. Gordon, GA
16S - Manpads Crewman	Ft. Bliss, TX
19E - Tank Crewman	Ft. Knox, KY
19K - M-1 Crewman	Ft. Knox, KY
63B - Light-Wheeled Vehicle Mechanic	Ft. Dix, NJ
63B - Light-Wheeled Vehicle Mechanic	Ft. Jackson, SC
64C - Motor Transport Operator	Ft. Dix, NJ
71L - Administrative Specialist	Ft. Jackson, SC
76Y - Unit Supply Specialist	Ft. Jackson, SC
94B - Food Service Specialist	Ft. Dix, NJ
94B - Food Service Specialist	Ft. Jackson, SC

The interview was concerned primarily with trainee progress and achievement measures in each course. Much of the information gathered is summarized in Table 15. There was surprising unanimity among the courses in these matters. All of the courses are group paced (GP), except for the self-paced (SP) 05C course, the mostly self-paced 16S course, and the lock-step (LS) 19E and 19K courses. Since group paced and lock step are virtually indistinguishable modes of procedure, the 05C and 16S courses are the only real exceptions and both are scheduled to become group paced in the near future.

Table 15

Summary of Interviews Held With Instructors and Supervisors of Selected Army Courses

Items of Information	05C	16S	19E	19K	63B Dix	63B Jackson	64C	71L	76Y	94B Dix	94B Jackson
Instructional mode	SP	SP/LS	LS	LS	GP	GP	GP	GP	GP	GP	GP
Type of testing done	E1,E2	E1,E2,E3	E1,E2,E3	E1,E2,E3	E1,E2,E3	E1,E2,E3	E1,E2,E3	E2	E2,E3	E1,E2,E3	E1,E2,E3
Tests pass to graduate	all	all	all	all	all	all	all	all	all	all	all
Passing score (E3)	–	100%	–	–	70%	–	70%	–	100%	70%	62.5-70%
Strictly timed tests	some	some	some	some	none	none	1	none	few	–	none
Trainees finish within time limit	all	all	all	all	95%	95%	all	96%	95%	NA	–
Coverage by EOCT	all	sample	all	all	sample	sample	all	all	sample	sample	sample
Final "grade"	P/F	PI	P/F	P/F	P/F	P/F	P/F	P/F	P/F	P/F	P/F
Same test for all	yes	yes	yes	yes	no	no	yes	yes	yes	no	no
Tests can be missed	no	no	yes[1]	yes[1]	no	no	no	no	no	no	no
Final grade calculation	NA	OT/ET[2]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Passing score	NA	PI≤1.25	NA	NA	NA	NA	NA	NA	NA	NA	NA
Retests recorded	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
No. of tests allowed	3	3	3	3	3	3	3	3	3	3+	3
Interval before retesting	none	variable	none	none	none	none	retrain	none	24 hr. or less	none	none
Challenges allowed	yes[3]	no	no	no	no	no	no	no	no	no	no
Tests monitored		yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Prompting permitted	no	no	no	no	no	no	no[4]	no	no	no	no[4]

[1] Medical disability
[2] Observed time/expected time
[3] Extremely rare
[4] Minor prompting allowed

Training performance data of considerable detail, e.g., task level or test item information, were often found to be recorded at Army schools. These data, generally not forwarded to centralized files, are not routinely available for research purposes. However, at most of the schools it was possible to make arrangements for these raw data to be forwarded to the Project A data base (LRDB) manager. The format in which detailed training performance data is available varies by installation. It may be maintained in the TRADOC Educational Data System (TREDS) system or on a local computer; it may be recorded in individual or class roster hard copy records; or it may be available only on the original individual test forms and score cards.

In line with the Army's current emphasis on competency training and testing, most tests now employed in these courses are hands-on or performance tests: TRADOC designation E1 or E2, rather than E3, pper-and-pencil, verbally oriented achievement tests. Although 9 of the 11 courses do have at least one written (E3) test, such tests constitute a minor portion of the testing programs. Because line between E2 and E3 tests is a fine one, some of the tests reported as E3 might be better classified as E2.

Courses are designed to make achievement or competence cumulative throughout the course. In all 11 courses, all trainees must pass, successively, all of the tests given in order to graduate. Only in 19E and 19K is a trainee allowed to skip a test, an then only if he is temporarily disabled. All of the performance tests (E1 and E2) given are scored on a pass/fail ("GO/NO GO") basis, although the criteria for pass/fail vary from course to course. The written (E3) tests in the 64C and 94B courses require 70 percent to pass; those in the 16S and 76Y course, 100 percent. In the other courses, the written (E3) tests are of minimal significance.

Time stress has been eliminated from most of the testing. The time limits on tests are strictly enforced only when time is the essence of the performance: in the 05C course for IMC (International Morse Code) for typing, and for preparing various pieces of equipment for operation, which must be done quickly under combat conditions; for the "10-Second Drill" in the 16S course; for the nuclear/biological/chemical drills in the 19E and 19K courses. The instructors interviewed generally believe that, for most tests, 90 percent or more of the trainees finish the required tasks well within the time limits of the tests, while the remainder would not be able to complete the test satisfactorily even if time were not a factor.

The coverage of the course by the tests varies. Virtually everything taught in every course is tested during the course, usually at the end of the module or annex. The nature of the course material and time, however, determine whether an end-of-course test (EOCT) will cover everything in the entire course. A relatively limited number of distinct tasks are taught in the 71L course for administrative specialists; this makes it quite possible to devise a test of reasonable length covering all tasks. The 94B course for food service specialists, on the other hand, includes a large number of tasks and a huge number of subtasks-- if each of the recipes in the cook's collection may be considered a subtask-- any one of which could generate a performance test. In such a case, a test of reasonable length can be constructed only by sampling arbitrarily from the vast array of possible performances. Whether such a test can be said to cover the entire course is debatable. Only in the 05C, 19E, 19K, and 71L courses does the EOCT appear to cover the entire course. In the other courses, the EOCT represents a sample of the course material.

In all but the 16S course the final grade is a GO/NO GO on the EOCT, and the other tests affect the final grade only insofar as the trainee has to pass all other tests before being eligible to take the EOCT. In the 16S course, the PI (Progress Index) is used as a "final grade." A PI = $\frac{\text{Observed time}}{\text{expected time}}$ of 1.25 or less is required for graduation. It remains to be seen whether the PI is a statistically satisfactory measure. For the other courses, a final grade might be constructed from the total number of first-time GO on all tests, for example, or a weighted sum of the number of tries, but the utility of such a measure is yet to be determined.

Turnover of military personnel frequently results in an absence of institutional memory. Interviews with SME about the development of existing training and measurement procedures and discussions about apparent gaps in training measurement have provided only limited information. At Ft. Dix, SME cited that they are not the proponent school for training in MOS 63B, 64C, and 94B as the reason for a lack of familiarity with rationales for particular aspects of training and measurement. At Ft. Gordon, SME presently assigned to the Directorate of Training Development have not been present at the time of development of training for MOS 05C.

The suitability for statistical purposes of any of the test scores available in any of these courses must be determined on a case-by-case basis. Complying with TRADOC guidance, many of the E2 and E3 tests have been developed in two or more alternate versions. For most of the written tests, these alternate versions are scramblings of the items and the distractors and qualify as equivalent forms. The alternative versions of the EOCT for both the 63B courses and both the 94B courses are not the same; some elements are common and some are different. Although the 63B course at Ft. Dix and Ft. Jackson are quite different in structure, they both use the test materials provided by the proponents of the course (Ordnance Center, Aberdeen Proving Ground); some comparisons between courses are therefore possible. The situation is quite different for the two 94B courses; the course taught at Ft. Dix uses the proponency's test materials; the course at Ft. Jackson, with the approval of the proponency, uses its own test materials. And so no direct comparisons are possible.

Frequently major differences in the way the tests are administered will vitiate comparisons between posts. In the 94B course at Ft. Dix, trainees practice small quantity cooking before they are tested; because of time constraints, trainees in the same course at Ft. Jackson are tested on the same materials after watching a live demonstration (and watching TV tapes), although most trainees have had no experience at cooking.

Similarly NO GO on E1 and E2 tests and failing grades on E3 tests are treated very differently from course to course and post to post. Usually, only two retests are permitted in most courses, but on occasion more may be given if, for example, the failure is slight or based on a misunderstanding, if the trainee has had a superior record up to that time, if the company or battalion commander recommends additional testing (or, if the trainee is in the National Guard or Army Reserve).

In most courses, the trainee can be retested immediately or after only a short delay unless retraining is clearly indicated. In the self-paced 05C course, the delay is up to the student. In some courses the trainee who fails a test is retested with another test; in others, with a different version of the same test; and in others, on only the items missed on the first test. Since the trainers and testers are making every effort to see that each trainee does get through the course (academic failures are quite rare), idiosyncratic treatment is bound to emerge in the testing procedure and this, in turn, limits the comparability of test results across trainees. As a consequence of the instructors' efforts, very few trainees fail to finish their courses. Attrition is low in all these courses, and very little of it is for academic reasons. Trainees are no longer recycled back to the beginning of the course, but there are occasional new starts made by trainees who have lost time through illness and who have been dropped back to one of the following classes.

In the self-paced mode of training, a trainee was sometimes allowed to "challenge" a phase of the training by taking the end-of-phase test and skipping the training for it if he/she passed. While this is still possible in the 05C course, it is rare.

All tests are monitored to prevent cheating. No major prompting is allowed on hands-on tests with the exception of a mild statement like: "Haven't you forgotten something?". This kind of prompt is not made part of the record. In many instances the trainee has a manual available during testing, and so prompting becomes supererogatory.

Development of Job Knowledge Tests

Job knowledge tests to be used as criterion measures of training performance in the 19 MOS are scheduled for development during the period October 83 - December 85. Development of these tests was begun during the visits to Army schools in which SME were interviewed,[1] where the tasks to be represented in the new measures were identified as follows:

(1) Lists of approximately 300-600 tasks for each MOS were obtained from the Army Occupational Survey Program (AOSP). The lists provide the percentage of soldiers performing each task, by skill level.

(2) Tasks performed by 5 percent or fewer of the soldiers in Skill Level 1 were excluded from further consideration.

(3) Where subtasks or elements of tasks were listed separately in the AOSP, they were combined to generate whole tasks with a natural beginning and end, e.g., the elements "remove old tire" and "install new tire" were replaced by "change tire."

(4) Two hundred twenty-five tasks were selected by stratified random sampling. Duty categories, for example, Redeye missile employment, Redeye operator maintenance, Redeye supply handling, were represented in proportion to the number of tasks in each category in the total list. Task titles were put on cards, one per card, for sorting by SME.

(5) From three to six SME, depending on the number available at each Army installation, eliminated those tasks that were obsolete or unfamiliar to them. They sorted the remaining tasks on a 3-point scale of importance and a 5-point scale of frequency of performance errors.

(6) The 100 tasks with the highest combined importance/error ratings were selected for analysis in group discussions by SME and research personnel to generate statements of correct procedure and to identify the locus and characteristics of errors in performance. These descriptions of correct procedures and errors will be used to the ex-

[1]At preparation of this report, knowledge test development had begun for the following MOS: 05C, 16S, 19E, 19K, 63B, 64C, 71L, 76Y, and 94B.

tent possible to construct item stems, correct alternatives, and construct distractors for multiple-choice knowledge test items.

Results

The ratings of tasks by SME for importance and error were analyzed to provide estimates of the consistency among raters in making the judgments. An appropriate form of intraclass correlation (ICC) where raters have not been selected randomly (Shrout & Fleiss, 1979) is:

$$ICC = \frac{\text{Mean Square Between - Residual Mean Square}}{\text{Mean Square Between - (K-1) Residual Mean Square}}$$

Reliability coefficients obtained using this formula are presented in Table 16 (Column A). Also included are the median Pearson r's for all paired comparisons of the raters (Column B) and an intraclass correlation based on a within-group design (Column C). The Column C coefficient was calculated because it is not affected by either small mean differences between tasks or lack of homogeneity of within-group variance. James, Wolf, and Demaree (1981) have suggested that the intraclass correlation may underestimate interrater reliability in situations where there is little difference in mean ratings between targets (tasks) even though there is almost perfect agreement among ratings for each target. Interrater reliabilities for a single rater ranged from low to moderate. However, the reliability of most interest is the reliability of the average rating across SME, because that is the information we have in hand. Using an average of four raters per MOS and applying the Spearman-Brown formula, the estimated reliabilities of the average ratings are shown in Column D.

Table 16

Task Rating Reliability Estimates

Post	MOS	Rating Type	A ICC	B Median r	C $\bar{r}_{wg}$	D[a] Rel. of Av.
Ft. Dix	94B	Importance	.34	.42	.75	.74
		Error	.08	.08	.27	.26
Ft. Dix	63B	Importance	.16	.10	.56	.31
		Error	.10	.10	.33	.31
Ft. Dix	64C	Importance	.06	.09	.22	.29
		Error	.00	.02	.36	.09
Ft. Knox	19E	Importance	.07	.11	.68	.33
		Error	.12	.16	.50	.43
Ft. Bliss	16S	Importance	.14	.32	.57	.65
		Error	.15	.18	.55	.47
Ft. Gordon	05C	Importance	.24	.25	.55	.57
		Error	.18	.21	.66	.52

[a]Column D shows the reliability of the average rating over four raters if the correlation in Column B is taken as the average.

The reason for the lack of high interrater reliability in some of the ratings of task importance and frequency of error in performance is not clear. Rating distributions were frequently quite dissimilar across raters, for example, many high ratings from one person and many low ratings from another in the same MOS, suggesting that the raters were using a different frame of reference, perhaps because of different prior experiences. Also because job assignments vary after an individual leaves AIT, and because many SME lack experience, precise judgments about importance and error rates may not always be possible.

Summary and Conclusions

The training school site visits have produced a large fund of information to be used in the development of training achievement criterion measures. For those schools visited, we now have in hand detailed information on the current criterion measures, the way in which they are used, the procedures used or not used to store training school information, the objectives of the school, and the content and design of the curriculum. The tests currently being used are being systematically examined to determine how thoroughly they reflect training objectives and content and how useful they will be as item content for the comphrehensive knowledge tests that must be developed as part of this project.

The existing training measures are one major source of item content for the comprehensive knowledge tests. Another major source is the description of relevant job tasks that has been developed. We now have in hand a list of

100-150 job tasks per MOS that were sampled proportionately from the categories of tasks contained in the Army's occupational survey item bank and which have been refined in terms of their importance and relevance to the MOS in question. Because the task descriptions were taken from the occupational survey item bank and the current Soldier's Manual, they are well anchored in the Army's design of the training curricula and the design of the job as it should be performed when the individual is required to perform in his or her specialty. While the importance and error ratings of the SME cannot be used to make precise discriminations among items, they are useful for identifying those tasks which are not currently job relevant and those which may be particularly prone to error.

As a consequence of the past year's effort, we are now in a reasonable position to begin generating the item pool for the comprehensive knowledge tests.

Next Steps

During the next 6-to-12 months the following activities will be paramount:

(1) The training school site visits will be completed, as per the original schedule in the Master Plan.

(2) The training objectives and training content will be (matched by the research staff) with the existing end-of-course tests and with the task descriptions to determine where new items must be written.

(3) The comprehensive knowledge test item pool for each item will then be generated.

(4) The items in each item pool will be submitted to SME and research staff review for a first determination of clarity, difficulty level, and relevance to specific training objectives.

(5) The items will also be pretested with small samples of incumbents.

(6) The edited item pool will then be administered to the criterion samples of trained vs. untrained enlisted personnel.

Associated Reports

The following report is by Rebecca L. Oxford-Carpenter and Linda J. Schultz of ARI entitled, "Toward Improving the Reliability and Validity of Army Training Measures." This paper provides both conceptual and practical suggestions about the kinds of research that could be done to improve the psychometric quality of training measures on Army installations.

Working Paper

RS-WP-83-7

PRELIMINARY THOUGHTS ABOUT RESEARCH ON RELIABILITY
AND VALIDITY OF ARMY TRAINING MEASURES

REBECCA L. OXFORD-CARPENTER

LINDA J. SCHULTZ

SELECTION AND CLASSIFICATION TECHNICAL AREA

APPROVED BY:
NEWELL K. EATON
Chief, Selection and
Classification
Technical Area

NOVEMBER, 1983

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

*To be published as Working Paper Selection and Classification Technical Area 83-7.

PRELIMINARY THOUGHTS ABOUT RESEARCH ON RELIABILITY AND VALIDITY OF ARMY TRAINING MEASURES

This working paper has been written to discuss, in a preliminary manner, some questions that must be addressed as part of the process of improving the psychometric quality, more specifically the reliability and the validity, of Army training measures. It precedes the development of a research plan which will define in a more precise manner the specific questions that will be addressed in a collaborative research effort conducted by the U.S. Army Research Institute and the Human Resources Research Organization. It is likely that the research plan will be much narrower in scope than this this general discussion. One possible modification may be to eliminate validity as an issue which can be adequately addressed in this particular effort. What this paper does attempt to do is outline some important issues that must be faced along the path toward improved Army training measures.

Some key questions are: What is the Army problem concerning training measures? What do reliability and validity mean? How does research on reliability and validity of training measures relate to Project A and to other projects? What are the objectives of research on reliability and validity of training measures? What research questions relate to reliability and validity of training measures? What does the literature say about reliability and validity indices that might be relevant to Army training measures? What further literature could be reviewed? What training measures could be researched? How could training data be gathered, by whom, and from where? What types of data analysis could be done? What kinds of scientific and operational outcomes might emerge? What problems might be foreseen?

What is the Army Problem?

The Army problem has several aspects. First, many Army trainers and test developers know little or nothing about psychometric quality assessment in general. Second, the state of the art is such that new developments are taking place constantly about quality of criterion-referenced tests, the Army's most prevalent type of training measure; yet the Army--largely due to lack of knowledgeable testing personnel and lack of resources--does not take advantage of these new developments. Third, little psychometric quality information, and even less information on differential reliability and validity (bias) by group, is available on Army training measures. Fourth, based on evidence from the field, many of these measures are psychometrically inadequate. Fifth, because of their inadequacy, such measures can detrimentally affect the Army as a whole and the careers of individual soldiers who may be misclassified or poorly selected for various Army jobs. Sixth, the overall picture of measurement on the training base is poorly documented, and as a result inaccurate information is often disseminated--particularly about the availability of quantitative vs. qualitative data. Seventh, the Army needs improved, multipurpose training measures. For example, many training measures that are criterion-referenced are limited in purpose to assessing whether the individual meets a given standard and may not be useful for ranking purposes or for predicting future job performance.

What Do Reliability and Validity Mean?

It is imperative that any type of assessment instrument used to evaluate an individual's performance, and which may impact on career decisions, demonstrate technical, psychometric qualities of high caliber. The psychometric qualities that have traditionally been recognized as crucial are test reliability and test validity. These qualities are so important because when used properly, valid and reliable tests provide evidence of achievement independent of subjective judgment (Schultz & Fortune, 1981).

To be valid, a test must first be reliable. It should be noted that the relationship between reliability and validity is such that test validity cannot exceed the square root of the test reliability; therefore, extremely high validity coefficients are not generally found in the literature, while very high reliability coefficients often are. Because a test must first be reliable, reliability is discussed first.

Reliability may be referred to as the consistency of measurement. Norm-referenced reliability types generally include internal consistency (split-half and other types), test-retest, and alternate-forms. Reliability is usually easier to document than validity.

Just as there are several kinds of test reliability, there are also several types of test validity. Validity concerns the extent to which the test measures what it purports to measure. Several types of validity may be relevant, such as content, concurrent, predictive, and construct validity. Any useful test generally has demonstrated validity of one of these kinds.

Content validity pertains to how well the test samples the domain of subject matter representative of skills from which conclusions are to be drawn. Concurrent and predictive validity represent two forms of criterion-related validity which look at the relationship between test scores and some independent external measures, i.e., criteria. Concurrent and predictive validity vary according to the time at which the criterion data are gathered, with concurrent criterion data gathered at the same time as the test in question is given and predictive criterion data gathered later. Construct validity concerns the use of tests for scientific inquiry and for the assessment of human traits, i.e., constructs. The construct validation process is used to clarify constructs and to test the basic theory whose traits are embedded in the test under investigation.

Sources regarding test reliability and test validity include Stanley (1971), Lemke and Wiersma (1976), Cook and Campbell (1979), Cronbach (1971), Jensen (1981), and Mehrens and Lehmann (1973). This section has defined reliability and validity in general terms, with a focus on traditional, norm-referenced varieties. As will be seen later in this report, new indices of a criterion-referenced nature have recently been developed. Some of the new indices are analogous to the traditional ones, but some are quite different. A complete discussion of new indices is given in the section on information in the literature about relevant reliability and validity indices.

How Does Research on Reliability and Validity of Training Measures Relate to Project A and Other Projects?

Research on reliability and validity of training measures closely relates to Project A, which is known as "Improving the Selection, Classification, and Utilization of Enlisted Personnel." Purposes of Project A are to (1) validate the Armed Services Vocational Aptitude Battery (ASVAB) against performance and (2) develop and validate new and other existing selection and classification procedures and measures in order to optimize the person-job match. Task 3 of Project A concerns measurement of school or training performance, in other words, performance in Advanced Individual Training (AIT). Although Task 3 does evaluate existing measures and its own new measures as part of its mission, Task 3 appears to focus more on validity than on reliability. Additional research focusing directly on reliability and validity of training measures could supplement Project A's effort on training measures.

Research on reliability and validity of training measures also ties in with the ARI in-house project entitled "Discriminant Utility of Training Criterion Measures." The discriminant utility project determines the discriminability of training measures in Army military occupational specialties (MOS) for the FY81/82 cohort, but it does not take the next logical step, which is to assess reliability and validity of these measures. Reliability would be a problem, because the data base does not contain the raw item data on individuals necessary to assess split-half (internal consistency) reliability. Also, there are no retests or alternate forms to provide other types of standard reliability indices. However, some rudimentary kinds of validity assessment might be possible by correlating the training data with the ASVAB or with the SQT.

The Computer Adaptive Testing (CAT) project is important because it is a joint services project that will validate the new CAT/ASVAB against existing training measures. The CAT project will gather training data on a sample of MOS. It is possible to share training-related reliability and validity information with CAT researchers.

What Are the Objectives of Reliability and Validity Research?

The suggested research objectives relate to recent, current, and new training measures in selected MOS. The objectives are:

- o To determine and document purpose, type, mode, and procedures of the measures. This is important because we need to have such basic data in order to know our subject.

- o To assess the reliability and validity of the measures in a variety of indices. We need to do this because to date there is very little information on reliability and validity and because no one has yet examined different indices of reliability and validity with Army training measures.

- o To determine the appropriateness of various indices using a set of criteria (statistical adequacy, practicality for Army use, and relevance to test and item type). It is crucial to decide which types of indices are appropriate for Army training measures.

- o To investigate the existence of differential reliability and validity by conducting analyses on training measure data from various racial/ethnic and sex categories. This is a key objective because test properties may be different across such categories.

What Research Questions Are Relevant?

Possible research questions include:

- o What are the purposes of recent, current, and new Army training measures in selected MOS?

- o Are some of the measures multipurpose?

- o If so, how are they used?

- What types of measures (norm-referenced, criterion-referenced, objective-referenced, other) are used?
- What modes (paper-and-pencil, hands-on, other) are used?
- What kinds of items (multiple-choice, completion, true-false, performance rating, etc.) are used?
- What kinds of cut-off procedures and scores are used?
- Is there any discrimination below and above the cut-off?
- If GO/NO-GO scoring is used, are quantitative data also available?
- What procedures are used for test construction, administration, scoring and reporting?
- Has any psychometric quality assessment already been conducted?
- If so, what kind?
- What reliability and validity data are available?
- What item analysis and variability (discriminability) data are available?
- How do various reliability and validity indices compare in terms of statistical adequacy?
- How do various reliability and validity indices compare in terms of practicality for Army use?
- How do various reliability and validity indices compare in terms of relevance to test type? (In other words, to what extent are classical reliability and validity indices relevant to the mostly criterion-referenced Army training measures, and to what extent are alternative indices relevant?)
- To what extent is there differential reliability and validity (bias) by racial/ethnic and sex category?

Clearly, the questions listed above have a direct relationship with the objectives mentioned earlier. Each question is directly linked with one or more of the objectives.

We now turn to a brief review of the literature on reliability and validity indices that might be relevant to Army training measures.

What Does the Literature Say About Reliability and Validity Indices That Might Be Relevant to Army Training?

The major function that a test serves will dictate to a great extent whether norm-referenced (NR) or criterion-referenced (CR) reliability and validity indices are appropriate. Norm-referenced test (NRT) performance is designed to rank individual examinees for purposes of screening, placement, selection, classification, evaluation of individual progress across a broad subject area, and overall program evaluation. In contrast, criterion-referenced tests (CRTs) are designed to compare an individual examinee's performance to an absolute performance standard, sometimes known as a "criterion," rather than to other examinees' performance. CRTs are often used to assign individuals to groups which reflect various levels of mastery concerning the objectives tested, to diagnose specific weaknesses and strengths, and to evaluate effects of instruction through pretests and posttests. It can be noted that for the purposes of screening, individual progress evaluation, and perhaps program evaluation there may be some overlap in function of CRTs and NRTs.

Certain indices of measurement for reliability and validity are associated with each of these measurement approaches. For example, the norm-referenced measurement (NRM) approach is traditional and relates to classical test theory. Earlier in this report, reliability was defined in general terms as consistency of measurement. Specific types of NR reliability coefficients include alternate forms, test-retest, and split-half. This means that the consistency of measurement relates respectively to the correlation between parallel test forms, the correlation between the scores on two or more administrations of a single test, and the correlation between halves of a single test. Pearson product moment correlation coefficients are usually used for alternate forms and test-retest reliability, while Spearman-Brown, Kuder-Richardson, and Cronbach's coefficient alpha are used for split-half and other internal consistency reliabilities.

Validity indices couched in the classical, NR approach concern the extent to which a test measures what it purports to measure. Several types of validity may be applicable to Army training measures. These types of NR validity include content, construct, concurrent, and predictive validity. These different validities were defined in an earlier section. Now we will discuss the appropriate statistical indices for each validity type. Because content and constuct validity require elements of personal judgment, these types of validity are often presented in verbal terms with no specific numerical expression. However, in the case of concurrent and predictive validity, which are both related to a criterion measure, statistical descriptions of the empirical relationships between predictor and criterion scores can be calculated. Concurrent and predictive validity differ in terms of the time of occurrence of the criterion scores. Pearson product-moment correlation coefficients are frequently computed as concurrent and predictive validity coefficients. However, point biserial, biserial, phi, and tetrachoric coefficients are also used to describe the degree of concurrent or predictive validity.

Several noted psychometricians, such as Hambleton and Novick (1973), Millman and Popham (1974), and Hambleton, Swaminathan, Algina, and Coulson (1975), have argued that classical (NR) test theory is not truly applicable to criterion-referenced testing. However, other well renowned measurement specialists, such as Woodson (1974), Livingston (1972, 1973, 1980), and Haladyna (1974) champion the use of NR measurement concepts and indices with CRTs. In addition, a variety of indices and procedures for calculating CR reliability and validity, along with several perspectives regarding the concept of CR variability, have been proposed by these and other measurement specialists.

With respect to calculating CR reliability, one of the most controversial techniques is that proposed by Livingston (1972, 1973, 1980). Livingston's (1972) procedure for calculating a CR reliability coefficient is based upon the squared deviations of scores from the performance standard (or cut-off score). Livingston's coefficient is thus purportedly analogous to NR reliability, which in some interpretations is based upon the squared deviations of scores from the mean. When Livingston's measurement assumptions are employed, i.e., the redefinition of variance, covariance, and correlation in terms of deviation of scores from the cut-off score , classical (NR) reliability theory appears as a special case of CR reliability theory when the mean equals the cut-off score (Stanley, 1971). Livingston's coefficient of CR reliability would appear to have intuitive appeal with Army training measures because of its emphasis on the performance standard, which is a key Army concern. However, Livingston's technique has been criticized by, among others, Shavelson, Block, and Ravitch (1972), who contend that Livingston's coefficient deviates so much from the traditional concept of reliability that it should more appropriately be labeled by some other name. Moreover, it is their contention that for the tests with which Livingston was concerned, traditional (NR) test statistics were applicable. Harris (1972) also criticizes Livingston's procedure, claiming that test floor and ceiling effects may be inconsistent with the procedure and charging that Livingston's work fails to advance CR reliability theory. Thus, it would appear that the applicability of Livingston's coefficient to the Army training measurement problem needs further investigation.

Another index for calculating CR reliability is suggested by Cohen (cited in Hambleton, Swaminathan, Algina, & Coulson, 1975). Cohen's coefficient Kappa is a measure of agreement which takes into account the element of chance when classifying examinees into mastery and nonmastery states based upon two test administrations. As such, it is an expression of consistency of measurement and reflects CR reliability. Coefficient Kappa is based upon the calculation of observed and expected proportions of agreement. Cohen has also proposed a weighted Kappa, which extends Kappa in order to permit the use of differential weighting of different kinds of misclassification. Potential utility to the Army might lie in being able to calculate a more accurate loss function than available with the use of other indices, thereby being able to capitalize upon more discriminating classification errors that could pose problems for selection into a training area.

Harris (cited in Hambleton, Swaminathan, Algina, & Coulson, 1975) has introduced an index of efficiency for a mastery test as a measure of CR reliability. The Harris index of efficiency proposes to determine the extent to which a CRT sorts students into mastery categories when the criterion for classification is based upon a cut-off score of an established number of items correct. This efficiency index is equivalent to a squared point biserial coefficient between total score and a dichotomous variable indicating the criterion group. Harris asserts that his coefficient serves the interest of decision-making accuracy with a large coefficient indicating high accuracy. However, Hambleton, Swaminathan, Algina, and Coulson (1975) claim that there may be instances in which the Harris index might lend itself to misinterpretation. The efficiency index does have implications for the Army with regard to its applicability to soldier placement as a result of correct assignment to mastery category. The CR reliability indices cited herein represent some of those offered in the measurement literature that might have merit for Army purposes.

In turning to CR validity, we find that of all the types of validity, content validity is recognized to have the most important bearing on CRTs. A number of procedures regarding content validation of CRTs and CRT items have been suggested in the literature. One such approach is that of item validation by content specialist or subject matter expert (SME) ratings. This method is widely used in the Army for training measures. This method, while intuitively appealing, is naturally judgmental and subjective. Assessment of the consistency of these ratings can be done by arranging specialists' ratings into a contingency table and then calculating a measure of agreement among item classifications. Such a measure of agreement might be Cohen's Kappa coefficient, as mentioned earlier in the reliabililty discussion. Coefficient Kappa mignt serve as a check to verify consistency among item raters. Other empirical techniques of item validitation include use of item difficulty and item discrimination statistics, according to Hambleton, Swaminathan, Algina, and Coulson (1975).

Cox and Vargas (cited in Oakland, 1972) have proposed a Difference Index (DI) which demonstrates some merit in detecting inferior quality items in CRTs. The DI actually compares results from two item analysis procedures. One procedure involves obtaining a posttest minus pretest DI by subtracting the percentage of students who passed an item on the pretest from the percentage who passed the same item on the posttest. The second procedure in the DI computation involves the more conventional technique which uses the difference between the upper and lower one-third determined by that item discrimination index which is based upon the distribution of student scores. Cox and Vargas then compute Spearman rank-order correlations between these two sets of DIs. Cox and Vargas contend that their DI index of item analysis produces results sufficiently different from traditional methods to warrant its consideration for use with CRM. An advantage of this technique for Army purposes is that it does not propose some psychometrically sophisticated computational procedure which yields difficult-to-interpret results.

Brennan and Stolurow (cited in Hambleton, Swaminathan, Algina, & Coulson, 1975) have established a process for both identifying CRT items which are in need of revisions and for determining item validity. As a component of this process, Brennan and Stolurow recommend the use of Cochran's Q test of the hypothesis of equal correlated proportions as a means of comparing difficulty levels of items intended to measure the same objective. Setting up confidence bands for each pair of significantly different items would provide information relative to which items were indeed significant. Hambleton, Swaminathan, Algina, and Coulson have noted this proposal as being constructive. The Army might consider its use for establishing content validity with CRTs in addition to the content specialist rating approach with SMEs.

Cronbach (cited in Hambleton, Swaminathan, Algina & Coulson, 1975) has advocated a rather elaborate method for establishing CRT content validity which entails the development of two separate test item pools from identical domain specifications by two independent test construction sources. The two resulting tests would then be administered to the same group of examinees and a correlation coefficient computed between the obtained sets of test scores. The obvious limitation of this method is the expense involved for its implementation. The Army may not be prepared to go to such lengths, especially when there are other desirable alternatives available.

Some other procedures which may be worth investigating are also offered in the literature. For example, Haladyna (1974) has endorsed a combined samples point biserial discrimination index as being extremely efficient with respect to determining the adequacy of CRT items. Popham (cited in Hambleton, Swaminathan, Algina, & Coulson, 1975) has proposed a two-pronged approach involving a priori and a posteriori elements for identifying CR items and establishing content validity. Finally, Woodson (n.d.) has advocated the use of a calibration sample to obtain empirical item characteristics and thus ultimately select items for the test--thereby contributing to the validity and to the determination of what the test measures.

This section has discussed traditional, classical, NR indices of reliability and validity and new, CR indices of reliability and validity. The debate continues about whether NR indices can be used with CRTs. Using Army training measures, it may be possible to try out certain NR and CR indices that seem promising. The state of the art in psychometric quality assessment can surely be advanced by such an effort, which is described in more detail in the following sections.

What Are Some Training Measures That Could Be Researched?

This section describes possible types of recent, current, and new training measures from several sources that could be researched. It also indicates the MOS with which these measures are associated.

Data from recent measures could be obtained from the discriminant utility project. The discriminant utility project covers training measures in approximately 227 MOS, data for which are contained in the Project A Longitudinal Research Data Base (LRDB) for the FY81/82 cohort. Of these MOS, it might be useful to select the ones that meet two criteria: they must overlap with the MOS selected for Project A research on training measures and they must have some useful variance. Of the 227 MOS, only four meet the two criteria just mentioned: 76W, 55B, 67N, and 05C (see Popelka, 1982). In collaboration with the researchers on the discriminant utility project, it is possible to obtain variability (discriminability) data on end-of-course training measures in these four MOS. It may also be possible to obtain some rudimentary validity data using the correlation between end-of-course training data and the SQT. Possible reliability indices will have to be determined. Both CRT and "traditional" indices could be investigated.

Recent and abundant test data are also available on 63B at Fort Dix for several years' time. It may be possible to conduct a mini-investigation of reliability and validity of end-of-course and interim training measures for 63B if the data prove workable.

Assessing the reliability and the validity of current training measures is problematic, because actual item data may not be available or may be impractical to use. If item data are usable and available, it would be possible to arrange with Task 3 staff to obtain raw data from two MOS in each of the three batches (X, Y, and Z batches) in order to assess the reliability of training measures. MOS now included in each batch are: 64C, 71L, 95B, and 13B (Batch X); 11B, 19E, 19K, 05C, 63B, and 91B (Batch Y); and 76W, 55B, 27E, 94B, 76Y, 16S, 67N, 12B, 54E, and 51B (Batch Z). MOS will be selected depending on the usability of training data (in terms of legibility, variability, and analyzability), on the density of the MOS, and possibly on findings from the discriminant utility project. Two potential candidates now appear to be 71L and 63B, both of which have usable training data and are dense MOS. A total of six, two per batch, could be selected. It might be useful to correlate current training scores for these MOS with scores on other measures, such as the appropriate ASVAB composite, in order to obtain some rough validity estimates. Interim and end-of-course training scores could be used where available.

For future, new job-knowledge training measures (i.e., measures developed specifically by Task 3 staff for Subtask 3.4 of Project A), it might be useful to obtain data from the same MOS selected for current measures--two MOS from each of three batches. These training measures will be developed and tried out over the period from now through the summer of 1985 and will be administered to the FY83/84 cohort in mid 1985. Again, reliability indices could be chosen from CRT or "traditional" types. Content validity could be assessed through comparison of test content with course content, SME reviews of tasks and items, and other related judgmental means. Predictive validity could be assessed via analysis of relationships between scores on training measures and actual job performance, data on which is to be gathered by Task 3 from the summer of 1985 through the spring of 1986.

How Could Data Be Gathered?

Data on recent measures has already been gathered. Much relevant data is on the LRDB and is ready for use. The wealth of recent 63B training data from Fort Dix has also been collected but would have to be organized and entered into a computer data base before it would be usable.

Task 3 staff members have arranged for submission of current training data by training personnel to Task 3. These data might be be used for some reliability and validity estimation.

Data on new measures will be gathered by Task 3 through a series of tryouts and field tests.

In short, it might not be necessary for more field data to be collected outside of preestablished channels in order to assess reliability and validity of recent, current, and new training measures. Exceptions might arise if item data are insufficient.

How Could Data Be Analyzed?

Data analysis, where possible, could be done using standard statistical packages such as SPSS, SAS, and BMD. Test and item analysis routines and regression and correlation programs could be used to obtain reliability and validity figures. However, standard packages may not suffice for certain criterion-referenced indices that could be used, so special programs might need to be written for those indices.

What Are the Scientific or Operational Outcomes?

Possible scientific or operational outcomes include:

- Determination of recent, current, or new testing types, modes, procedures, and psychometric quality of recent, current, or new training measures in selected MOS.
- Determination of most appropriate reliability and validity indices for different types of training measures.
- Identification of effects of group membership on reliability and validity.
- Recommendations for future psychometric quality assessment of Army training measures.
- Multiple reports aimed at various audiences (TRADOC, ARI, DoD, psychometric community, interservice groups, and professional associations).

What Problems Can Be Foreseen?

Several problems are predictable. First, the nature of many Army training measures is such that variance is at a minimum. Therefore, classical norm-referenced psychometric quality coefficients may be of little use with a large number of these tests. Researchers will have to look carefully for the appropriate psychometric quality statistics to use.

Second, coordination is an issue. New research on reliability and validity of training measures could fill any crucial psychometric gaps that Task 3 does not address. Coordination with Task 3, with other Project A tasks, and with other projects would need to be done.

Third, most new training measures developed by Task 3 and most new MOS-specific measures developed by Task 5 are to be job-knowledge tests. Precise distinctions between these two sets of measures have not yet been fully clarified. Therefore, it is too early to tell whether reliability and validity indices that are relevant to new Task 3 measures will be equally appropriate for new Task 5 measures. Generalizability of research results concerning reliability and validity of training measures to other similar kinds of measures--such as MOS-specific measures--would have to be approached with caution.

Summary

We have discussed the key elements in the investigation of reliability and validity of recent, current, and new Army training measures. The ultimate goal is to improve psychometric quality of training measures so that selection and classification procedures can be optimized.

REFERENCES

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation - Design and analysis issues for field settings. Chicago, IL: Rand McNally.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), Educational measurement, 2nd ed. (pp. 443-507). Washington, DC: American Council on Education.

Ghiselli, E. E. (1964). Theory of psychological measurement. New York: McGraw-Hill.

Haladyna, T. M. (1974). Effects of different samples on item and test characteristics of criterion-referenced tests. Journal of Educational Measurement, 11, 93-99.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 10, 159-170.

Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. (1975, April). Criterion-referenced testing and measurement: A review of technical issues and developments. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Harris, C. W. (1972). An interpretation of Livingston's reliability coefficient for criterion-referenced tests. Journal of Educational Measurement, 9, 27-29.

Jensen, A. R. (1981). Straight talk about mental tests. New York: Free Press.

Lemke, L, & Wiersma, W. (1976). Principles of psychological measurement. Chicago, IL: Rand McNally.

Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 9, 13-26.

Livingston, S. A. (1973). A note on the interpretation of the criterion-referenced reliability coefficient. Journal of Educational Measurement, 10, 311.

Livingston, S. A. (1980). Comments on criterion-referenced testing. Applied Psychological Measurement, 4, 575-581.

Mehrens, W. A., & Lehmann, I. J. (1973). Measurement and evaluation in education and psychology. New York: Holt, Rinehart & Winston.

Millman, J., & Popham, W. J., (1974). The issue of item and test variance for criterion-referenced tests: A clarification. Journal of Educational Measurement, 11, 137-138.

Oakland, L. (1972, April). An evaluation of available models for estimating the reliability and validity of criterion-referenced measures. Paper presented at the annual convention of the American Educational Research Association, Chicago, Il.

Schultz, L. J., & Fortune, J. C. (1981). The three i's: Sources of test bias. Education, 102, 117-123.

Shavelson, R. J., Block, J. H., & Ravitch, M. M. (1972). Criterion-referenced testing: Comments on reliability. Journal of Educational Measurement, 9, 133-137.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), Educational measurement, 2nd ed. (pp. 356-442). Washington, DC: American Council on Education.

Woodson, M. I. C. E. (n.d.). Classical test theory and criterion-referenced scales. Berkeley, CA: University of California. (DRIC Document Reproduction Service No. ED 083 298).

Woodson, M. I. C. E. (1974). The issue of item and test variance for criterion-referenced tests. Journal of Educational Measurement, 11, 63-64.

VIII. CONSTRUCTION OF MOS-SPECIFIC CRITERION MEASURES

Patrick Ford (HumRRO)

As noted in the Research Plan for Project A, job selection research, in general, and military research, in particular, have been frequently criticized for the lack of job relevant validation criteria. Consequently, there has been considerable pressure to include such criteria measures in a comprehensive validation effort. Inclusion of such measures is also dictated by our model of overall soldier effectiveness. That is, successful execution of the specific job tasks for which an individual was trained is a significant component of overall effectiveness. It is necessary that we make every effort to assess this component of effectiveness as well as the state of the art will allow.

At the same time, we must also recognize that standardized hands-on task performance measurement is expensive and the R&D costs for developing such measures are also high. Consequently trade offs must be made. As previously argued, we have opted to devote considerable research effort to a smaller subset of MOS (i.e., nine) rather than compromise the amount of resources devoted to criterion development in each MOS beyond the point where the crucial research questions could be answered. The general strategy also includes the development of behaviorally anchored rating scales and paper-and-pencil knowledge test measures to determine if the less expensive method can serve as a substitute for the more expensive.

Specific Objectives

The specific objective of this activity is to develop reliable, valid, and economical measures of first- and second-tour job task performance of enlisted personnel in a sample of nine MOS. These measures will serve both as:

(1) Data collection instruments for establishing the relationships among various kinds of predictors and criterion measures, and

(2) Prototypes for the development of performance measures for additional MOS and/or MOS clusters.

Two different kinds of performance measures will be developed. The first will be direct measures of task performance (e.g., the average time it takes a soldier to troubleshoot and repair a malfunctioning electrical component). For measures of this kind, the incumbents must be evaluated under carefully structured and standardized conditions. The second kind will consist of two measures that are based on indirect evidence of performance, knowledge tests, and ratings by supervisors or peers.

First Year Activities

During the first year our efforts relative to constructing MOS-specific criterion measures were focused principally on: (a) developing the specifc samples of job tasks from which to build hands-on performance measures and job-specific knowledge tests, (b) using the critical incident method to develop rating scale measures of MOS-specific task performance, and (c) using the accumulated task descriptions to develop a taxonomy of MOS-specific task performance categories, or factors, that will guide predictor selection and subsequent MOS clustering analyses.

Development of Task Samples for Performance Measurement

We began by generating a task sample for each of the four MOS in Batch A (13B, 64C, 71L, and 95B) by selecting and consolidating task statements from the Soldier's Manual (SM) and CODAP survey task descriptions.

These two job analysis sources were consolidated through a four-step procedure:

(1) Identify CODAP activities performed at SL1,

(2) Group CODAP statements under SM tasks,

(3) Group CODAP-only statements,

(4) Conduct Subject Matter Expert (SME) review.

Identify CODAP Activities Performed at Skill Level 1

The assumption for this step was that every activity included in the occupational survey questionnaire that had a nonzero response frequency, after allowing for error in the survey, was performed at skill level 1. The procedure for estimating the error was to compute the average response frequency for the survey and use that proportion to determine the boundaries of a confidence interval about zero. Activities with frequencies above the confidence interval were considered to have nonzero frequencies. For example, the confidence interval for 13B SL1 was $\pm$ 2.7. All statements with frequencies of 2.7 percent or lower were considered to be zero and were deleted from consideration; statements above 2.7 percent were considered part of the SL1 task domain. The results of this initial screen are shown in Table 17.

TABLE 17

CODAP Statements Deleted From Task Domain

	13B	64C	71L	95B
CODAP Statements	669	677	822	776
Delete by "Zero" Frequency	67	169	329	210
(Confidence Interval)	(2.7)	(3.0)	(4.0)	(4.2)
Delete SME Review:				
Change in Doctrine	19		58	177
Nontask				20
Collective task	24			
Balance	559	508	435	369

Group CODAP Statements Under SM Tasks

A CODAP statement (i.e., an item in the survey questionnaire) was placed under an SM (Soldier's Manual) task if the statement duplicated the SM task or was subsumed under the SM task as a step or variation in conditions. The effort first tried to identify SL1 tasks (either MOS specific or Common) with which the CODAP statement could be matched. If this could not be done, higher skill levels (HSL)--SL 2, 3, and 4--were successively reviewed and the CODAP statements matched with those SM tasks, if possible. Thus the grouping concentrated on matching CODAP statements with doctrine statements (i.e., Soldier's Manual tasks) wherever possible even if doctrine did not specifically identify the activity as a SL1 responsibility. All SL1 SM tasks were included regardless of whether or not they had parallel CODAP statements. A sample of such a grouping for 13B is included as Appendix E. The number of SM tasks with CODAP statements and the number of CODAP statements that matched the SM tasks are shown for each of the four MOS by skill level in Table 18.

Group CODAP-Only Statements

Since some CODAP statements could not be matched with any SM task, or any subset of elements from an SM task, the third step was to edit the remaining CODAP statements so that although they were similar in format to the SM task statements, they were still a clear portrayal of additional task content not contained in the SM. In some cases a CODAP statement became a task statement by itself. In other cases a new task statement was developed which could appropriately subsume several CODAP statements. The results of this step are shown in Table 18.

TABLE 18

Tasks in Performance Domain

	13B	64C	71L	95B
SM MOS TASKS				
SL1 w/CODAP (#)	67 (91)	21 (220)	39 (130)	98 (175)
SL1 w/out CODAP	55	1	0	40
SL2 w/CODAP (#)	24 (15)	*	*	15 (27)
SL3 w/CODAP (#)	28 (39)	2 (3)	50 (93)	0 (0)
SL4 w/CODAP (#)	19 (21)	3 (5)	1 (1)	2 (3)
COMMON TASKS				
SL1 w/CODAP (#)	22 (42)	61 (69)	61 (61)	46 (49)
SL1 w/out CODAP	48 **	10	10	23
SL2 w/CODAP (#)	13 (34)	20 (29)	10 (10)	5 (5)
SL3 w/CODAP (#)	5 (10)	5 (5)	2 (2)	3 (4)
SL4 w/CODAP (#)	2 (4)	10 (12)	1 (1)	2 (2)
CODAP ONLY (#)	73 (303)	33 (165)	29 (137)	70 (104)
TOTAL DOMAIN	356 (559)	166 (508)	203 (435)	304 (369)

* MOS combines SL1 and SL2. **One common task in MOS SM.

Note: Numbers in parentheses are CODAP statements subsumed by SM tasks.

Conduct SME Review

The final step in the consolidation was to confirm the grouping of CODAP statements with SME at the proponent school. At least three senior NCO or officers reviewed the grouping for each MOS. The review focussed on the placement of each CODAP statement and the appropriateness of the task titles for the CODAP-only tasks.

Some CODAP statements were deleted from the domain based on the SME review. As shown in Table 17, three reasons accounted for the deletions. The review of 13B identified changes in the doctrine (content specification) for the MOS that had occurred since the CODAP survey had been administered that would account for some of the CODAP-only tasks. Tasks that no longer applied (such as "Conduct ESC inspection") were deleted. The review of 95B identified administrative labels (such as "Question missing") that had been misconstrued as tasks. Also, in 13B, some ARTEP (collective) tasks were included. If the SME concluded that the collective tasks contained only individual tasks that were already in the domain, the statements were deleted. An example is "Fire high angle mission."

The result of the consolidation of SM and CODAP was a task domain for Skill Level 1 of each of the four MOS. The domain included:

(1) All SL1 tasks from the MOS SM and the SM of Common Tasks and their support CODAP statements.

(2) All HSL tasks with supporting CODAP statements

(3) All CODAP-only tasks.

These domains constitute a product in themselves in that they portray in precise task descriptive terms a definition of the job-world that an SL1 incumbent will face.

Narrow Domain

The task domains that were assembled were still too broad to assure that any task selected for hands-on test development would be both frequently performed and critical to job performance. The domains were, therefore, narrowed further through a six-step process. The goal was to arrive at a maximum number of tasks that could be managed feasibly in a systematic review by SME for criticality and clustering decisions. Because each of the four MOS presented unique structures, resources, and requirements, not all of the six steps that follow were performed for each MOS:

(1) Combine system specific tasks

(2) Delete tasks that pertain only to restricted duty positions

(3) Delete HSL tasks that have been officially designed not relevant to SL1 job performance by proponent

(4) Translate CODAP frequencies into task frequencies

(5) Delete HSL and CODAP-only tasks with atypically low frequencies

(6) Collect preliminary criticality ratings.

The tasks deleted as a result of each step are summarized in Table 19, as are the number of tasks in the final set selected for criticality evaluation.

Combine systems specific tasks. The Soldier's Manual for 13B treated the same operations performed on different equipment systems as separate tasks. For example, "Measure the quadrant with the range quadrant" applies

TABLE 19

Effects of Narrowing Domain

	13B	64C	71L	95B
TASKS IN DOMAIN (CODAP)	356 (559)	166 (508)	203 (435)	304 (369)
COMBINE SYSTEMS		NA	NA	NA
SL1	37			
SL2	10			
SL3	10			
SL4	5			
RESTRICTED DUTY POSITION		NA		NA
SL1 (CODAP)	46 (8)		19 (66)	
SL2	0		--	
SL3	0		14 (18)	
SL4	0		1 (1)	
CODAP-only	0		8 (18)	
DESIGNATED NA FOR SL1		NA	NA	NA
SL2 (CODAP)	0			
SL3	5 (6)			
SL4	13 (20)			
LOW FREQUENCY				
SM MOS Tasks			NA	NA
SL2 (CODAP)	1 (1)	--		
SL3	3 (4)	2 (3)		
SL4	1 (14)	3 (5)		
Common Tasks				
SL2 (CODAP)	9 (29)	13 (14)		
SL3	5 (10)	4 (4)		
SL4	1 (1)	9 (9)		
CODAP-only	34 (105)	16 (83)		
PRELIMINARY SORT	NA	NA	NA	
SL1				93 (148)
SL2				18 (30)
SL3				2 (3)
SL4				4 (5)
CODAP-only				59 (193)
TASKS DELETED (CODAP)	180 (185)	47 (118)	42 (103)	176 (279)
TASKS FOR CRITICALITY	176 (374)	119 (390)	161 (332)	128 (90)

howitzers and is treated as six tasks. From a training perspective that is appropriate because the performance steps vary somewhat among the howitzers. From the perspective of this project, however, treating such tasks as one rather than six tasks was preferable. The justification was that a soldier could only be held accountable for performing the task on the one kind of howitzer in his unit. If the task "Measure the quadrant with the range quadrant," should be selected, project staff may have to prepare as many as six forms of the test, but it should represent only one task in the criterion space.

Delete for restricted duty positions. The criterion for deleting a duty position task was that an Additional Skill Identifier or Special Skill Identifier and at least one week of special training were specified as being required for task performance. Only the 13B and 71L domains included duty positions that met that criterion. There were four duty positions for 13B: Artillery Mechanic (M198), Assembler; 155mm Atomic Projectile, Assembler; 8-Inch Atomic Projectile; and Nuclear Security Guard. The only 71L duty position that met the criterion was postal clerk.

Delete Higher Skill Level (HSL) tasks designated not applicable to Skill Level 1 (SL1). A set of MOS tasks for 13B had been reviewed by an Artillery Center Critical Task Board just before the SME review of the task domain. The results of that Board were distinctive in that the Board assigned levels of performance to each task by skill level rather than assuming a clear break between skill levels. Eighteen HSL tasks that had been in the task domain for SL1 had been rated in the lowest category (not applicable for Skill Level 1). Those tasks were deleted. Ratings for three tasks (after combining for weapons systems) which had not been in the domain (because no CODAP data covered them) had ratings that indicated that Skill Level 1 soldiers should at least have some knowledge of the task. Those tasks were added to the domain and are included in the domain totals in Table 18.

Translate CODAP frequencies into task frequencies. CODAP statements did not always correspond directly with task statements. In some cases, the CODAP statements represented steps within the tasks. In other cases, the CODAP statements represented various conditions. For example, CODAP frequencies covered statements like "Drive vehicle 2 1/2 tons or less in administrative convoy" and "Drive tractor-trailer combination vehicle in tactical convoy" when the 64C MOS task was "Operate Vehicle in Convoy." In still other cases, the CODAP statement was equipment specific while the task for testing purposes was generic. For example, CODAP frequencies covered "Prepare semifixed ammunition" and "Prepare separate loaded ammunition," but the task for consideration was "Prepare ammunition."

The algorithm for assigning frequencies to tasks is shown in Figure 8. Generally, when CODAP and task statements matched, the frequency for the matching statement was applied to the task. If there was no match, the most frequent step or condition was the basis for the task frequency. However, in some cases, frequencies were aggregated to account for equipment differences.

Delete low frequency HSL and CODAP-only tasks. The purpose of this screen was to identify tasks with atypically low frequencies. The general

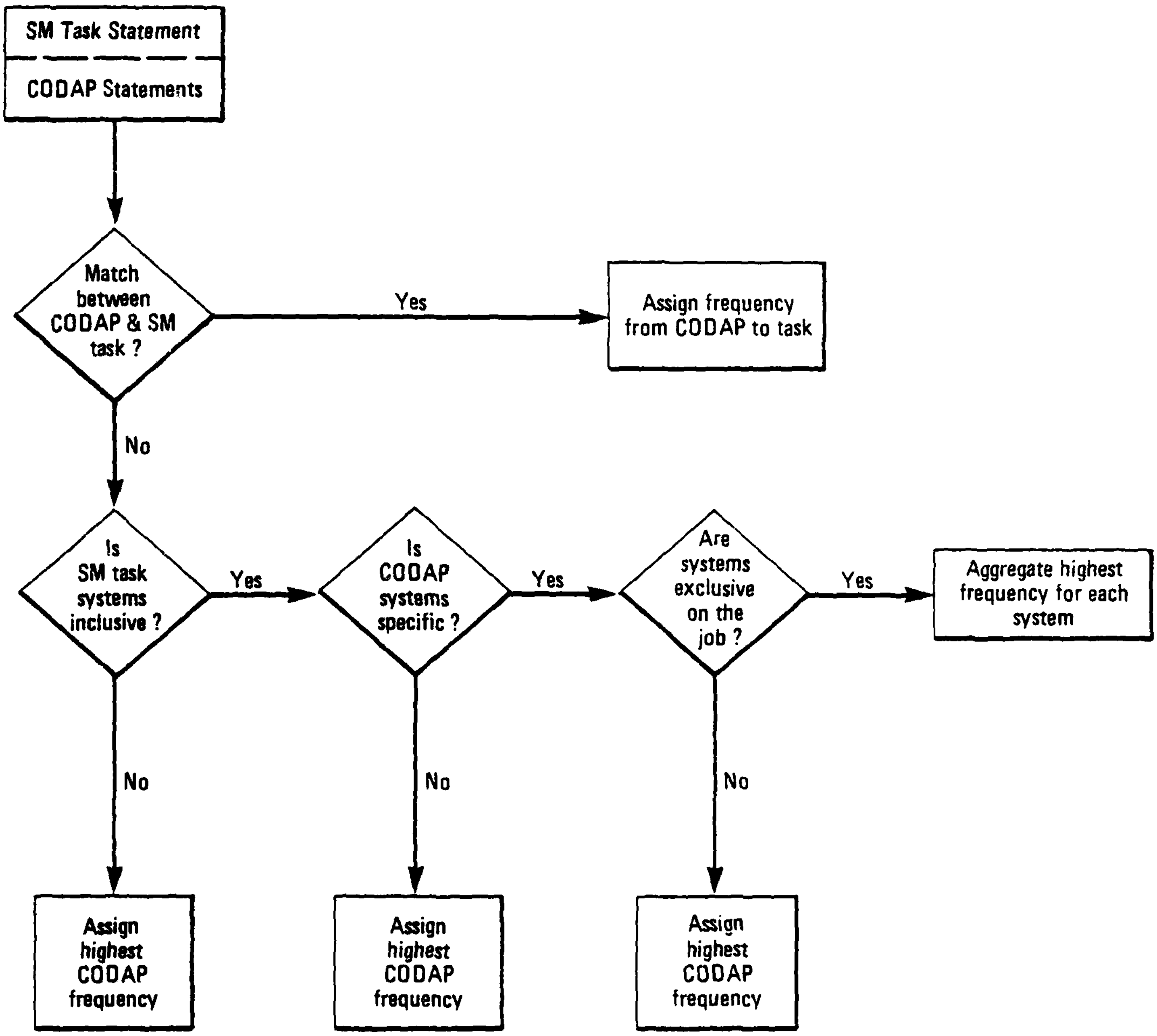


Figure 8. **Method for Assigning Frequencies to Tasks**

approach was to compare frequency distributions of the Skill Level 1 tasks (MOS and Common) with the HSL and CODAP-only tasks. HSL and CODAP-only tasks were then eliminated until the two groups were not significantly different with respect to location, dispersion, and form.

A four-step procedure identified the atypically infrequent tasks to be elminated:

(1) List the response frequencies of Skill Level 1 tasks.

(2) List the response frequencies of HSL/CODAP-only tasks.

(3) Test groups for difference using Mann-Whitney U test.

(4) If groups were different and the HSL/CODAP-only group had tasks with lower response frequencies, eliminate lowest frequency tasks until group differences were not significant at .01 level.

Collect Preliminary Criticality Ratings

Because the 95B Skill Level 1 domain was so large, it was narrowed through a preliminary sort on criticality. Ten senior 95B NCO were given 304 cards with task titles and brief descriptions of the scope of each task. They were asked to sort the tasks into two groups of approximately the same size. One group contained the more critical tasks, the other the less critical. They then ranked the group of more critical tasks from most to least important within that group. The ratings assigned to each task by the NCO were combined and plotted against respective CODAP frequencies to select the most critical, most frequently performed task for 95B SL1.

Scale Criticality and Cluster Remaining Tasks

The tasks remaining for each MOS after narrowing the domain were the candidates for selection. Since only a subset of the candidates could be covered in the large-scale data collection, further information was gathered to enable selecting a subset that contained the most critical tasks and represented the functional areas of the narrowed domain.

This information was gathered through a two-stage data collection effort with 15 senior NCO and officer SME at each of the four proponents. During the first phase the SME ranked each task. Each SME was given a card for each task. The card had the task title and a brief description of the scope of the task. The SME selected the one task that was most important for a European combat defensive situation and the one task that was least important for that situation. The SME repeated the process until all tasks were ranked.

During the second phase the SME sorted the tasks into groups based on the performance requirements of the tasks. The SME worked with the same cards as in the criticality phase. The results for the 15 raters were analyzed by means of a hierarchical clustering program.

For the final task selection project staff selected tasks to represent the clusters, giving priority to high criticality/high frequency tasks.

Summary and Next Steps

What we have produced at this point is a very carefully specified set of tasks that will form the content of the hands-on performance measures and the paper-pencil job knowledge measures for the Batch A MOS. Great care was taken with the above procedural steps to insure that the content of the two MOS-specific performance measures accurately reflects relevant job content that is highly representative of what people actually do and that is critical for effective performance in the MOS.

In the coming year we will repeat the process for Batch B MOS and will begin to develop the actual exercises that will test for proficiency on the specified tasks.

MOS-Specific Behaviorally Anchored Performance Rating Scales

As noted earlier two alternative methods will be used to assess MOS-specific job performance. One method will use paper-and-pencil measures of job knowledge that reflect the tasks being assessed by the hands-on measures. Construction of these tests will begin during FY84.

The second method uses the critical incident technique to develop behaviorally anchored rating scales for task performance factors in specific MOS. Work began on these measures during the last quarter of FY83 and is continuing at the present time.

Procedure

To develop behaviorally anchored rating scales for the four MOS in Batch A (13B, 64C, 71L, 95B), critical incident workshops have been conducted with 8-15 NCO from each MOS in each of four locations. The workshops were conducted in the same way as previously described. The only substantive difference is that the workshop participants were asked to focus on task performance itself rather than on behavior examples that might represent other major aspects of total effectiveness.

Results

The number and location of the workshops, the number of participants, and the number of critical incidents that were generated are shown in Table 20. From these incidents an initial set of performance factors for each MOS have been constructed by having the project staff group critical incidents into categories that are judged to represent similar elements of task performance. These initial sets of factors are shown in Appendix F. The retranslation phase of the procedure and the completion of the rating scales will be done in FY84.

Next Steps

During the next contract period the remaining workshops will be conducted and the development of specific performance factors for each MOS in the sample will be completed. The remaining steps will include the [illegible] important retranslation step (see Research Plan), which [illegible]

Table 20

Behavioral Analysis Workshops Conducted

Location	Dates	MOS	Number of Participants	Behavioral Examples Generated
Ft. Ord	25-26 August	64C	10	80
		71L	7	59
		13B	15	195
		95B	14	213
Ft. Polk	29-30 August	64C	15	240
		71L	14	210
		13B	11	150
		95B	15	235
Ft. Bragg	12-13 September	64C	9	175
		71L	10	154
		13B	16	195
		95B	17	225
Ft. Campbell	15-16 September	64C	9	175
		71L	10	154
		13B	16	195
		95B	15	238
Totals		64C	48	716 (14.9)*
		71L	43	641 (14.9)
		13B	54	775 (14.4)
		95B	61	911 (14.9)
				3,043

*Average incidents per participant.

valid and reliable performance dimensions. The judgments obtained in the retranslation step will be used to construct rating scales for each of the MOS-specific performance dimensions. These new rating scales will then be pilot tested on small samples of incumbents. The result will be the first set of behaviorally defined rating scales that has ever been used to assure specific technical performance in a skilled job.

Associated Reports

As noted in the Introduction, a selected research report follows:

Issues and Strategies in Measuring Performance in Army Jobs *

William Osborn

Human Resources Research Organization

How well soldiers can perform their job duties is of central importance to the Army. As implied in the previous paper, job skill alone is not the only measure of soldier effectiveness. But in light of the congressional action which stimulated this project--a mandate to tie selection procedures to actual job performance--the direct and valid measurement of that performance is seen as the anchor point in criterion development.

We are all familiar with the customary ways of assessing job proficiency: hands-on performance tests, paper-and-pencil tests of job knowledge, and more indirect methods involving ratings of job performance. I will turn in a moment to a discussion of these, as well as other aspects of the research and development work planned in this part of the project. But before specific methods and procedures can be sensibly discussed, it is necessary to set the stage conceptually. This is important because the merit and feasibility of various approaches to job analysis and performance measurement influenced the development of the research strategy for the overall project. And, in turn, that strategy to a large extent constrained the possible directions that job-specific performance measurement could take.

The problem faced by project designers was this: With limited dollars and limited troop support how can one develop valid selection procedures for each of more than 200 occupational specialties, and do so in a way that will satisfy the mandated criterion standard of job performance. Clearly, cost

*Paper presented at the 91st Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

prohibits empirically validating predictors against a thoroughly measured criterion of proficiency and performance for each occupational specialty. Three broad alternatives remain.

One is to validate predictors directly for every job using criterion measures that are highly general and inexpensive--generic rating scales and indicators of performance such as disciplinary actions, commendations, or rate of advancement. The shortcomings of this approach are that (a) it fails to meet the requirement for a criterion of job performance, and (b) there is no solid basis for defending the substitutability of these general criteria for that of measured job performance.

A second approach is to attempt to validate predictors against common dimensions of job performance that have been abstracted for measurement and then reconstructed analytically to provide a separate criterion measure for each job. Generic performance tests developed and used in this way offer an interesting but unproven method of criterion measurement. Even if possible, such measures might well lack the necessary depth of coverage to capture important distinctions in performance among jobs. Moreover, because job performance is measured more abstractly, this approach, like the first, would lack the apparent relevance or face-validity common to job-specific performance tests.

The third approach--and the one adopted for this project--is to validate predictors against criteria intensively measured for each of a sample of jobs chosen to represent the occupational domain, and then attempt to generalize the predictor equations to other jobs through linking profiles of similar job requirements. While fewer jobs can be addressed initially with this approach, criterion performance can be thoroughly assessed maximizing the opportunity for differential validity to emerge.

The early success of the approach hinges on two things: one is selecting a small but highly representative sample of job specialties on which to focus, the other is choosing a valid and comprehensive mix of criterion measures.

Choosing the number and kind of job specialties to be sampled involved compromise. Again, with project resources fixed, the greater the depth of performance measurement the fewer the jobs covered, and, conversely, the more shallow the measurement the greater the number of jobs that could be addressed. Selecting a range of job types was constrained to some degree by the statistical need for an adequate number of incumbents in each specialty. These tradeoffs were made and a sample of 19 Army occupational specialties chosen. The sample represents about 8% of the Active Army's 238 entry-level jobs but over 40% of all soldiers. The jobs range from combat to noncombat and from technical to nontechnical; they reflect current variations in aptitude requirement as well as the Army's structure of occupational specialties used in career management. Most importantly, perhaps, the sample represents quite well clusters of Army job specialties derived from judged similarity of performance requirements.

We knew from the offset, however, that job proficiency could not be assessed fully or realistically for even this sample of 19 specialties. A measure of job knowledge and the objective indicators and ratings of overall soldier achievement were to be obtained for all 19, and an additional set of job-specific performance measures administered in a representative subsample of nine of the specialties. This strategy provides feasible criterion measures for the full sample while enabling an empirical evaluation of the relevance and completeness of those measures in terms of job-specific performance for a subset of the specialties.

This then is the context within which the measurement of job-specific performance will proceed.

Measurement Methods

The immediate objective of this phase of the project is to develop reliable, valid and economical measures of first and second tour job performance of soldiers in the sample of nine occupational specialties. These measures will serve two purposes. Their primary purpose is to provide a criterion for validating a variety of predictors and other criterion measures. Their secondary purpose is to serve as prototypes for future measures of Army job proficiency.

Two different kinds of performance measures are planned. The first are direct measures of task performance obtained under structured and standardized conditions: for instance, time to isolate and repair a fault in a piece of electronic equipment, accuracy in filling out a form, or time and accuracy in engaging a target. The second kind will consist of measures that are based on indirect evidence of performance such as job knowledge and ratings by supervisors or peers.

Both kinds of measures are needed. Instruments of the second, more economical type are needed for operational use in monitoring performance, and for the Army's continuing efforts to improve selection and classification--which will not end with this project. Instruments of the first type are needed in order to develop the second. They also are needed to calibrate periodically the accuracy of selected predictor instruments. The careful calculations of utility that will be made in this project would be open to serious challenge if they were based solely on less direct measures of performance.

Issues concerning the choice and mix of methods generally center around trade-offs between the cost and validity of alternative approaches. The more precisely one specifies the performance to be observed and the conditions under which it is to be observed, the higher the cost.

Frederickson (1962) and Engel (1970) offer simple taxonomies of performance evaluation measures. Both tend to distinguish measures along two continua of remoteness or indirectness relative to actual job performance: the remoteness of the test behavior observed and the remoteness of the observer or scorer. Job performance tests are generally viewed as the most direct method since they call for application of knowledge and demonstration of skill by eliciting behaviors that are equivalent, or nearly equivalent, to those required in the job setting. But the directness of this method--with its inherent relevance, content validity and fairness--comes at a price. Many personnel managers believe the benefits of performance testing do not justify the demands on facilities and personnel (Harris & Mackie, 1962) nor the wear and tear on equipment (Angell Shearer, & Berliner, 1964). Also, the level of professional skill available in the military to develop and administer performance tests has been questioned (Vineberg & Taylor, 1972). And yet another shortcoming of performance tests--obvious but not widely discussed--is that the greater administrative time they require usually restricts coverage of job domain; one can measure fewer job tasks per unit of time than is possible with less direct measures.

The shortcomings of performance tests, especially that of cost, have led to the widespread use of job knowledge tests. Job knowledge tests consist of questions about task performance, usually delivered in a paper-and-pencil multiple-choice format. They are indirect measures to the extent that the behaviors measured do not constitute task performance but only mediate it.

Despite their evident economy, a question lingers concerning the degree to which knowledge tests can adequately gauge a person's job performance capability--either in terms of the range of job behaviors that can be validly represented by knowledge items, or in the sense that knowledge testing in a paper-and-pencil mode presumes at least minimal literacy. In some earlier Army research (Shirkey, 1966; Urry, Shirkey and Waldkoetter, 1965; and Yellen, 1966), correlations between job knowledge test scores and work-sample criteria were found to be too low to support the use of knowledge tests alone to assess individual proficiency for the jobs of medical specialist, supply specialist, cook and track vehicle mechanic. Similar results were reported in a study of general vehicle repairmen (Engel and Rehder, 1970), and in a review of Air Force research on maintenance performance (Foley, 1974). On the other hand, knowledge tests do appear to have adequate validity for jobs (e.g., personnel specialist) in which cognitive skills predominate, provided that only knowledge actually required on the job is covered in the test (Urry, Shirkey & Nicewander, 1965; Vineberg & Taylor, 1972). Adequate validity also was observed in a more recent study (Osborn & Ford, 1977) in which the knowledge tests were evaluated against a hands-on mastery criterion for low-skill manual tasks. Controlling for mental ability and level of task mastery, correlations on the order of .70 were found between various kinds of knowledge tests and hands-on task performance. These high correlations, it is important to note, were attributed to two factors: (1) the skilled aspect of the tasks tested consisted essentially of recalling functions, not of manual performance, making a knowledge medium appropriate; and (2) the knowledge items were meticulously tied to the critical steps in task performance through careful task analysis.

Affective classes of behavior, such as motivation to perform a task, can be assessed by performance tests if one uses unobtrusive measures (Osborn, 1979). But to embed a task in some simulated job context sufficiently broad to permit the task to be performed voluntarily requires time and expense not typically justifiable. Standardization and scoring problems also militate against attempting to test motivational behaviors *in situ* (e.g., Harris, Campbell, Osborn, & Ford, 1975).

Similarly, time pressures, inadequate supplies and equipment and various organizational factors can influence the performance of soldiers who otherwise know how and want to do the job correctly. An indirect measure, usually in the form of ratings by a supervisor or peer, is therefore considered a more feasible method of tapping the affective or "will do" aspects of job behavior. Supervisor or commander ratings typically do not correlate highly with job knowledge or job sample test performance (e.g., Engel & Rehder, 1970; Vineberg & Taylor, 1972), but this does not rule out their use for measuring aspects of performance not represented in knowledge or hands-on tests. Such ratings can be particularly useful when developed in ways that anchor the rater's judgments to specific, relevant job behaviors (e.g., Borman, Dunnette & Johnson, 1974; Borman, Hough & Dunnette, 1976; Campbell, Dunnette, Arvey & Hellervik, 1973).

It seems that different methods of performance measurement have different advantages and disadvantages. Despite their cost, hands-on performance tests, correctly developed and administered, cannot be equalled in job relevance, fairness, or acceptability to examinees (Schmidt, Greenthal, Hunter, Berner, & Seaton, 1977); nor is there a known substitute for a performance test in measuring proficiency on tasks involving psychomotor skill. Knowledge tests, if used for the right kinds of job tasks and linked methodically to knowledge-based task elements, have wide applicability, acceptable validity and administrative feasibility. Performance ratings are the most feasible but the most indirect measures of job proficiency, yet they permit the measurement of affective dimensions that cannot be efficiently tapped by other means.

Our plan in the present project is to try all three methods, evaluating their construct validity in a field test and selecting on the basis of those results the optimal mix of measures to use for collecting job-specific criterion data on the main cohort.

Development of Measures

Time does not permit a detailed account of plans for developing the measures. I would, however, like to give you an overview of development procedures, highlighting some of the more important steps and issues.

Work has begun on the first developmental step, that of job and task analysis. Two kinds of analyses are underway: one suitable to developing the performance and knowledge tests, the other to developing the behavioral based ratings.

Since tests are to be tied to job tasks, job performance must first be broken down into a list of tasks, a sample of those tasks selected, then each further analyzed into its elemental parts. A task inventory approach to job analysis has its shortcomings, in that important contextual factors and transition behaviors can be lost in the "seams" when a job is partitioned into tasks (Osborn, Harris & Ford, 1974). Yet the approach has advantages for testing. Chief among them is the exceptional degree of control and standardization that can be achieved in testing performance task by task. Another is analytic convenience: judgments about mission relevance, frequency of performance, kinds of behavior, and other data important to achieving content validity can be obtained systematically and reliably on a task by task basis. A final advantage particular to this project is that the Army's occupational analysis system is based on a task inventory approach.

The Army Occupational Survey Center, working with the technical schools responsible for job and task analysis, routinely develop task lists for the various job specialties. A survey form based on the task list is used to obtain data from job incumbents on, among other things, tasks performed and frequency of performance. These task lists and related survey data were used as a starting point in our work. Drawing on doctrinal expertise in the service schools, the task lists are being refined in light of such factors as job structure and recent or impending changes in equipment. Once differences between doctrine and practice have been resolved, each list will be narrowed to the duty positions and skill level of relevance. These tasks will then be further screened using frequency-of-performance data from the field and mission-criticality estimates from the schools--tasks low in both frequency and criticality being dropped. We expect the resulting list to be in the neighborhood of 100 tasks for each job specialty. These will then be subjected to a final screening involving new task data. Using rating protocols under development, subject-matter-experts will be asked to do two things. The first is to sort tasks into groups, <u>not</u> on the basis of job function or related equipment, but in terms of similar performance requirements; the idea being to derive from these data a crude approximation of factored intercorrelations of actual task proficiency. The second requirement of the job experts will be to order the tasks as to importance in carrying out a prescribed unit mission. Those important and frequently performed will then be sampled over the clusters of similar tasks to identify the 25 to 35 that should and can be feasibly tested within the limits of project resources.

In the final stage of analysis for test development, detailed descriptions will be prepared for each selected task. Existing descriptive data from Field Manuals, Technical Manuals, Soldier's Manuals, school lesson plans, and other task analysis research will be used in detailing the procedures, conditions and standards for performing a task. Accuracy and completeness of the descriptions will be confirmed through job expert reviews.

Quite a different job analysis will be conducted to support development of the behavior-based performance rating scales. In a series of workshops with experienced job incumbents--approximately 100 in each job specialty--examples of effective and ineffective job performance will be obtained. These will be in the form of brief stories or vignettes drawn from the participants experience and stated in terms of soldier performance on the job. When edited and cast into a common format the performance examples are ready for scale development

The two types of job analysis are proceeding concurrently. Once they are completed, development of the respective performance measures will begin.

The first step in test development will be to decide in the case of each task whether a hands-on or a knowledge test will be prepared. For purposes of experimental field testing, to be described in a moment, both kinds of tests will be developed for several tasks. We expect to develop hands-on tests for about half of the tasks which is all that administration time will support; included in this set will be those tasks with a skilled psychomotor or perceptual component. Knowledge tests will be developed for nearly all tasks in the sample, excluding only those in which psychomotor or perceptual skills predominate.

Development of a hands-on test begins with the task descriptive data and proceeds through four stages. The scoring approach is determined first. Here a decision is made as to whether test scoring will be directed at the product or the process of task performance, or both. The question that must be answered is what must the scorer observe to evaluate a soldier's performance. If task performance results in a product then scoring can focus on the measurable characteristics of that product. However, since the relationship of task process to task outcome is not always evident, some elements of performance--safety precautions, for instance--must be observed and scored as they occur.

In the second stage of development these product and process characteristics are translated into a score sheet or list of observable items to be evaluated by the scorer. These items are molecular enough to be scored reliably by a trained evaluator with or without scoring aids. Dimensions of overall task performance such as time and sequence are added to the scoring protocol where important.

The third stage in developing a hands-on test involves preparing detailed specifications for test administration, including the precise conditions (environment, equipment, terrain, etc.) for testing, instructions to examinees, and instructions to scorers.

Once the draft test package has been reviewed by job experts, development moves to the fourth and final stage in which through a small scale field tryout the test is evaluated for administrative feasibility, scorer reliability and acceptability to examinees. Final revisions are guided by the results of this trial run.

Development of a knowledge test begins with the same task descriptive data. In fact, methodically linking test questions to task procedures is the key to valid tests of job knowledge. And the sequence of decisions and actions to be followed in that linkage hinge on the causes of failure to perform the task correctly. Each essential behavior within the task is analyzed with the help of a job expert for potential causes of a failure: Is it because the soldier doesn't know <u>when</u> to perform a step? Is it because the soldier doesn't know <u>where</u> to perform it? Is it because the soldier doesn't know <u>what</u> the end result looks like? Is it because the soldier doesn't know <u>how</u> to execute the behavior?

For each likely cause of error, the <u>correct</u> location, or sequence, or product, or procedure is identified and then described in words or pictures. A question is then framed, and, finally, real-world response alternatives (distractors) are selected to complete the test item. The important point is that by considering these four questions about each aspect of task performance, we can pinpoint both what is important to ask in a knowledge test of task performance, and how to ask it. This procedure normally produces from four to eight questions, depending on task complexity, that tap the critical aspects of task performance. It also helps prevent test questions that so often are used merely because they are easy to ask.

Development procedures for the behavior-based rating scales pick up where the behavioral analysis left off. The edited examples of effective and ineffective job performance are content analyzed to form preliminary performance dimensions. After being reviewed by job experts for plausibility and wording, the preliminary dimensions are returned along with the performance examples to the group of experienced job incumbents who provided the examples. They are asked to sort the examples according to content into the dimensions and then to rate each on a seven point scale of effectiveness. When analyzed, these data will yield the final set of dimensions or rating scales, with points along each scale illustrated or anchored by actual performance examples.

A final form of performance measure, not yet mentioned, will be added to round out experimentally those already described. This is a rating scale for each of the job tasks to be tested. A simple seven point scale of of quality of performance will be used to obtain from the supervisor and peers summary judgments of an incumbent's task proficiency.

When all job-specific performance measures are prepared, they will undergo evaluation in a field test. Approximately 150 soldiers from each job specialty will be tested. A soldier will take the hands-on and knowledge tests, rate his peers on summary task performance and on the behavioral scales, rate the acceptability of the methods used to measure performance, and complete a job experience questionnaire; in addition, supervisors will provide ratings of soldier performance on both the task-specific and behavioral scales. Data collection will take nearly two days per soldier and two weeks for a group of 150.

Reliability, administrative feasibility, cost, and acceptability to soldiers will be assessed for each of the measures. Though the sample size will be too small for sensitive measurement of subgroup bias, differences in performance among ethnic and gender subgroups over type of measure will be examined.

The most interesting analyses will be of relationships among the various types of criterion measures. The planned overlap among the task-specific measures, mentioned earlier, is illustrated in Figure 1. Since performance will be measured by two methods on all tasks and by all three methods on about a third of the tasks, convergent validity of the three types of measures can be analyzed. Moreover, combining these task performance data with those from the job-specific behavioral ratings and the Army-wide measures of soldier effectiveness (which will also be collected during the field test) will provide the first opportunity to examine empirically the dimensionality of the criterion space. The factor structure within and across criterion measures will provide a basis for evaluating the construct validity of measurement methods.

Results of the field testing will bring us to the final development step in job-specific performance measurement: That of choosing for each job specialty the set of measures to be administered in the main cohort. It is assumed at this point that hands-on measures and behavioral ratings will, for reasons mentioned, comprise the core of the criterion instruments. The extent to which other more economical measures may be used to augment or perhaps substitute for the core measures will be determined from field test data.

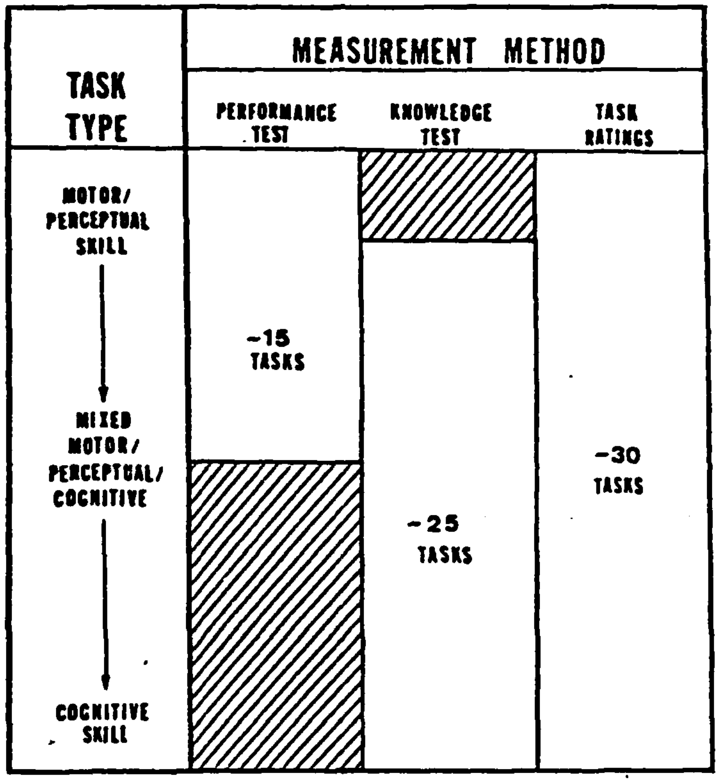


Figure 1. Planned Overlap of Task-Specific Measurement Methods

References

Angell, D., Shearer, J. W., & Berliner, D. C. (1964). Study of training performance evaluation techniques (TR-NAVTRADEVCEN-1449-1). Orlando, FL: Naval Training Device Center.

Borman, W. C., Dunnette, M. D., & Johnson, P. D. (1974). The development and evaluation of a behavior-based naval officer performance assessment package. Minneapolis, MN: Personnel Decisions, Inc.

Borman, W. C., Hough, L. M., & Dunnette, M. D. (1976). Development of behaviorally based rating scales for evaluating the performance of U. S. Navy recruiters (TR-76-31). San Diego, CA: Navy Personnel Research and Development Center.

Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. (1973). The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 57, 15-22.

Engel, J. D. (1970). An approach for standardizing human performance assessment. Paper presented at THEMIS Conference, Lubbock, TX:

Engel, J. D., & Rehder, R. J. (1970). A comparison of correlated job and work sample measures for general vehicle repairmen (HumRRO Technical Report 70-16). Alexandria, Virginia: Human Resources Research Office.

Foley, J. P., Jr. (1974). Evaluating maintenance performance: An analysis (AFHRL TR-64-57(I)). Wright-Patterson Air Force Base, OH: Air Force Human Resources Library.

Frederickson, U. (1962). Proficiency tests for training evaluation. In R. Glaser (Ed.), Training research in education. Pittsburgh, PA: University of Pittsburgh Press.

Harris, A., & Mackie, R. R. (1962). Factors influencing the use of practical performance tests in the Navy. ONR Technical Report 703-1.

Harris, J. H., Campbell, R. C., Osborn, W. C., & Boldovici, J. A. (1975). Development of a model job performance test for a combat occupational specialty: Volume I. Test development; Volume II. Instructions and procedures for conducting a functionally integrated performance test (HumRRO Final Report FR-CD(L)-75-6). Alexandria, VA: Human Resources Research Organization.

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1983). Improving the selection, classification, and utilization of Army enlisted personnel; Project A: Research Plan (ARI Research Report 1332). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Osborn, W. C., Harris, J. H., & Ford, J. P. (1974). Functionally integrated performance testing. Proceedings of the 16th Military Testing Association Conference.

Osborn, W. C., & Ford, J. P. (1977). Research on methods of synthetic performance testing (HumRRO, Final Report FR-CD(L)-76-1). Alexandria, VA: Human Resources Research Organization.

Osborn, W. C. (1979). Analysis and classification of performance testing problems. Proceedings of the 21st Military Testing Association Conference.

Schmidt, F. L., Greenthal, A. L., Hunter, J. E., Berner, J. G., & Seaton, F. W. (1977). Job sample versus paper and pencil trades and technical tests: Adverse impact and examinee attitudes. Personnel Psychology, 30 (2), 187-197.

Shirkey, E. C. (1965). Preliminary validity report of the MOS evaluation test for medical specialist (USAEEC Research Study No. 31). Ft. Benjamin Harrison, IN: U. S. Army Enlisted Evaluation Center.

Shirkey, E. C. (1966). Preliminary validity report of the MOS evaluation test for general supply specialist (USAEEC Research Study No. 38). Ft. Benjamin Harrison, IN: U. S. Army Enlisted Evaluation Center.

Urry, V. W., Shirkey, E. C., & Nicewander, W. O. (1965). A preliminary validity report of the MOS evaluation test for personnel specialist, MOS Code 716.1 (USAEEC Research Study No. 21). Ft. Benjamin Harrison, IN: U. S. Army Enlisted Evaluation Center.

Urry, V. W., Shirkey, E. C., & Waldkoetter, R. O. (1965). Preliminary validity reporting of the MOS evaluation test for cook (USAEEC Research Study No. 20). Ft. Benjamin Harrison, IN: U. S. Army Enlisted Evaluation Center.

Vineberg, R., & Taylor, E. (1972). Performance in four Army jobs by men at different aptitude (AFQT) levels: 4. Relationships between performance criteria (HumRRO Technical Report 72-73). Alexandria, VA: Human Resources Research Office.

Yellen, T. (1966). Preliminary validity report of the MOS evaluation test for track vehicle mechanic, MOS Code 632.2 (USAEEC Research Study No. 49). Ft. Benjamin Harrison, IN: U. S. Army Enlisted Evaluation Center.

IX. DEVELOPMENT OF ARMY-WIDE JOB PERFORMANCE CRITERION

Walter C. Borman (PDRI)
Barry J. Riegelhaupt (HumRRO)
Lawrence M. Hanser (ARI)

This part of the effort is devoted to the identification, refinement, and development of Army-wide performance measures. Army-wide performance measures are those indicators of general performance and effectiveness not related directly to the performance of MOS-specific tasks.

The central goals of this activity are: (a) to identify aspects of soldier effectiveness that apply to all MOS; (b) to identify and/or develop valid indicators to measure these aspects of effectiveness; and (c) to establish the indicators as criteria of solder effectiveness and, where appropriate, as in-service predictors of future performance or other aspects of soldier effectiveness. In-service predictors are measures obtained after a soldier enters the Army; they predict the soldier's later performance or effectiveness in his/her military career. Measures must be identified and/or developed for both first- and second-tour performance.

Definition of Army-wide effectiveness within the general overall model of soldier effectiveness requires careful specification of the relevant criterion space. "Outcome indicators" and objective administrative indexes such as attrition, disciplinary actions, special awards, schools attended, etc., are clearly Army-wide criteria, and measures of these types of criteria are of concern in the research. A second focal point is the development of general performance and soldier effectiveness measures. An indivdiual's "worth to the Army" is conceptualized as including a relatively broad set of soldier effectiveness criteria such as organizational commitment, organizational socialization, and morale.

Special behavior-based rating scales are being prepared to measure soldier effectiveness on all important dimensions identified in the initial model development work, and supervisory, peer, and self ratings will be gathered to provide a set of Army-wide effectiveness criteria.

The Preliminary Model

To generate the initial model for the general effectiveness domain we made some preliminary hypotheses about constructs that might be considered. These constructs focus on the areas of organizational commitment, organizational socialization, and morale.

Organizational Commitment. The concept of organizational commitment (Porter, Steers, Mowday, & Boulian, 1974; Steers, 1977) refers to the strength of a person's identification with and involvement in the organization. It incorporates three kinds of attitudinal and cognitive elements: acceptance and internalization of organizational values and goals, motivation to exert effort toward the accomplishment of organizational objectives, and firm intentions of staying in the organization. It connotes a sense of loyalty to the organization as a whole and a desire to fulfill more general role requirements that come with organizational membership.

Organizational Socialization. Organizational socialization is the process by which an individual acquires the social knowledge and skills necessary to assume an organizational role (Van Mannen & Schein, 1979). Some part of this knowledge and skill is, of course, job specific. For example, training programs designed to improve the effectiveness with which a person performs job-related tasks are part of the process of organizational socialization. But there are also many other skills and knowledge necessary for effective functioning as an organizational member, which are not job specific. When the socialization process is successful, a person will acquire not only job-related skills but also new patterns of behavior with subordinates, peers, and superiors in the organization; new attitudes, beliefs, and values in line with organizational norms; and new ways of using time not formally dedicated to performing job-related tasks.

Morale. The concept of morale has traditionally been regarded as an extremely important element in military organizations. The concept of military morale is multifaceted. It seems to involve feelings of determination to overcome obstacles, confidence about the likelihood of success, exaltation of ideals, optimism even in the face of severe adversity, courage, discipline, and group cohesiveness. (Motowidlo, Dowell, Hopp, Borman, Johnson, & Dunnette, 1976).

In sum, we expect that the criterion domain of general soldier effectiveness and worth to the Army is heavily saturated with elements of organizational commitment, successful socialization, and morale. Our preliminary hypotheses, then, were that soldiers who show high levels of commitment to the Army, acceptance of Army norms, and morale are more effective soldiers in this broader sense and are also of more value to the Army.

These three broad constructs can be viewed in another way that leads to more specific hypotheses about this domain. From the combination of morale and commitment emerges a general category that can be labeled "Determination." It is basically a motivational and affective category that reflects the spirit, strength of character, or "will do" aspects of good soldiering. Morale and socialization spawn "teamwork," behaviors that have to do with effective relationships with peers and the unit. Commitment and socialization give rise to "allegiance." This taps into acceptance of Army norms with respect to authority, faithful adherence to orders, regulations, and the Army lifestyle, and adjustment and socialization to the point of wanting to continue in the soldiering role and stay in the Army. Each general category of effectiveness subsumes five more specific dimensions.

Figure 9 shows how we attempted to integrate these ideas. As shown, the most abstract and broad construct, "Soldier Effectiveness," is defined according to somewhat narrower notions of "Morale," "Socialization," "Commitment," which, with judicious mingling of conceptual elements, produce more concrete categories of "Determination," "Teamwork," and "Allegiance," which, finally, subsume more specific dimensions of soldier effectiveness. Fifteen preliminary dimensions of soldier effectiveness are discussed in the first of the associated reports at the end of this chapter.

Development of General Effectiveness Measures

The principal means being used to build new measures of general soldier effectiveness is the behavioral analysis or behaviorally anchored rating scale (BARS) technique. It is dependent on the gathering of critical

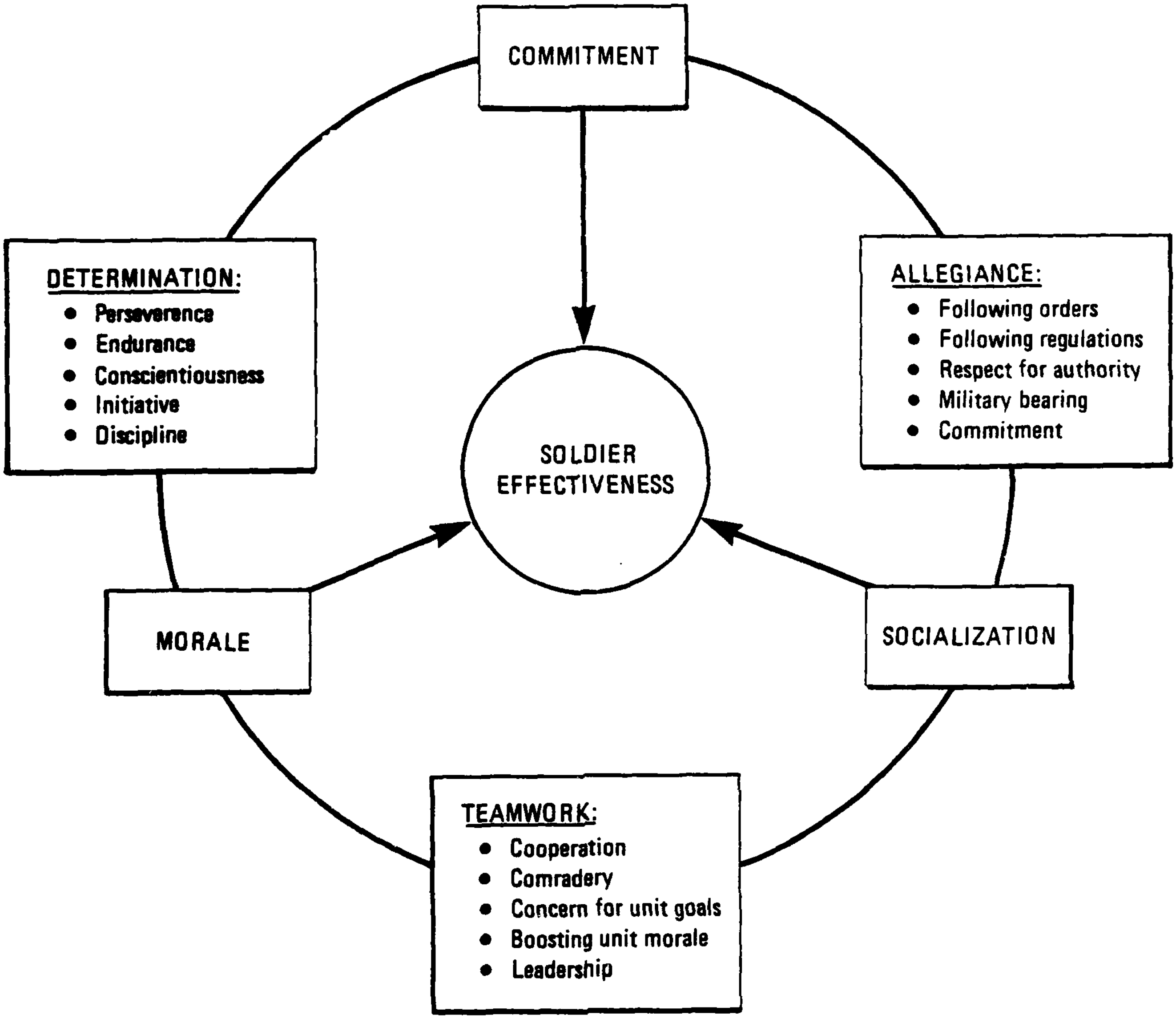


Figure 9. **A Preliminary Model of Soldier Effectiveness**

incident descriptions of job behaviors described earlier. It is also the principal means by which the model or theory of general effectiveness will be refined, revised, and developed. The general protocol used in the workshops is included as Appendix G.

Although the conceptual model showed promise as a way of depicting several important dimensions of soldier effectiveness, the intention was not to accept these dimensions at face value. Instead, we plan to use the inductive behavioral analysis method (Campbell, Dunnette, Arvey, & Hellervik, 1973; Smith & Kendall, 1963) to develop dimensions of soldier effectiveness, and then to revise the model based on what is learned in this analysis.

Results

At this time, we have conducted two BARS development workshops with a total of 14 experienced Army officers (captains and majors). In the workshops these officers generated 245 examples of first-tour soldier effectiveness, and we have performed a preliminary content analysis to explore possible dimensions emerging to define soldier effectiveness. Several other workshops will be conducted with officers and NCO to ensure good coverage of the entire target domain, but these 245 examples provide some idea of what that domain will look like.

Dimensions Emerging From the Examples

Twenty-two relatively fine-grained and specific dimensions were derived from the content analysis. They appear in the first of the associated reports at the end of this chapter.

Comparison of Model Dimensions and the Empirically Derived Dimensions

To obtain an initial idea about how the empirically derived dimensions might fit into the 15-dimension framework provided by the model of soldier effectiveness, we sorted each behavioral example into one of the model's dimensions. Then, behavioral incident membership in dimensions within the two systems was cross-referenced to provide a rough comparison between the two dimensional systems. This comparison is shown in Figure 10.

Results of this cross-referencing show first that two dimensions in the model of soldier effectiveness are not reflected in any examples. Commitment and Comradery had no incidents sorted into them. Second, Conscientiousness and Following Regulations are probably too broad, with incidents from 7 of the 22 empirical dimensions appearing in each of those categories of the model.

Third, four of the empirical dimensions do not have any representation in the model. Job Knowledge/Skill, Financial Management, Stealing/Lying, and Physical Fitness are not reflected in the model's dimensions. Fourth, there are some near one-to-one matches between the two dimensional systems. Military Bearing and Military Appearance; Boosting Unit Morale and the second Leadership dimension in the empirically derived dimensions; Perseverance (in both systems); and Discipline and Drug/Alcohol Abuse provide good matches. In the last case, however, Discipline is defined much more broadly than the content represented in the Drug Abuse examples.

Figure 10

Comparison of the Model's Dimensions and the Empirically Derived Dimensions

Model Dimensions \ Empirical Dimensions	1. Promptness	2. Job Skill	3. Finances	4. Stealing/Lying	5. Physical Fitness	6. Clean Quarters	7. Drug Abuse	8. Equipment	9. Detail	10. Following SOP	11. Initiative	12. Perseverence	13. Improve Skills	14. Appearance	15. Accept Orders	16. Courtesy	17. Follow Regs.	18. Leadership I	19. Leadership II	20. Leadership III	21. Leadership IV	22. Concern for Others
1. Following Orders						0			0						■							
2. Following Regulations	■							X	0					0			■	X		X		
3. Respect Authority														0		■	X					
4. Military Bearing														■								
5. Commitment																						
6. Cooperation											X										■	■
7. Comradery																						
8. Concern Unit Goals		0							0													0
9. Boosting Unit Morale																			■			0
10. Leadership										■								X				
11. Perseverence											0	■						0				
12. Endurance								0			0	0										
13. Conscientiousness						■		X	■		0	0	■					X				
14. Initiative		0									X							X		0		
15. Discipline							■															
16. Other		■	■	■	■						0									0		

■ = Very high overlap (>75% of behavioral examples falling in empirical dimension fall in model dimension)
X = Intermediate amount of overlap (15-74% of examples falling in empirical dimensions fall in model dimensions)
0 = Small overlap (at least 1 example but <15% of examples falling in empirical dimensions fall in model dimensions)
Blank = No overlap.

Finally, and this is perhaps the most salient result, there seems to be considerable overlap between the dimension content in the two systems, but often the configuration of that content differs. Essentially, elements of the dimensions in the two systems are put together differently.

The most important objective in developing dimensions is to achieve the purposes of the project, and this overriding concern will guide future efforts to integrate empirical information with the theoretical model. Dimensions will be developed and defined to reflect in a comprehensive and, at the same time, efficient manner the domain of soldier effectiveness. Dimensions will be derived to provide raters using rating scales based on the dimensions with an easy-to-understand, highly face-valid rating format that reflects accurately the behavioral requirements of this domain.

This approach has the advantage of forcing a broad perspective on the criterion domain. It points out potentially important elements of individual effectiveness that might be overlooked by purely inductive approaches to job and task analysis. For this reason, we believe the model is useful for guiding efforts to impose structure upon the complexity of what "soldier effectiveness" might mean in the Army.

Development of Archival Records as Army-wide Criterion Measures

A major activity within our overall program of performance criterion development is to explore the use of archival administrative records in the formation of first-tour criteria and in-service predictors of soldier effectiveness. The Enlisted Masterfile (EMF), the Official Military Personnel File (OMPF), and the Military Personnel Records Jacket (MPRJ) are the records sources that contain administrative actions that could be used to form measures of first-tour soldier effectiveness.

As mentioned in the Research Plan, a serious difficulty in using administrative records to form soldier effectiveness criteria is that the material in the records very often reflects only exceptionally good or exceptionally poor performance. Measures of performance based on infrequently appearing personnel actions could have very little variance, i.e., everyone has about the same score on them. A strategy for dealing with the skewness in records data that results from low base rates is to combine records of different kinds of events and actions into more general indexes. When scores on administrative measures that reflect the same underlying constructs are combined, the base rate might improve to a level where significantly higher correlations with other variables would be possible. Consequently, before administrative records composites can be formed and assigned to performance constructs we must determine which administrative indexes have sufficient variance and acceptable base rates to warrant inclusion in composite formation, and which records distinguish effective from ineffective soldier performance. As such, we must identify which administrative actions reflect Army-wide soldier effectiveness and from which archival sources it is most feasible to obtain them.

First Year Activities

Accordingly, during the past 6 months we have begun a detailed examination of the three archival data sources and an analysis of the feasibility of developing criterion indexes from them. The Enlisted Masterfile is a computer file corresponding to every enlisted individual currently on the U.S. Army payroll. It contains a large number of variables for each individual ranging from pay grade to Skills Qualification Test (SQT) scores to appraisal ratings in the form of the Enlisted Efficiency Report (EER). A complete description of the variables available from the EMF is given in the Longitudinal Research Data Base (LRDB) plan shown in an earlier chapter.

An initial review of the EMF was carried out by interviewing several key Army personnel who have knowledge of and/or responsibility for the EMF. The variables which appeared to hold the most promise are: (1) reason for separation, (2) reenlistment eligibility, (3) reenlistment eligibility bar, and (4) weighted Enlisted Evaluation Report score. With the exception of the weighted EER, these measures may more appropriately be considered outcomes that result from performance, rather than evaluations of performance per se. In theory, the EER variable on the EMF, which is a weighted average of a soldier's last five EER should be an excellent variable. As a practical matter, however, its usefulness may be limited. Since EER are only done on soldiers in grades E5 and above, only a small percentage of the first-tour cohort is likely to have had even one EER at the time of data collection. Secondly, in the past few years EER scores have tended to cluster at the maximum of 125. Thus, distinguishing effective from ineffective performers on the basis of EER scores may not be possible. A definitive answer regarding suitability of EMF variables for use as criteria is dependent on our own comprehensive examination and analysis of the existing computer records and existing EMF documentation. That analysis is currently in progress.

Information in the Official Military Personnel File (OMPF) is stored on microfiche. Depending upon their purpose, documents are filed in one of three sections:

(1) The performance (P) fiche. The P fiche is the portion of the OMPF where performance, commendatory, and disciplinary data are filed.

(2) The service (S) fiche. The S fiche is the OMPF section where general information and service data are filed.

(3) The restricted (R) fiche. The R fiche is the OMPF section for historical data that may be biased against the soldier when viewed by selection boards or career managers. For this reason release of information on this fiche is controlled.

The initial examination of microfiche records was conducted by a combined team of four research staff members. They conducted a 3-day site visit at the Enlisted Records and Evaluation Center (EREC) at Ft. Benjamin Harrison. A total sample of 465 individual soldiers was drawn from a variety of MOS. If a microfiche packet could be found for the individual, each record in the packet was examined by a staff member and a variety of information items was recorded. A summary of the major findings is as follows:

(1) Of the 414 microfiche packets that could be located, 278 contained only a service fiche while 136 contained both a service and performance fiche.

(2) Of the 136 soldiers in our sample who had performance fiche, 44 of them (32 percent) were prior service members. Of these 44 soldiers 20 had EER in their files. Six of the soldiers had 2 EER apiece for a total of 26 EER. The distribution of EER scores was:

Frequency	Score
13	125
3	123-124.9
5	121-122.9
5	121

(3) A total of 52 Articles 15's was issued to the 136 soldiers who had a performance fiche.

(4) Sixty-three awards were received by the 136 soldiers. 41 of these awards were for completion of a training course.

(5) Twelve letters of appreciation/commendation appeared on the performance fiche.

(6) Of the 136 soldiers, 26 were credited with having attended a school. Two of these soldiers attended two schools apiece.

After examining the microfiche and the regulations governing their composition, as well as interviewing cognizant officials, we reached two conclusions:

(1) The data which exist in the OMPF are not nearly as complete or timely as we would like them to be. For grades E5 and below, which are the grade levels that enlisted personnel in the FY83/84 first-tour cohort research will be, there is an 8-12 month backlog from the time a personnel action is taken to the time it appears on microfiche at EREC.

(2) Whether performance-related material for a given soldier appears in the OMPF depends in large part on his or her commanding officer CO. If a commendatory or discipli-disciplinary action is taken on a soldier the CO has three choices. He/she can either send it to EREC to be filmed on the soldier's performance fiche, his restricted fiche, or neither. We did not see the restricted fiche and, given their sensitive nature, it is questionable at this time whether we will gain access

to these fiche. Keeping in mind the 8-12 month backlog, even if we are granted permission to view the restricted fiche, administrative index data may not be available when we need them. The CO's third alternative is of greatest concern. While AR 640-10 lists specific disposition for each document authorized to appear in the OMPF, the individual CO has discretionary power regarding which commendatory letters, letters of reprimand, and Articles 15's, for grades E5 and below, get forwarded to EREC for inclusion on the OMPF. It is therefore possible for a soldier not to have a performance fiche but to have one or more Articles 15's in his Military Personnel Records Jacket (MPRJ).

Because of the limitations in the microfiche records, determination of the discrepancy in type, quantity, quality, and timeliness of the information contained in a soldier's MPRJ (201 file) and the information that exists in the OMPF appear to be of vital importance. The MPRJ (201) file is the primary mechanism for storing information about an individual's service record. It is the most complete and up-to-date record, and it physically follows the individual wherever he or she goes. It is located at the Military Personnel Office (MILPO) that services the soldier's unit.

It is our educated guess that the MPRJ will be a much richer and more timely source of information than the microfiche records. It would, however, be considerably more expensive to collect data from these personnel files than from records at EREC. Information in the files has not been condensed, and the files are not stored in a centralized location. They move with the individual. However, when data from both sources have been collected and compared, a decision can be made concerning the trade-offs of data collection at EREC versus individual field units.

To carry out the feasibility study on the microfiche vs. 201 files we plan to collect records data from MPRJ and OMPF for a sample of 750 soldiers, 150 soldiers in each of five MOS at five Army posts, to determine the type, quantity, quality, and timeliness of information that appears in these two sources. We will also be able to determine whether there are significant differences in frequency of administrative actions across MOS and posts.

A records collection form and guidelines appear in Appendix H. Once this data collection is completed we will identify those administrative records useful for measuring soldier effectiveness and know from which sources to collect them.

Associated Reports

Task 4 personnel have been engaged in a wide range of activities during the first year of the project. The reports which follow represent a diverse sample of the issues that have confronted Task 4 this past year.

One of the major activities assigned to Task 4 included the development of measures of Army-wide performance. Toward this end, the paper by Borman et al. focuses on efforts at developing a behaviorally based definition of soldier performance. The constructs which have emerged from this work will form the core around which Army-wide measures of performance will be developed. These constructs will also serve to aid in the identification of promising areas for predictor development.

The paper by Eaton et al. represents an initial attempt at expressing the impact of personnel decisions in terms of practial utility. A number of methods for doing so have appeared in the literature. This paper provides a brief review of how these methods may apply to the military as well as some sample calculations and further suggestions.

Task 4 personnel have also been exploring the possibilities of combining administratively collected information into a measure or measures of performance. Hanser and Grafton discuss some of the difficulties which might be encountered when working with these types of archival data. They also suggest methods which may be useful for avoiding some of these difficulties where possible and dealing with them if they cannot be avoided.

Putting the "Dollars" into Utility Analyses *

Newell K. Eaton, Hilda Wing & Karen Mitchell
U.S. Army Research Institute
for the Behavioral and Social Sciences

Testing to improve selection and classification decisions is a normal part of entering into employment in most large organizations and in many small ones (Friedman & Williams, 1982). The benefits of such testing have been questioned on an increasingly frequent basis, both from the perspective of fairness and in terms of the benefits accrued from sometimes costly selection programs. Cost benefit analyses of selection and classification procedures frequently stress the price of recruiting and testing, while neglecting the dollar benefits of resulting increases in productivity. This oversight leaves managers unable to estimate the dollar value of selection and classification procedures to the organization.

The utility of implementing various personnel policies is important data for management decision-making. The costs of implementing personnel policies must be weighed against anticipated benefits. Implementation costs are usually couched in real dollar terms. Costs associated with salaries, space, overhead for test administration, fees, per diem paid to applicants, computer time and personnel for scoring, etc. are easy to estimate. Benefits accrued from implementation of personnel policies, however, are not as clearly identifiable in dollar terms. Judgments of the net positive impact of implementing given personnel policies are, therefore, difficult to make.

Early approaches to estimating the benefits associated with testing programs focused on the correlation between predictor measures and criterion performance. Kelly (1923) defined the "Index of Forecasting Efficiency" as $E = 1 - \sqrt{1 - r}$. A second index, the "Coefficient of Determination" was proposed by Hull (1928). This index is simply r^2, the proportion of criterion variance accounted for by the predictor measures. Both approaches suggest that only testing programs with high validity coefficients are beneficial.

However, important management information is neglected by mathematical procedures involving only validity or relative variance. Taylor and Russell (1939) incorporated information about selection ratios, the proportion of applicants selected to those who apply. Their computations dichotomize criterion performance into satisfactory and unsatisfactory, and clearly show that with sufficiently small selection ratios, even tests with low validities can lead to important increases in the numbers of successful performers. Increases in the proportion of successful performers predicted by the selection instrument can be linked to the increased usefulness or utility of an instrument.

---

*Paper presented at the 91st Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

Brogden (1949) and Cronbach and Gleser (1965) also addressed utility estimation. Their formulations dealt with continuous levels of criterion performance rather than the dichotomy used by Taylor and Russell (1939). They linked normally distributed performance levels to the values estimated for those performance levels, in dollars. A useful formula for the gain in productivity, or utility (U$), obtained by using valid selection procedures includes (1) Ns, the number of individuals selected (2) SD$, the standard deviation of performance, scaled in a utility metric such as dollars; and (3) the average performance expected on the criterion by the selected group as estimated from a valid predictor, given by Rxy Zx:

$$U\$ = Ns\ SD\$\ Rxy\ Zx \quad (1)$$

To account for the cost (Ct) of testing each of Na applicants the formula was adjusted to

$$U\$ = Ns\ SD\$\ Rxy\ Zx - Na\ Ct$$

A more complete description of such formulations can be found in Cronbach and Gleser (1965) and Hunter and Schmidt (1982).

While the values of most of the variables on the right hand side of the Brogden-Cronbach-Gleser formulas are known, the estimation of SD$, the standard deviation of performance scaled in dollars, is problematic. A recent review by Hunter and Schmidt (1982) reports that only two published efforts have attempted the computation of SD$ using cost accounting methods.

An alternative to cost accounting estimates is to estimate the dollar values to the organization of individual or unit performance at the 50th percentile level, the 85th percentile level (one standard deviation above the mean), and sometimes, the 15th percentile level (one standard deviation below the mean). The dollar difference between the 15% and 50%, and 50% and 85% points provides an estimate of SD$. This technique was used by Cascio and Silbey (1979) with second level managers in food and beverage sales (Mean = $30,000, SD$ = $9,500); by Schmidt, Hunter, McKenzie, and Muldrow (1979) with computer programmers (SD$ = $10,413); by Hunter and Schmidt (1982) with budget analysts (SD$ = $11,327); and by Bobko, Karren, and Parkington (1983) with insurance counselors (Median = $96,000, SD$ = $56,950). In the last study, actual sales data were also available and yielded sales-based statistics which were close to those obtained by the rating method (Median = $117,300, SD$ = $52,308).

The SD$ estimations reported above were derived in contexts where performance was easily measurable in dollars (although Mack, Schmidt, and Hunter, Note 4, did work with park rangers). The SD$ estimation questions developed by Hunter and Schmidt (1982) asked for estimates of the "value to the agency" of various performance levels. The questions were preceeded by instructions to "consider the cost of having an outside firm provide these products and services" (Schmidt et al., 1979, for computer programmers), or to "consider what the cost would be of having an outside consulting firm produce these products and services" (Hunter & Schmidt, 1982, for budget analysts). Both Cascio and Silbey (1979) and Bobko et al (1983) framed questions in terms of performance value and estimates of total yearly dollar sales.

Hunter and Schmidt (1982) reviewed performance value and SD$ estimation work and looked at SD$ in relation to annual employee salary. They estimated that the true average for SD$ falls between 40% and 70% of annual salary. In a more recent paper (Hunter and Schmidt, 1983) they suggested the use of 40% of annual wage as an estimate of SD$. This rule-of-thumb approach to estimating SD$ enables decision makers to make rapid assessments of test utility. Such assessments could indicate the desirability of further research or utilization of test procedures. They may not be relevant in settings where complex, expensive equipment is involved, or where work cannot be contracted out, such as combat positions in the military.

An alternative utility estimation technique is to focus instead on the effectiveness of the system of which the individual is a unit. One can call techniques like these "system effectiveness techniques". Such techniques incorporate the concept of the number of extra units required for various levels of improved system performance, and their overall cost, as estimates of utility. Research reported by Arnold, Rausdenberger, Soubel and Guion (1982) described procedures for estimating system value in terms of the number of units required to reach fixed levels of performance. They costed out the price of additional units required at lower levels of unit productivity to describe the economic importance of selection procedures for steel workers. Specifically, their research suggested that a worker at the 99th percentile was at least twice as productive as one at the 1st percentile. The difference between the 99th and 1st percentiles is six standard deviation units. Arnold et al divided the estimated cost per worker ($18,000 per year) by six to obtain their estimated SD$ = $3,000.

Wallace (1982) used a similar approach in estimating the cost effectiveness of valid selection procedures for Army tank commanders. The performance of such weapons systems in combat can be indexed by relatively simple performance parameters such as the probability of hits. Wallace used simulation techniques to determine that, in terms of hit rate, a high performance tank commander is as effective as five lower performance tank commanders. Using Wallace's data, the difference between the high performer (at about the 70th percentile) and the low performer (at about the 35th percentile) was estimated to be one standard deviation unit. Wallace asserts that each tank system costs approximately three million dollars per year. If only high performance tank commanders were used, fewer tank systems would be required. Wallace estimated that savings resulting from selection procedures which lead to a greater number of selected high performers may be hundreds of millions of dollars per year in weapons systems expenditures.

If one works from Wallace's estimates for tank commanders, SD$ may well be substantially in excess of a tank commander's salary. Of course, such techniques seem most appropriate where variation in the performance of the system is due primarily to the variation in performance of one job within the system. This would be true for the tank system and the tank commander position used in Wallace's work. This research also points up the feasibility of deriving estimates of performance value and SD$ by obtaining estimates of required numbers of units at different performance levels to yield equivalent systems results. When performance can be indexed by the quantity of the output, differing numbers of units at varying performance or output levels can be equated. Further, by considering the cost of additional units needed to achieve additional output, one can compute the value of improved performance yielding the same additional output with the original number of units.

The purpose of this research is to examine both SD$ estimation techniques and systems effectiveness techniques using data from complex, expensive equipment systems - Army tank systems. Systems effectiveness will be discussed in terms of changes in the numbers of units required, at differing levels of performance, to achieve the same level of output. Additions or decrements in the number of units, at varying performance levels, can be linked to additions or decrements in the effectiveness of the system. Procedures will be presented for linking these changes to changes in the utility of the system, in dollars or other utility metrics.

Method

First, the standard SD$ estimation techniques were used to examine the SD$ and utility of improved performance of tank commanders, the key personnel in a tank system. Second, mathematical derivations were developed linking the Brogden-Cronbach-Gleser formula with the systems effectiveness concepts described above.

The SD$ Estimation Procedures

Pilot work for the research was based on a questionnaire modeled after that of Hunter and Schmidt (1982). The language of the questionnaire was targeted to the research sample and references to contracting out for equivalent services were deleted. Information was provided on tank system costs associated with salaries, recruiting and equipment. Neither three psychologists acting as reviewers nor five tank commander (TC) supervisors were able to respond meaningfully to the questionnaire. The consensus was that there was no frame of reference for estimating performance value because one could not define and/or purchase the services of a tank commander in combat.

Because the first questionnaire didn't work, a second questionnaire, with additional referent information, was developed for use with Army personnel assigned as M1 TCs. The revised questionnaire provided information on the training and responsibilities of TCs and the cost of a tank. Like the first questionnaire, the second form requested information about the value of performance at the 50th ("average") and at the 85th ("superior") percentile. This is the basic information requested in the SD$ estimation technique. Also, the revised questionnaire asked for the numbers of superior TC's required to match the performance of a company (17) of tanks manned with average TCs. Finally, respondents were asked to indicate the certainty with which they made these judgments. The questionnaire follows.

VALUE ASSESSMENT QUESTIONNAIRE

The purpose of this questionnaire is to obtain your estimates, or opinions, of the relative value to the Army of "average" and "superior" tank commanders. Because these are estimates or opinions, you may not be too sure of your answer. PLEASE MAKE YOUR BEST GUESS, then indicate how sure you are of your answers.

For the purposes of this questionnaire an "average" tank commander is an NCO or commissioned officer whose performance is better than about half of his fellow TC's. A "superior" tank commander is one whose performance is better than 85% of his fellow tank commanders.

The first question deals with relative value. For example, if a "superior" clerk types 10 letters a day and an "average" clerk types 5 letters a day then, all else being equal, 5 "superior" clerks have the same value in an office as 10 "average" clerks.

In the same way, we want to know your estimate or opinion of the relative value of "average" vs. "superior" tank commanders in combat.

1. I estimate that, all else being equal, ____ tanks with "superior" tank comanders would be about equal in combat to 17 tanks with "average" tank commanders.

2. How certain are you of your answer?

a. Positive. The answer is exact.
b. Pretty sure. The answer is correct within 1 tank either way.
c. Fairly sure. The answer is correct within 2-3 tanks either way.
d. Somewhat doubtful. The answer is correct within 4-5 tanks either way.
e. Totally uncertain. The answer is off by more than 5 tanks either way.

Next, we want you to try to estimate the value to the Army of an "average" tank commander during combat.

You should know that an "average" tank commander

- has more than 4 years experience on tanks.
- commands a tank worth more than $1 million.
- is responsible for three other crewmen.

Of course, the value to the Army of a tank commander in combat may be considerably more, or less, than his salary and benefits. On making your estimate you may wish to consider the time and training the tank commander has received, his salary and benefits, the responsibility of his position, and the cost of his tank and supporting units.

3. Based on my judgment, I estimate the value to the Army of an "average" tank commander in combat at $____________ dollars per year.

4. How confident are you of your answer?

a. within $1,000.
b. within $10,000.
c. within $100,000.
d. Other ______________.

Now we would like to ask the same questions about the "superior" tank commander. Remember, we are calling a "superior" tank commander one who is better than 85% of his fellow tank commanders.

5. Based on my judgment, I estimate the value to the Army of a "superior" tank commander in combat at $____________ dollars per year.

6. How confident are you of your answer?

a. within $1,000.
b. within $10,000.
c. within $100,000.
d. Other ______________.

Subjects. The revised second questionnaire was administered to 48 M1 tank commanders enrolled in advanced training at a Continental U.S. Army post and to five other enrollees (one M48 Armor crewman and four U.S. Marine tank commanders). All subjects were male. Racial/ethnic identification of Ss was not obtained. It can be presumed that the sample was representative of the corps of CONUS Army TC's. The median number of years experience as a tank crew member was nine.

Systems Effectiveness Technique - Derivation

Let the cost of a single unit in a system be CU.

Let the overall systems effectiveness be Y. This may be achieved with varying numbers of units depending on the performance of the units. Or

$$Y = n1Y1 = n2Y2 \quad \ldots \; niYi \qquad (2)$$

where ni = number of units at performance level i, and

Yi = mean performance of units at level i, on ratio scale.

Examples of performance scales useable in this formula are probability of hits per firing (Army tank commander), number of convictions per year (detective), number of pupils achieving given standard (teacher), or other frequency-type variables.

Reducing cost while holding systems effectiveness constant

In (2), if Y1 is less than Y2, then n1 must be greater than n2. The savings in dollars (S$) is the decrease in units required at the higher level of performance, times the cost per unit.

$$S\$ = CU\ (n1 - [illegible] \qquad (3)$$

From (2) $n1Y1 = n2Y2$

Thus $$n1 = \frac{n2Y2}{Y1}$$

Substituting into (3)

$$S\$ = \frac{CU\ n2\ (Y2 - Y1)}{Y1} \quad (4)$$

Improving systems effectiveness while holding cost constant

The number of extra units operating at the initial performance level which are needed to achieve the improved performance of the system obtained with ni units operating at the Y2 performance level more than the Y1 performance level is

$$DELTA\ N = \frac{N1\ (Y2-Y1)}{Y1}$$

In a system where improved performance Y2 is obtained from each of the initial n1 units, the overall improvement in system performance is

$$n1Y2 - n1Y1 = n1\ (Y2 - Y1) \quad (5)$$

The dollar value of improved performance is equivalent to the extra number of lower performing units needed times the cost per unit.

$$U\$ = \frac{Cu\ n1\ (Y2 - Y1)}{Y1} \quad (6)$$

Simply stated, the value in dollars of achieving an overall improvement in performance in a system equals the cost of adding the number of units required to effect the improvement with units operating at the initial performance level.

Estimating U$ using SD in performance units

The basic Brogden-Cronbach-Gleser formula works in any metric (m); U and SD need not be expressed in dollars. The overall improvement in system performance is

$$Um = NsSDm\ Rxy\ Zx$$

In output units of performance, Y, this is equal to (5)

$$n1\ (Y2 - Y1) = N\ SDy\ Rxy\ Zx$$

Substituting into (6)

$$U\$ = \frac{Cu\ N\ SDy\ Rxy\ Zx}{Y1} \quad (7)$$

Formula (7) more conveniently describes the utility in dollars of selection; this formula uses SDy, the standard deviation in output units of performance, rather than SD$, the standard deviation of performance in dollars.

Estimating SD$ using cost and performance data

Setting (1) and (7) equal

$$U\$ = Ns\ SD\$\ Rxy\ Zx = \frac{Cu\ N\ SDy\ Rxy\ Zx}{Y1}$$

and solving for SD$

$$SD\$ = \frac{Cu\ SDy}{Y1} \quad (8)$$

Or, SD$ equals the cost per unit times the ratio of the SD of performance to the initial mean level of performance. This proportion may be easily estimated in many situations. It is interesting to note that this parallels the Hunter & Schmidt notion that SD$ may be linked to some percentage of salary. Here, Cu is the cost of the unit in the system - equipment, support, and personnel - rather than salary. However, estimates from both (7) and (8) are appropriate only when the performance of the units in the system is to a major extent a function of the job under investigation. To the extent that it is not, corrections to these formulae would be required.

Analyses

Estimation of SD$

There are several ways to estimate SD$. The first is to use the Schmidt and Hunter (1983) suggestion of 40% of annual salary. In 1983 the base pay for Army enlisted personnel with ten years of service at the ranks expected for tank commanders ranged from $15,500 to $17,000. Non-taxable allowances for such items as housing could amount to more than $8,000 for a married serviceman with dependents. An estimate of an equivalent civilian salary would be no more than $30,000 per year.

The second and third ways to estimate SD$ were taken directly from the estimates given for the values of average and superior TCs Estimates given for the value of an average tank commander were tabled and the median was determined. The same procedure was followed to obtain the median estimate of the value of a superior tank commander. Confidence estimates for these two judgments would also be tabled and analyzed. For each judge, the difference between the values given for the average and superior tank commander was calculated. These differences were tabled and the median was determined. The difference between the median estimates for the average and superior TCs is the second means to estimate SD$ while the third means is to use the median of the differences between the estimates for the average and the superior TCs.

The fourth way of estimating SD$ was determined with the systems cost approach, as follows. The estimate given for the number of superior TCs equivalent to a company of 17 average TCs were tabled, along with the confidence judgments for these estimates. The median of the number of superior TCs required and of the confidence of these judgments, were determined. Based on these a value was determined to represent the number of superior TCs judged as equivalent to 17 average TCs. The fourth way to estimate SD$ requires the calculation of the fraction of a superior TC equivalent to one average TC, or the inverse of the ratio of the number of superior TCs judged as equivalent to a company of average TCs, to the number of average TCs in this company. The inverse of this fraction times the (median) estimated value of the average TC provides an additional estimate of the value of a superior TC. The difference between this additional estimate and the median estimate of the average TC provides an estimate of SD$.

The fifth way of estimating SD$ uses (8). The cost of a tank per year, Cu, as estimated by Wallace (1982) is $3,000,000. We preferred the more conservative figure of $300,000 per year. Criterion related validity research on tank crew performance (e.g., Eaton, Bessemer & Christiansen, 1979) suggest that meaningful values for the ratio SDy/Y1 range from .2 to .5. We selected the more conservative value of .2.

Estimation of Utility or U$

To determine the utility of improved selection procedures for Army TCs, one may use either (1) or (7). For both, N or Ns is the total number of TCs, which can be estimated at 2500. Let us assume a selection ratio of 50% with a validity coefficient of .30. This leads to a predicted level of performance for the selected group of .24 standard deviation units above the mean. As in the fifth way to estimate SD$, let SDy/Y1 be .2.

The five estimates of SD$ as developed above can be substituted into (1) to yield five estimates of U$. If Cu = $300,000, then (7) also yields an estimate of overall utility.

Results

Estimating SD$

Percentage of salary. If a typical TC's salary is $30,000 per year, then 40% of this is $12,000, for a first estimate of SD$.

Estimates of value. The estimates of value for both average and superior TCs were skewed and very broad. For the average TC, the median estimate was $30,000. With the first quartile at $17,000 and the third quartile at $100,000. The distribution appeared somewhat bimodal, with five judges each giving estimates of $20,000 and $30,000, and six judges giving the estimate of $100,000. The mean of the distribution was $340,198 with a standard deviation of $952,798. The distribution is shown in Figure 1.

For the superior TC, the median estimate was $50,000 with the first quartile at $30,000 and the third quarter at $300,000. The distribution was highly skewed and almost trimodal, with five judgments each at $20,000 and $50,000 and six judges at $100,000. There were nine estimates of $1,000,000 or more, including one at positive infinity which we set equal to $100,000,000. The distribution is shown in Figure 2. The value given yielded a mean of $2,899,806 and standard deviation of $14,071,535. Of interest were the confidence estimates (questions 2 and 4). Most (70%) of the subjects indicated their estimates were accurate within $10,000, with 11% not responding, and 19% indicating confidence of more than $10,000.

The difference between the median estimates of superior and average performance value (the second way of estimating SD$) is $20,000. The median difference between average and superior TC performance (the third way of estimating SD$) is $15,000. The first quartile was $7,000 and the third was $85,000.

Inspection of the mean $ values given average and superior performance indicated that a SD$ of about $2,500,000 would be obtained.

Estimates of equivalence. The modal response given as the number of superior TCs judged as equivalent to a company of 17 average TCs was 10. The median estimate was 9 and the median confidence was that the estimate was no more than 1 tank off. Nearly all estimates were included in the range of 4 to 12 superior TCs as equivalent to 17 average TCs, and that this estimate was no more than 3 tanks off. The response "10" was judged to be a representative value of central tendency.

If 10 superior TCs are judged equivalent to 17 average TCs, and an average TC is "worth" $30,000 per year, then a superior TC is "worth" 17/10 times $30,000, or $51,000. This figure is almost the same as the median estimate of the value of a superior tank commander. Hence, the fourth way of estimating SD$, of obtaining the difference between the superior equivalents to average TCs, is $21,000, close to the value for the second way.

Estimates from system effectiveness. Substituting into (8) with Cu = $300,000 and SDy/Y1 as .2 produces SD$ = $60,000. This fifth way produces a value almost three times as large as the next largest value, and five times the size of the value calculated at 40% of annual salary.

All of the values calculated for SD$ are displayed in Table 1.

Estimation of U$

Using Ns = 2500, RxyZx = .24 (as expected from a selection procedure with validity of =.30 and a selection ratio of 50%), the utility in dollars, or U$ was calculated via (1) with each of the five above described values for SD$. Using Cu = $300,000 and SDy/Y1 = .20, as above, utility was also calculated with (7). The results are displayed in Table 2.

Discussion

The various methods of estimating SD$ and U$ do not lead to the same results. While using 40% of annual salary leads to the lowest value, estimating values based on medians of average and superior performance produces larger values which fall within the 40 - 70% range specified by Hunter and Schmidt. Although the raters indicated confidence in their judgments, the distribution is far from comforting. The extreme skewing of judgments, and bi/tri modal nature, suggested that they may have not been made based on the same considerations and values. This would indicate the difficulty of making such judgments when the cost of contracting out the work is unknown and/or other financially intangible factors are involved. The higher values for SD$ and U$ result from the systems cost technique. These are probably as reliable as any estimates. It is also plausible that these systems cost estimates are below the "true" values. A tank probably costs more than $300,000 per year to operate, if not the $3,000,000 proposed by Wallace (1982); the selection ratio is probably less than the 50% used here, which implies a higher level of performance, hence greater ability, than assumed here, and the ratio of SDy to Y1 is probably more than .2.

One could argue that the value of improved performance of tank commanders, the basis for the salary and estimation procedures, is but a part of that value of the tank systems, the basis of the system costs procedures. If one assumes the estimates to be accurate, they suggest the tank commander to contribute about one-third of the systems cost. An empirical question is the size of one contribution of the TC to the system: we do not yet know the answer. This is an important issue because the TC is a high responsibility position, such as are other operations of expensive equipment and managers and supervisors. Both tank analytical and rational judgments, as well as the lore within the armor community, however, suggests a far greater contribution of the tank commander. It was the assumption of the overweighing impact of the tank commander to system performance that led to our initial thoughts on the systems effectiveness formulae derivations.

The two additional limitations we see of the systems cost technique are, first, the problem of estimating the cost of a unit in the system and, second, its restriction to those situations when the quality of performance can be indexed by quantity. Wallace's (1982) estimate of unit cost per tank (CU) may seem excessive but, it is an estimate which can be and generally frequently is made in accounting departments. It also can be adjusted if it appears out of line. As Hunter and Schmidt (1982) note in discussing supervisor estimates of SD$, ball park estimates of SD$ can be made with minimal fuss; the same statement is true of estimates of Cu. Second, quality as indexed by quantity may not be meaningful in some situations. Prospective users should question, however, whether qualitative indices may not be made into quantitative ones. For example, a police department may decide that conviction of one murderer is equivalent to the conviction of several burglars. It is likely that such equations are being used to compare performance of different individuals, albeit informally and, possibly, inconsistently.

Table 1

Estimate of SD$

Procedure	Dollar Figure
Salary	
1. Forty percent of annual salary	$12,000
Estimation	
2. Difference between median estimates of value	20,000
3. Median of differences between estimates of value	15,000
4. Difference between estimate of fractional superior equivalent and median average	21,000
Systems Cost	
5. Cost times ratio of standard deviation to mean performance	60,000

Table 2

Estimate of Utility of 2500 tanks

Procedure	Formula	SD$	Utility
Salary			
1. Salary Percen age	(1)	$12,000	$ 7,000,000
Estimation			
2. Difference between medians	(1)	$20,000	$12,000,000
3. Median of differences	(1)	$15,000	$ 9,000,000
4. Superior equivalents	(1)	$21,000	$12,600,000
Systems Cost			
5. Cost times performance with SD$ or SDy	(1) or (7)	$60,000	$36,000,000

Figure 1. Estimation of value in dollars per year of average tank commander.

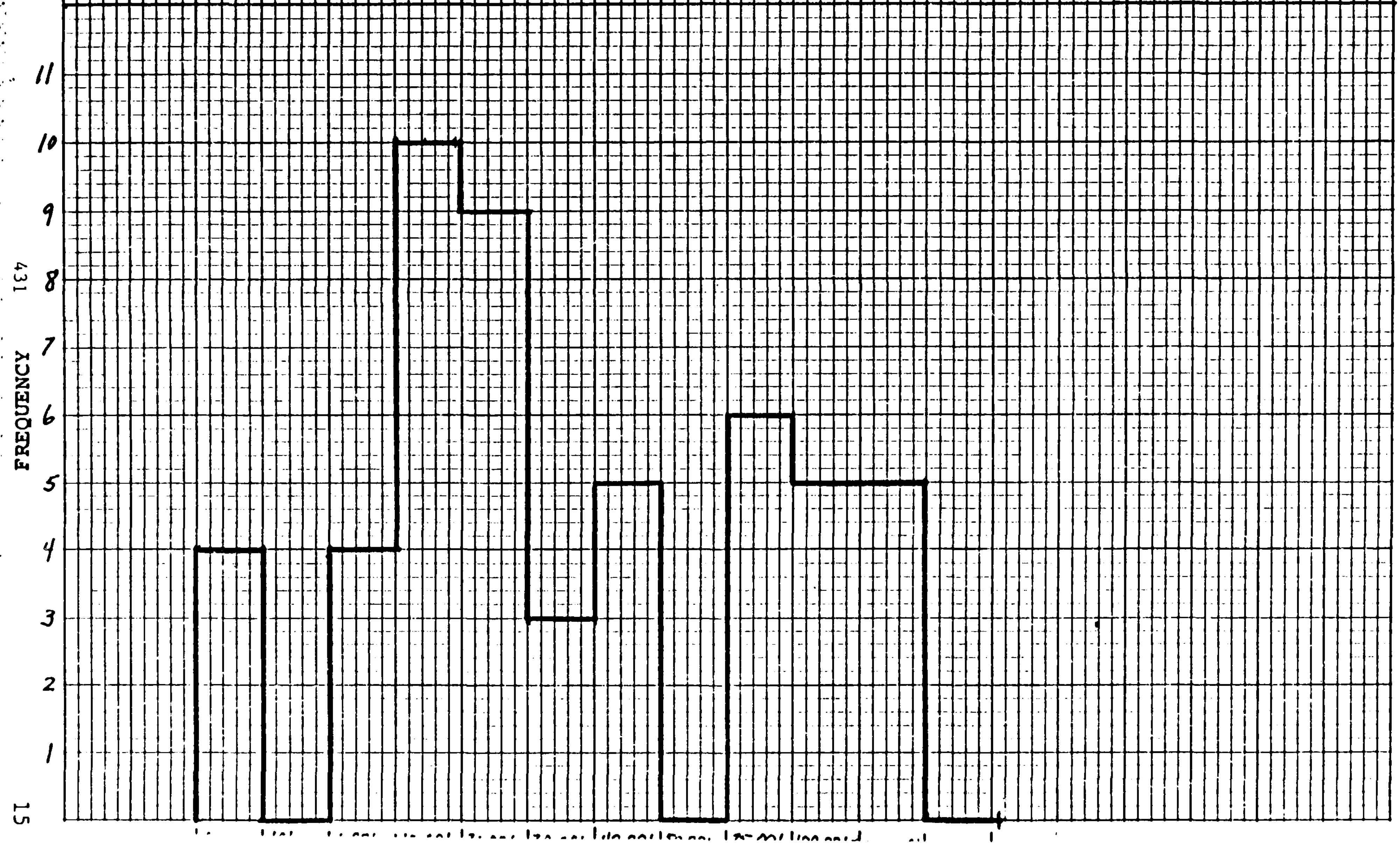
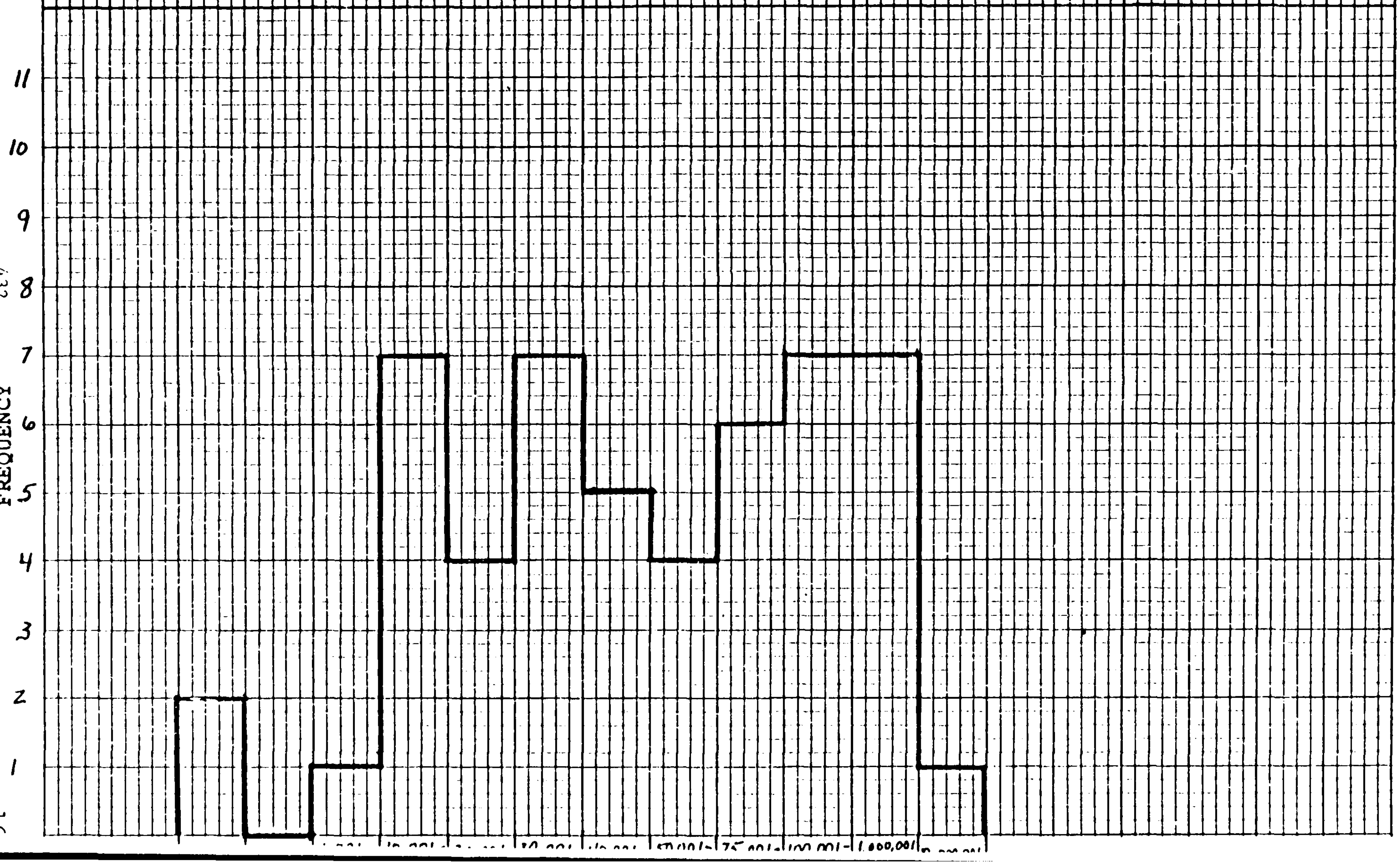


Figure 2. Estimation of value in dollars per year of superior tank commander.



References

Arnold, J. D., Rauschenberger, J. M., Soubel, W. G. & Guion, R. M. Validation and utility of a strength test for selecting steelworkers. Journal of Applied Psychology, 1982, 67, 5, 588-604.

Bobko, P., Karren, R., & Parkington, J. J. The estimation of standard deviations in utility analyses: An empirical test. Journal of Applied Psychology, 1983, 68, 170-176.

Brogden, H. E. When testing pays off. Personnel Psychology, 1949, 2, 171-183.

Cascio, W. F. & Silbey, V. Utility of the assessment center as selection device. Journal of Applied Psychology, 1979, 64, 2, 107-118.

Cronbach, L. J. & Gleser, G. C. Psychological Tests and Personnel Decisions. Urbana: University of Illinois Press, 1957, 1965.

Eaton, N. K., Bessemer, D. W., & Kristiansen, D. M. Tank crew position assignment. (Technical Report 391). Alexandria, VA: U. S. Army Research Institute, October 1979.

Friedman, T. & Williams, E. B. Current use of tests for employment. Ability testing: Uses, consequences and controversies. Part II: Documentation section. Washington, DC: National Academy Press, 1982, 99-169.

Hull, C. L. Aptitude Testing. Yonkers, N. Y.: World Book, 1928.

Hunter, J. E. & Schmidt, F. L. Fitting people to jobs: The impact of personnel selection on national productivity. In E. A. Fleishman & M. D. Dunnette (Eds.), Human performance and productivity: Volume 1. Human capability assessment. Hillsdale, N.J.: Erlbaum, 1982.

Hunter, J. E. & Schmidt, F. L. Quantifying the effects of psychological interventions on employee job performance and work-force productivity. American Psychologist, April 1983, 473-478.

Kelly, T. L. Statistical method. New York: Macmillan, 1923.

Mack, M. J., Schmidt, F. L., & Hunter, J. E. Estimating the productivity cost in dollars of minimum selection test cutoff scores. Washington, DC: U.S. Office of Personnel Management, Office of Personnel Research and Development, 1981.

Schmidt, F. L., Hunter, J. E., McKenzie, R. & Muldrow, T. The impact of valid selection procedures on workforce productivity. Journal of Applied Psychology, 1979, 64, 609-626.

Taylor, H. L. & Russell, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 1939, 23, 565-578.

Wallace, J. R. The Gideon Criterion: The effects of selection criteria on soldier capabilities and battle results. (Research Memorandum 82-1). Fort Sheridan, IL: U.S. Army Recruiting Command, January 1982.

Developing a Model of Soldier Effectiveness:

A Strategy and Preliminary Results *

Walter C. Borman
Personnel Decisions Research Institute

Stephan J. Motowidlo
College of Business Administration
The Pennsylvania State University

Lawrence M. Hanser
U. S. Army Research Institute for
the Behavioral and Social Sciences

Paper prepared for symposium:

Integrated Criterion Measurement for
Large Scale Computerized Selection and Classification

American Psycholcgical Association Convention
Anaheim, California

* Paper presented at the 91st Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

ABSTRACT

This paper introduces a model of individual effectiveness that extends beyond successful performance on specific job tasks and on directly job-related effectiveness dimensions. The model of soldier effectiveness suggested here contains elements of morale, along with organizational commitment and socialization. The notion is that these broad constructs represent important criterion behaviors that contribute to an individual soldier's "worth to the Army" and to his/her unit's organizational effectiveness. Fifteen dimensions springing from the model are named and defined.

The paper also presents preliminary results of behavioral analysis or BARS (Smith and Kendall, 1963) research to develop dimensions of soldier effectiveness using this comparatively inductive procedure. Fourteen Army officers in two workshops generated a total of 245 behavioral examples of soldier effectiveness in these early stages of the research project. Although by no means a formal test of the soldier effectiveness model, the content of the examples generated showed considerable similarity to many elements of the model. Exceptions were noted and discussed. Also discussed were certain advantages to taking a broader perspective in studying individual effectiveness, particularly in this kind of organization, as well as risks inherent in considering criterion elements that are not directly job-related.

Developing a Model of Soldier Effectiveness:

A Strategy and Preliminary Results

This paper focuses on an effort to build a model of soldier effectiveness. Model development is proceeding in research attempting to define U. S. Army first-tour enlisted performance. The performance definition work will in turn lead to development of criteria of effectiveness to be used in a large-scale test validation research program in the Army (Eaton and Shields, 1982).

First, the model development effort seeks to develop performance dimensions relevant to all soldiers, no matter what their specific job and duty assignment. Thus, one aspect of the research will focus on developing a set of generic performance constructs that reflects the portion of the performance criterion space in common across all or the vast majority of Army enlisted jobs.

A second focus is on identifying constructs related to other possible elements of soldier effectiveness. The notion here is that being a good soldier means much more than doing the job properly, that is, performing assigned tasks in a technically correct manner. A model of soldier effectiveness may properly include other elements that contribute importantly to a soldier's effectiveness and "overall worth to the Army." This paper explores these other elements and presents initial empirical results that help define in behavioral terms the domain of soldier effectiveness.

Regarding the model, we have made some preliminary hypotheses about constructs that might be considered under the broad soldier effectiveness domain. These constructs revolve around the areas of organizational commitment, organizational socialization, and morale.

The Preliminary Model

Organizational Commitment -- The concept of organizational commitment (Porter, Steers, Mowday, & Boulian, 1974; Steers, 1977) refers to the strength of a person's identification with and involvement in the organization. It incorporates three kinds of attitudinal and cognitive elements: acceptance and internalization of organizational values and goals, motivation to exert effort toward the accomplishment of organizational objectives, and firm intentions of staying in the organization. The concept transcends job involvement and motivation to perform the specific tasks that comprise the job. It connotes a sense of loyalty to the organization as a whole and a desire to fulfill more general role requirements that come with organizational membership. We argue that organizational commitment may reflect one aspect of this broad conception of soldier effectiveness.

Organizational Socialization -- "In its most general sense," say Van Maanen and Schein, (1979, p. 211), "organizational socialization is the process by which an individual acquires the social knowledge and skills necessary to assume an organizational role." Some part of this knowledge and skill is, of course, job-specific. For example, training programs designed to improve the effectiveness with which a person performs job-related tasks are part of the process of organizational socialization. But there are also many other knowledges and skills necessary for effective functioning as an organizational member that are not job-specific. When the socialization process is successful, a person will acquire not only job-related skills but also new patterns of behavior with subordinates, peers, and superiors in the organization; new attitudes, beliefs, and values in line with organizational norms; and new ways of using time not formally dedicated to performing job-related tasks.

Such individual changes are frequently crucial for assuring that the behaviors of different individual organization members will be smoothly coordinated toward accomplishing the organization's mission. As a result, soldier effectiveness might reasonably be regarded as partly a reflection of successful socialization; that is, people whose behavior and attitudes more closely coincide with Army norms might be regarded as more effective soldiers and considered of more worth to the Army.

Morale -- The concept of morale has traditionally been regarded as an extremely important element in military organizations. Munson (1921), a former brigadier general on the General Staff, writes:

That their mental state, their will to do, their cooperative effort, their morale--all of which are synonymous--bear a true relation to their output, productivity, and the success of their joint undertaking, is so obvious and has been proven so often as to require no supporting argument. (p.2)

The concept of military morale is multifaceted. It seems to involve feelings of determination to overcome obstacles, confidence about the likelihood of success, exaltation of ideals, optimism even in the face of severe adversity, courage, discipline, and group cohesiveness. (Motowidlo, Dowell, Hopp, Borman, Johnson, & Dunnette, 1976). Borman, Johnson, Motowidlo, and Dunnette (1975) report results of a study in the Army designed in part to identify behavioral dimensions of morale (see also Motowidlo & Borman, 1977). They found that the following dimensions efficiently describe behavioral expressions of morale among soldiers: community relations; teamwork and cooperation; reactions to adversity; superior-subordinate relations; performance and effort on the job; bearing, appearance, marching, and military courtesy; pride in unit, Army, and country; and use of time during off-dury hours. Because morale seems to figure so prominently as a determinant of unit effectiveness, behavioral dimensions like these may also in part represent important elements of individual soldier effectiveness.

In sum, we expect that the criterion domain of soldier effectiveness and worth to the Army is heavily saturated with elements of organizational commitment, successful socialization, and morale. Our preliminary hypotheses, then were that soldiers who show high levels of commitment to the Army, acceptance of Army norms, and morale are more effective soldiers in this broader sense and are also of more value to the Army.

These three broad constructs can be viewed in another way that leads to more specific hypotheses about this domain. From the combination of morale and commitment emerges a general category that can be labeled Determination. It is basically a motivational and affective category that reflects the spirit, strength of character, or "will-do" aspects of good soldiering. Morale and socialization spawn "Teamwork," behaviors that have to do with effective relationships with peers and the unit. Commitment and socialization give rise to "Allegiance." This taps into acceptance of Army norms with respect to authority, faithful adherence to orders, regulations, and the Army life-style, and being adjusted and socialized to the point of wanting to continue in the soldiering role and stay in the Army. Each general category of effectiveness subsumes five more specific dimensions.

Figure 1 shows how all of this fits together. As shown, the most abstract and broad construct, "Soldier Effectiveness," is defined according to somewhat narrower notions of "Morale," "Socialization," and "Commitment," which, with judicious mingling of conceptual elements, produce more concrete categories of "Determination," "Teamwork," and "Allegiance," which, finally, subsume more specific dimensions of soldier effectiveness. Table 1 contains 15 preliminary dimensions of soldier effectiveness.

Figure 1

A Preliminary Model of Soldier Effectiveness

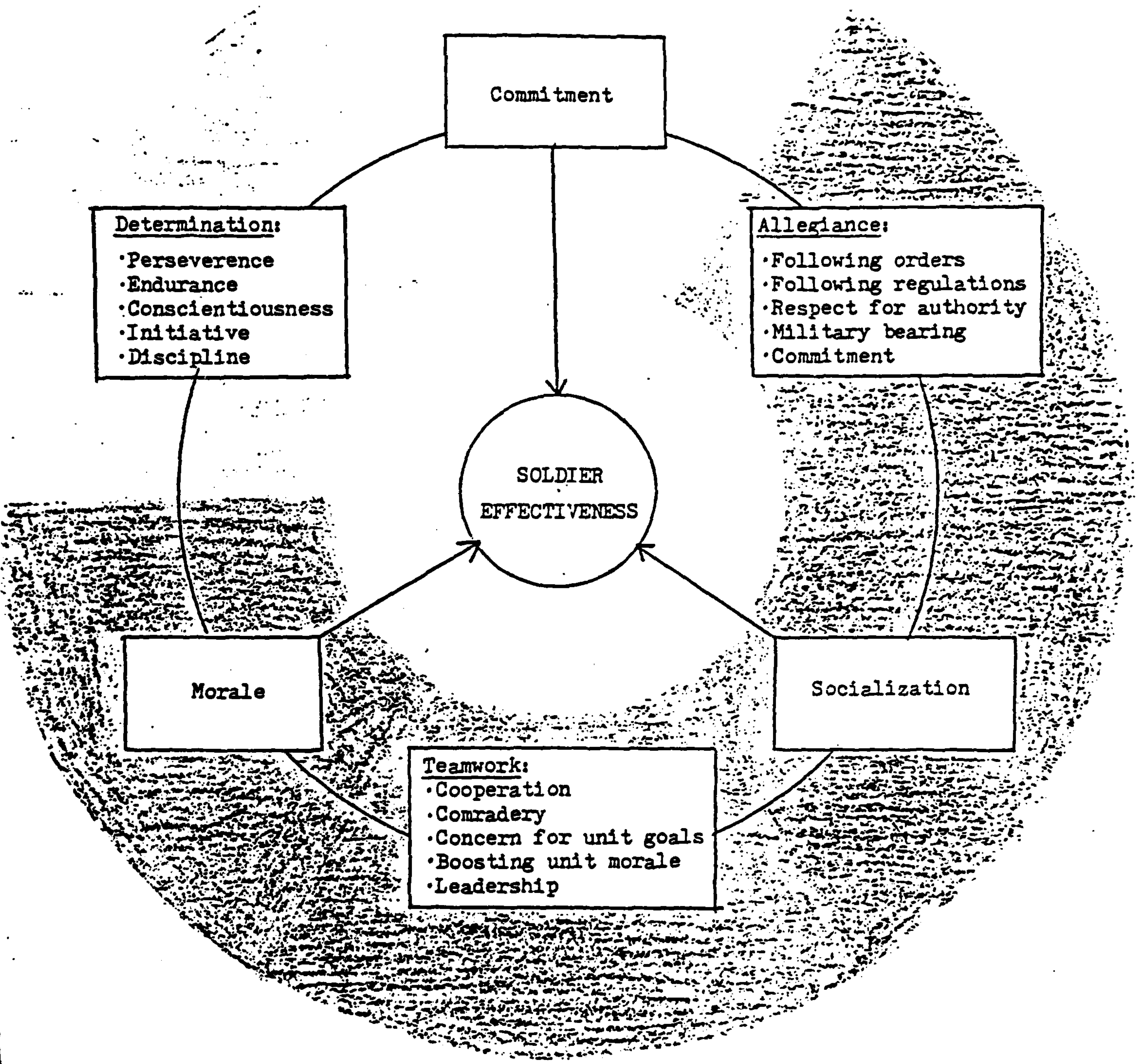


Table 1. Fifteen Preliminary Behavioral Dimensions Defining Overall Soldier Effectiveness

I. Allegiance

1. Following orders: responds willingly and eagerly to orders; carries out orders promptly and thoroughly; accepts direction from superiors without undue hesitation; versus responds half-heartedly to orders; carries out the letter but not the spirit of orders; refuses to obey orders.

2. Following regulations: complies with rules and regulations; conforms appropriately to standard procedures; tries to correct nonstandard conditions; versus frequently violates rules and regulations; ignores standard procedures when personally inconvenient; follows the letter but not the spirit of rules and regulations.

3. Respect for authority: defers appropriately to superiors' expertise and judgment; shows good military courtesy and respectful demeanor to superiors; speaks respectfully about superiors in conversations with others; versus habitually questions superiors' expertise and judgement; fails to salute properly or show military courtesy and respect in the presence of superiors; speaks disrespectfully about superiors in conversations with others.

4. Military bearing: grooms and dresses to maintain a crisp military appearance; stands, walks, and marches with an erect military posture; shows pride in the uniform and military insignia; versus grooms and dresses sloppily or without regard to military custom; stands, walks, and marches in a slouchy, casual, or careless manner; shows indifference toward the uniform and military insignia.

5. Commitment: successfully adjusts to military life; shows pride in being a soldier; wants to stay in the Army; versus fails to adjust to military life; shows indifference, dissatisfaction, or embarrassment about being a soldier; wants to leave the Army.

II. Teamwork

6. Cooperation: voluntarily pitches in when necessary to help other unit members with their job and mission assignments; willingly accepts personal inconvenience to aid other unit members with important problems; takes the trouble to listen and support other unit members with personal difficulties; versus pitches in only reluctantly when asked for job- or mission-related assistance; refuses to help other unit members with important problems if personally inconvenient; shows insensitivity and impatience with other unit members who have personal difficulties.

7. Comradery: is popular and well-liked by other unit members; forms close friendships with other unit members; spends off-duty time in group activities with other unit members; versus is unpopular or disliked by other unit members; frequently quarrels or fights with other unit members; remains aloof and spends off-duty time in solitary activities.

8. Concern for unit objectives: puts unit objectives before personal interests; makes personal sacrifices for the unit as a whole; works hard to meet unit objectives even when there is no personal gain; versus refuses to help meet unit objectives when they conflict with personal interests; shows more concern for personal interests than for the welfare of the unit; works for unit objectives only when there is personal gain.

9. Boosting unit morale: helps the unit stick together through hard times; encourages others to keep going when things seem bleak and hopeless; cheers others up when in unpleasant situations; versus shows no concern for unit solidarity; cynically criticizes others who refuse to give in for being foolish and unrealistic; constantly reminds others of the negative or unpleasant aspects of their situation.

10. Emergent leadership: shows good judgment in suggesting ideas for how others in the unit should proceed; persuades others to accept his/her ideas, opinions, and directions; others turn to this soldier for guidance and advice; versus never or rarely has good ideas for how others in the unit should proceed; presents opinions timidly and indecisively or is pushy and strident in rendering opinions, persuading/guiding others, etc.; others ignore this soldier's ideas, opinions, and directions.

III. Determination

11. Perseverence: struggles tenaciously to reach objectives even when the odds of success seem hopeless; sustains maximum effort over long periods of hard duty with unflagging stamina; comes back with redoubled effort when temporarily set back by failure; versus gives up on objectives that cannot be easily reached; tires out quickly and takes frequent rest breaks; easily discouraged by minor setbacks and frustrations.

12. Endurance: shrugs off severely uncomfortable or unpleasant conditions as though they were trivial; adapts and makes the best of hardship conditions without complaint; refuses to let troubles get him/her down; versus exaggerates the severity of minor discomfort and unpleasantness; constantly complains and grumbles about the lack of amenities; loses perspective and becomes demoralized by insignificant troubles.

13. Conscientiousness: spends extra time and effort to get the job done; consistently completes job and duty assignments promptly on or ahead of schedule; carries out assignments with thoroughness and careful attention to detail; versus refuses to take extra steps to make sure the job gets done; is frequently slow or late in completing assignments; works sloppily and ignores important details.

14. Initiative: volunteers for assignments; anticipates problems and takes action to prevent them; performs extra necessary tasks without explicit orders; versus refuses to volunteer for assignments; waits passively until difficulties occur and reacts only to the immediate problem; does only what explicitly ordered to do.

15. Discipline: devotes full concentration to the job at hand without yielding to the temptation of distractions; controls self-indulgent appetites and does not allow them to interfere with the performance of duty; keeps emotions in check and almost never loses his/her temper; versus easily distracted by opportunities to play, socialize, or pursue other leisure activities; lets too much eating, drinking, sleeping, or other self-indulgent appetites interfere with the performance of duty; fights or destroys property in uncontrolled emotional outbursts with little provocation.

METHOD

Behavioral Analysis to Define the Soldier Effectiveness Domain

Although the conceptual model showed promise as a way of depicting several important dimensions of soldier effectiveness, the intention was not to accept these dimensions at face value. Instead, we plan to use the inductive behavioral analysis method (Campbell, Dunnette, Arvey, & Hellervik, 1973; Smith & Kendall, 1963) to develop dimensions of soldier effectiveness, and then to revise the model based on what is learned in this analysis.

Briefly, behavioral analysis or behaviorally anchored rating scale (BARS) development involves getting persons knowledgeable about a job to generate behavioral examples of effective and ineffective performance in all aspects of the job. Several examples from the domain of interest here, soldier effectiveness, appear later in Table 2.

The many examples generated are content analyzed, and categories of performance are formed. Then, in "retranslation," persons familiar with the job evaluate each example, placing it in a category and rating the level of effectiveness reflected. Those examples that show good agreement regarding the effectiveness ratings and the categories judges sort them into are used to develop behavioral anchors of effectiveness serving as benchmarks defining the different effectiveness levels on each category.

At this time, we have conducted two BARS development workshops with a total of 14 experienced Army officers (Captains and Majors). In the workshops these officers generated 245 examples of first-tour soldier effectiveness, and we have performed a preliminary content analysis to explore possible dimensions emerging to define soldier effectiveness. Several other workshops will be conducted with officers and NCOs to ensure good coverage of the entire target domain, but these 245 examples provide some idea of what that domain will look like.

RESULTS

Dimensions Emerging from the Examples

Twenty-two relatively fine-grained and specific dimensions were derived from the content analysis. They appear in Table 2 along with a couple of representative examples characterizing each dimension.

Table 2. Dimensions Derived from Content Analysis of 245 Behavioral Examples of Soldier Effectiveness [a]

1. Promptness vs. Tardiness

- During a six month time period, this soldier was always on time for duty at the appropriate location.

- Although this soldier had been reminded about the SQT testing date and knew it required 100% attendance, he failed to show up on time. His whole unit had to schedule a make-up session.

2. Job Knowledge and Skill

- This soldier was very knowledgeable about the military movements required for a Change of Command parade. In his unit's preparation for the parade, on several occasions he provided help and advice to fellow unit members regarding parade movements, resulting in quicker and smoother preparation for the event.

- While on night patrol this soldier lost his way. Even though he had in his possession a compass and a map of the area, because of his inability to use them, he was unable to find his way back to his unit until daylight.

3. Personal Financial Management

- Although this soldier had been notified by the telephone company that his bill was overdue, he did not pay it. As a result, the telephone company contacted the CO, who then had to take time to straighten out the matter.

- Although this soldier had been counseled on his finances, he continued to purchase merchandise on credit that he could not afford. As a result, he was unable to pay the bills and became seriously indebted.

4. Stealing, Lying, Sociopathy

- When this soldier discovered that he was missing some tools he had signed out for, he promptly reported it to his platoon sgt. As a result, property account was maintained and the lost equipment was replaced.

- When this soldier damaged a one-of-a-kind piece of equipment that was needed for a field exercise, he neglected to tell his supervisor until just before the exercise started. Consequently, his group was not able to participate in the exercise.

---

a. Acronyms are contained in the Appendix

5. Physical Fitness

- This soldier was identified to go to PNCOC. Although initially overweight, this soldier pushed himself to meet Army weight standards prior to the course. As a result, he met standards a week ahead of time.

- Although counseled repeatedly and assigned to Remedial PT, this soldier refused to work on improving his physical condition. As a result, he was not able to meet the physical fitness standards and was barred from reenlistment.

6. Maintaining Clean and Neat Quarters/Environment

- When this soldier was walking through a neighboring battalion, he picked up some litter and went out of his way to deposit it in a waste can.

- When it was this soldier's turn for daily clean up, he swept only the easily accessible areas and did not bother to replace the plastic bags lining the trash cans.

7. Drug/Alcohol Abuse

- While participating in a German-American carnival, this soldier became drunk and refused to leave when asked to do so by the German bartender.

- This soldier was picked up off the base, while driving under the influence of alcohol. Previously, this soldier had been briefed about the use of drugs and alcohol, and on this occasion he knew he had drunk too much to drive.

8. Maintaining Own Equipment

- After a two-hour fire fight in a tactical exercise, the unit returned to base camp. Most of the unit members relaxed rather than preparing for the next fight, but this soldier cleaned his weapon without being told to do so.

- Because this soldier did not follow the proper procedures in cleaning his weapon, he broke its stock.

9. Attention to Detail

- Prior to a motor movement over mountainous terrain, this soldier was directed to check all APCs for proper torque on drive sprocket bolts. He checked them carefully and reported that the bolts were in good shape. The vehicles did make the trip successfully without incident.

- While raising a vehicle to remove a wheel, this soldier neglected to check the emergency brake and did not use check blocks. When the jack slipped, the vehicle rolled into another vehicle, causing minor damage to both.

10. Following Standard Operating Procedures on Tasks

- While on guard duty, the soldier noticed someone in an unauthorized area. He properly detained the individual and reported the incident to the commander of his relief.

- The unit reenlistment NCO failed to maintain reenlistment records in accordance with the proper REGS and Co. SOP. The problem was noted prior to an inspection but meant that the ISG had to work with this soldier to prepare for the inspection.

11. Initiative/Volunteering

- After fighting a forest fire on a military reservation for fully 48 hours, this soldier volunteered to help out a relief unit, continuing to fight the blaze for another six hours.

- When asked to stay late with another soldier to prepare the NBC room for inspection, this soldier decided to leave since he knew he was not required to stay.

12. Perseverance

- When this soldier was tasked with digging a fighting position for a training exercise, he continued to work on it even after he had passed the OIC inspection. His position was so well constructed that it was chosen to be a permanent fixture in the installation.

- When this soldier was assigned to guard a bivouac area at night on an FTX, he fell asleep at one of the training stations even though he knew he was supposed to be walking the post.

13. Effort to Improve Soldiering and Job Skills

- This soldier couldn't understand the sample problems during land navigation class. After class he returned to the learning center and continued working with the programmable learner until he understood the problems. As a result, during the FTX, his patrol was able to make maximum points for finding each of the required locations.

- The soldier studied and practiced performing critical tasks during off-duty hours and as a result attained a maximum score on his SQT.

14. Military Appearance

- At a guard mount inspection, this soldier's boots were highly polished, his hair neatly trimmed, and his uniform was neat and clean.

- This soldier appeared at the first formation unshaven, with unshined boots, and wearing a wrinkled uniform without a belt. His personal appearance and uniform failed to meet Army standards.

15. Accepting Orders from Superiors

- On several occasions when directed to perform a duty, this soldier questioned his NCO as to why he had to perform it. His continual questioning of orders caused delays and generated ill feelings within his unit.

- When this soldier's unit was preparing to deploy to California for a training exercise, he told his squad leader, company commander, and other soldiers that he would not go because it was all worthless.

16. Military Courtesy

- The soldier was in a group sitting around a tree when a senior officer walked toward them. He called the group to attention and saluted the officer.

- On several occasions when in the presence of officers and NCOs, this soldier did not rise, use a respectful tone of voice or respond with "Sir/Sgt." As a result, this soldier was cited for disrespect towards his superiors.

17. Following Regulations

- After duty hours, while on duty as the Company CQ, this soldier received a Red Cross emergency message. He passed the message only to the soldier that it involved, and did not attempt to notify anyone in the chain-of-command.

- During an FTS in Germany this soldier took his vehicle downtown although he knew it was against regulations.

18. Leadership: Taking Initiative to Lead Others; Taking Charge When Placed in Leader Position

- This squad leader made a point of being well prepared for FTX. As a result when the Brigade Commander gave a surprise inspection during FTX, he was the only leader who was able to brief the commander on the whereabouts of his personnel, their responsibilities, individual and squad sectors of fire, etc.

- A soldier was told that he was to be in charge of a PT formation. He refused to perform the assigned mission and instead reported to sick call. At sick call it was confirmed that he was only feigning illness.

19. Leadership: Motivating Others to Push On vs. Encouraging Them to Goof Off

- This soldier's squad was anticipating working long hours for several days to get ready for an inspection. Some of the squad members began complaining, at which time the soldier barked out, "Hey! Are we in this together or not?" The complaining died down and the squad did well in the inspection.

- When this soldier's battalion was left on post support duty, he constantly complained about it and advised others not to finish a detail ahead of schedule because there would just be more work.

20. Leadership: Correcting Performance of Others

- An SP4 was standing outside his company without his hat on. This soldier, also an SP4, approached him and told him firmly but tactfully to return to his room and get his hat.

- When an SP4 in the back of the formation began making sarcastic and sexist remarks about a female squad leader, this soldier quickly took him aside and told him to be quiet. As a result the SP4 ceased the disruption and later apologized to the squad leader.

21. Leadership: Instructing Others

- Before a Map Reading class, this soldier gave several fellow squad members some refresher training on terrain features. As a result these soldiers were more interested in the class and their performance improved.

- When tasked with learning Morse code, this soldier built a training device so that soldiers could send code to one another in the unit bay. As a result everyone learned the codes more quickly and the unit's overall performance in code training improved.

22. Displaying Concern for Individual Others and the Unit

- A unit member suffered heat stroke and was being medivaced to the hospital. The soldier, accompanying this man, noticed that his body temperature was remaining very high. He removed the man's clothes and drenched him with water, lowering his temperature sufficiently to avoid brain damage and possible death.

- During a water safety class, this soldier bragged about his own swimming skills and laughingly chided another soldier who could not swim.

Figure 2

Comparison of the Model's Dimensions and the Empirically Derived Dimensions

Model Dimensions \ Empirical Dimensions	1. Promptness	2. Job Skill	3. Finances	4. Stealing/Lying	5. Physical Fitness	6. Clean Quarters	7. Drug Abuse	8. Equipment	9. Detail	10. Following SOP	11. Initiative	12. Perseverence	13. Improve Skills	14. Appearance	15. Accept Orders	16. Courtesy	17. Follow Regs.	18. Leadership I	19. Leadership II	20. Leadership III	21. Leadership IV	22. Concern for Others
1. Following Orders						0			0						■							
2. Following Regulations	■							X	0					0			■	X		X		
3. Respect Authority														0		■	X					
4. Military Bearing														■								
5. Commitment																						
6. Cooperation											X										■	■
7. Comradery																						
8. Concern Unit Coals		0							0													0
9. Boosting Unit Morale																			■			0
10. Leadership										■								X				
11. Perseverence											0	■						0				
12. Endurance								0			0	0										
13. Conscientiousness						■		X	■		0	0	■					X				
14. Initiative		0									X							X		0		
15. Discipline							■															
16. Other		■	■	■	■						0									0		

■ = Very high overlap (>75% of behavioral examples falling in empirical dimension fall in model dimension)
X = Intermediate amount of overlap (15-74% of examples falling in empirical dimensions fall in model dimensions)
0 = Small overlap (at least 1 example but <15% of examples falling in empirical dimensions fall in model dimensions)
Blank = No overlap.

Comparison of Model Dimensions and the Empirically Derived Dimensions

To obtain an initial idea about how the empirically derived dimensions might fit into the 15-dimension framework provided by the model of soldier effectiveness, the first author sorted each behavioral example into one of the model's dimensions. Then, behavioral incident membership in dimensions within the two systems was cross-referenced to provide a rough comparison between the two dimensional systems. See Figure 2 for a depiction of this.

Results of this cross-referencing show first that two dimensions in the model of soldier effectivenss are not reflected in any examples. Commitment and Comradery had no incidents sorted into them. Second, Conscientiousness and Following Regulations are probably too broad, with incidents from 7 of the 22 empirical dimensions appearing in each of those categories of the model.

Third, four of the empirical dimensions do not have any representation in the model. Job Knowledge/Skill, Financial Management, Stealing/Lying, and Physical Fitness are not reflected in the model's dimensions. Fourth, there are some near one-to-one matches between the two dimensional systems. Military Bearing and Military Appearance, Boosting Unit Morale and the second Leadership dimension in the empirically derived dimensions, Perseverance (in both systems), and Discipline and Drug/Alcohol Abuse provide good matches. In the last case, however, Discipline is defined much more broadly than the content represented in the Drug Abuse examples.

Finally, and this is perhaps the most salient result, there seems to be considerable overlap between the dimension content in the two systems, but often the configuration of that content differs. Essentially, elements of the dimensions in the two systems are put together differently.

Of course, how the behavioral content is configured to form dimensions is largely arbitrary. The most important objective in developing dimensions is to achieve the purposes of the project, and this overriding concern will guide future efforts to integrate empirical information with the theoretical model. Dimensions will be developed and defined to reflect in a comprehensive and at the same time, efficient manner the domain of soldier effectiveness. Dimensions will be derived to provide raters using rating scales based on the dimensions with an easy-to-understand, highly face valid rating format that reflects accurately the behavioral requirements of this domain.

DISCUSSION

The model described here was designed to portray elements of soldier effectiveness in a context broader than successful performance on job-related tasks. It is an effort to tap elements of criterion behaviors that are important for organizational effectiveness, but that are not specifically a part of the individual's job. The model presumes that soldier effectiveness involves morale, commitment, and socialization and suggests more specific dimensions that underlie effectiveness in the soldiering role regardless of what the individual's particular job might be.

This approach has the advantage of forcing a broad perspective on the criterion domain. It points out potentially important elements of individual effectiveness that might be overlooked by purely inductive approaches to job and task analysis. For this reason, we believe the model is useful for guiding efforts to impose structure upon the bewildering richness of what "soldier effectiveness" might mean in the Army.

Of course, the model might not be correct. Some facets of effectiveness which it suggests might not actually be important elements of success as a soldier, and there could well be other important facets of effectiveness that it does not address. Consequently, we emphasize that the model is useful primarily as an initial guiding device and that subsequent empirical work will refine the model, trim away irrelevant criterion elements, and add new ones found to be important components of soldier effectiveness.

The preliminary set of behavioral examples reported in this paper indicates that most of the elements of effectiveness described by the model are in fact recognized by officers as important for soldier effectiveness. However, Commitment and Comradery, for example, were not represented in the incidents gathered so far. Possibly, these dimensions do not belong in the model.

Another possibility is that these dimensions are important even though they are not frequently reflected in behavioral incidents. There might conceivably be something about those facets of the effectiveness domain that make it difficult for people to recall examples of behavior unless instructed to think specifically about those particular facets.

Although there are potential advantages to broadening the scope of the performance domain to include elements of effectiveness that are not job-specific, this approach also carries an inherent risk. As we move from the relative concreteness and immediacy of effectiveness in specific job-related tasks, the importance of less job-related elements such as Respect for Authority and Military Bearing for organizational effectiveness becomes less obvious and direct.

Even though officers frequently cite examples of behavior that reflect these elements in ways that indicate they assume they are important for soldier effectiveness, it is not obvious that soldiers who are exceptionally good or poor in those areas necessarily contribute to or detract from the success of the Army's overall mission. It is much more obvious that soldiers who perform their jobs well or poorly contribute to or detract from organizational effectiveness. On the other hand, although such dimensions might seem somewhat removed from effective contribution to the Army's mission, we believe they may help shed light on patterns of behavior that have important implications for Army organizational effectiveness.

References

Borman, W. C., Johnson, P. D., Motowidlo, S. J., & Dunnette, M. D. Measuring motivation, morale and job satisfaction in Army careers. Minneapolis: Personnel Decisions, Inc., 1975.

Campbell, J. P., Dunnette, M. D., Arvey, R., & Hellervik, L. The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.

Eaton, N. K., & Shields, J. L. U. S. Army Soldier Selection, Classification, and Utilization Research Program. Paper presented to 90th Annual Convention of the American Psychological Association, August 1982.

Motowidlo, S. J., & Borman, W. C. Behaviorally anchored scales for measuring morale in military units. Journal of Applied Psychology, 1977, 62, 177-183.

Motowidlo, S. J., Dowell, B. E., Hopp, M. A., Borman, W. C., Johnson, P. D., & Dunnette, M. D. Motivation, satisfaction, and morale in Army careers: A review of theory and measurement. (ARI Tech. Rep. TR-76-A7). Arlington, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, December 1976 (NTIS No. AD-A036390)

Munson, E. L. The management of men. New York: Holt, 1921.

Porter, L. W., Steers, R. M., Mowday, R. T., & Boulian, P. V. Organizational commitment, job satisfaction, and turnover among psychiatric technicians. Journal of Applied Psychology, 1974, 59, 603-609.

Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for ratings scales. Journal of Applied Psychology, 1963, 47, 149-155.

Steers, R. M. Antecedents and outcomes of organizational commitment. Administrative Science Quarterly, 1977, 22, 46-56.

VanMaanen, J., & Schein, E. H. Toward a theory of organizational socialization. In B. M. Staw (Ed.), Research in organizational behaviors (Volume I). Greenwhich, CT: JAI Press, 1979.

Appendix

Acronyms used in Table 2.

APC - Armored Personnel Carrier
CO - Commanding Officer
CQ - Charge of Quarters
FTX - Field training exercise
NCO - Non-Commissioned Officer
OIC - Officer in Charge
PNCOC - Primary Non-Commissioned Officer Course
PT - Physical Training
REGs - Regulations
SOP - Standard Operating Procedure
SP4 - Specialist 4th Class
SQT - Skill Qualification Test

Dusting Off Old Data:

Encounters with Archival Records

Lawrence M. Hanser and Frances C. Grafton

U.S. ARMY RESEARCH INSTITUTE
FOR THE
BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, VA 22333

August 1983

* Paper presented at the 91st Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

This paper could have been titled "One researcher's dream, another researcher's nightmare." You will recognize the intent of that remark as I proceed. This paper really addresses three separate, but related issues. The first of these has to do with dredging up old data for analysis or re-analysis, the second has to do with analyzing large data sets, and the third is related to the applicability, or inapplicability, of textbook solutions to real world problems. For us these issues are so bound together that our discussion of one inevitably bleeds into a discussion of the others. We don't pretend to provide any new or revolutionary ways for dealing with these problems. Indeed, many of the comments we will make will be familiar to you. For example, how many times have you told your students to clearly label each computer printout, yet failed to do so yourself? Hundreds of such little bits of wisdom exist. I can assure you that we have failed to apply most of them, and those that we did apply were often the result of dumb luck. Our intent in this paper is to revive what you already know, but don't apply, and perhaps offer some new suggestions along the way.

Our experience stems from many collective years of work with very large data sets. Let me provide you with the context from which our present discussion flows. A little over a year ago ARI undertook a large validation project -- the project you've been hearing about throughout this symposium. During 1982, prior to the awarding of the contract for this project, Frances and I were tasked with providing the Defense Advisory Committee on Military Personnel Testing with information about the validity of our selection tests. They asked for detailed information about the criteria against which ASVAB had been validated in the past, as well as for information concerning the validity of ASVAB for use with minority groups.

We'd like to provide you with a few statistics to give you some feeling for the size of the problem and the context within which the problem had to be solved. Every month around 70,000 individuals apply to enter one of the US Armed Forces. For each of those individuals, a Military Enlistment Processing Station puts together a 690 byte record. Some quick arithmetic will tell you that the size of the monthly applicant data set is ... well, very big. (Notice the precision of numbers. We throw out more cases than most psychologists begin with!). Remember, I said 70,000 people apply. Of these, about 10,000 to 11,000 actually enlist in the Army. The rest either enlist in one of the other Services or are not enlisted at all. The total number of enlistees during 1981, for the Army alone, was approximately 130,000. While the large sample sizes that we deal with do create some unique problems, for the most part the impact is simply to magnify the same problems that other personnel researchers face.

You may think, boy they must have some kind of fancy automatic system to handle all those data. Well..., yes and no. While we are moving in that direction, we have not arrived there yet. The entrance battery, ASVAB, is machine scored but the scores are often manually keyed rather than transferred by machine into a file. Other entry level information which is collected at the recruiting and enlistment processing stations is also manually entered into the system. As recently as last year some scores were entered twice -- a set of data was keyed into a local system, printed from the local system, and keyed from the printout into another system (note the possibility for error).

Back to some statistics. Each individual who enlists in the Army is sent to school for training in a Military Occupational Specialty (MOS). There are over 300 MOS in the Army, and first time enlistees can qualify to enter most of them. While there are many MOS, there are far fewer schools, so that it's not too difficult to track an individual from enlistment to a school. Training would of course be the first most logical place to collect criterion data. While in training, soldiers are located in relatively few locations in the continental U.S., and are generally clustered by MOS, so that data collection would be relatively convenient and cost effective. In fact, some of the scores earned by soldiers during training are entered into the Training and Doctrine Command's Educational System (TREDS), which is stored in Washington, D.C. Other training scores are sometimes stored at the individual schools, and some are not stored at all. One problem with measures of success in training as criteria is that the Army has moved in the direction of mastery learning so that the training scores which are routinely collected by the schools typically display insufficient variance for our purposes. Project A research is partly directed at solving this problem.

After the completion of training, tracking individuals is a bit more complex because assignment can be to one of hundreds of military posts worldwide. Fortunately the Army does this quite well; we know where most everybody is. Unfortunately, knowing where everybody is and collecting criterion measures on them are two quite different problems. An MOS may have a large number of incumbents worldwide but have very few in any given location, making the cost of a special data collection effort prohibitive. Like most organizations, the bottom line requires that we take the prudent course of action and use the best available criterion that we have, scores on Skill Qualification Tests.

Tracking the location of an individual is one thing; tracking the location of data available about an individual and gaining access to it is quite another. A file which is available to us from the Military Enlistment Processing Command (MEPCOM) in Northern Illinois, contains ASVAB test scores and other entry level information. The official machine readable military personnel file is maintained by the Military Personnel Center (MILPERCEN) in Alexandria, Virginia. Microfiche copies of the individual soldier's hand-carried paper personnel files are kept in Indianapolis. However, much personnel data that would be of interest to us from the paper files never makes its way into machine readable form. Scores earned on Skill Qualification Tests, on-the-job measures of performance which have been used as criteria in the past, are on file at Ft. Eustis, Virginia. If you imagine some massive computer network which allows for file merging or rapid transfer of information from one file to another -- don't even think it. These installations, for the most part, don't even use the same brand of computer.

It is really a bit unfair of us to paint this kind of picture. We view the world as researchers. Predictor and criterion scores have to be matched for each individual before we can use them, and as you well know, we often find it useful to have other descriptive information, such as race and sex, along with an individual's test scores. The key is that we view the world as researchers. These databases were not developed, and are not maintained to serve the whim of personnel research. They were developed to serve the everyday, ongoing operational needs of the Army.

I've strayed away from the topic a bit -- these kind of stories are fun to tell, but let me return to the point, the validity research we were tasked to do by the Defense Advisory Committee on Military Personnel Testing. The first problem that we had to deal with was how to choose a sample for the analyses. As with most real world problems, we had been given a very short time period in which to complete our report. It was clear from the start that we would have to narrow the number of MOS with which we would work. We began by deciding that we would have to work with data which had already been collected. While ASVAB scores are available on an individual at enlistment, the criteria we intended to use, Skill Qualification Tests, are not administered until approximately 18 months after enlistment. This is of course a common problem in validity research -- criterion data not being immediately available. Since our research was begun in early Summer 1982, the time lag between enlistment and the availability of a criterion measure meant that our enlistment sample had to have entered the service by January 1981. Recall also that this had to be done in the real world. While we enlist approximately 130,000 in a year, it doesn't occur in one day, but over the period of a year. Thus, in order to span just a single year's enlistment population, our sample would had to have begun entering the Army in January 1980. But that would have been too simple. As fate would have it, new forms of ASVAB were made operational on October 1, 1980. Actually many other similar variables entered into our choice of the sample years. Let's just say that when the dust settled, we had chosen to use enlistees from 1976 through 1978 who had criterion scores from 1978 and 1979.

We were still faced with the problem of narrowing the number of MOS to be included -- 300+ were just too many. In many MOS the number of entrants in a given year can be quite small if not zero. After examining the available criterion data, we chose to analyze 6 MOS. Not the least of reasons for our choices was sample size. The number of individuals tested from one MOS to another varies greatly. In 1978 the largest group tested was infantry with 13,524, and the smallest was Lance missile repairman with 1. Similar numbers were tested in 1979. Since the committee had specifically asked that we look at minority groups, we also considered this when choosing the MOS. This tended to be more of a problem with regard to women than Blacks, as might have been expected. We also tried to include a range of types of specialties. We ended up including infantry, drivers, cooks, radiotelephone operators, and military police. Even though we had taken available sample sizes into account, when we split our samples into subgroups, our sample sizes became quite small in some cases.

I previously noted that every applicant is tested with ASVAB, the predictor battery, prior to enlistment, and the data recorded by MEPCOM. MEPCOM generates the data file that contains ASVAB scores, but they do not provide archival storage. The ASVAB and other identifying information for our sample was obtained from the Defense Manpower Data Center (DMDC). During the time period in which our sample enlisted, three parallel forms of ASVAB were used: 5, 6, and 7. Form 5 is used exclusively in the Department of Defense high school testing program. We were concerned only with forms 6 and 7. ASVAB 6/7 consisted of 16 subtests which were combined into 10 Aptitude Area Composites, 9 of which were used as qualifying scores for entrance into different Army training programs. The SQT dat[illegible]re obtained from the SQT Directorate at Ft. Eustis.

We also decided that we would need to look at copies of the actual SQT test booklets to provi[illegible] an adequate description of the criteria. This is where using old data really caused some problems. We discovered that the only remaining copies of the SQT test booklets were stored at Fort Monroe in Virginia. Further, to our dismay, multiple copies had not been kept, so that we were sent test booklets only after promising that we would return them promptly. In some cases we were provided with what proved to be the sole remaining copy of a given test. I can assure you that these were not easy to liberate for our use.

Now that you have a broad idea of the context in which our research had to be done, we'd like to talk more specifically about the difficulties we encountered and how we dealt with them. We have also developed some rules to live by when analyzing data. These rules fall into two major categories, analysis and housekeeping.

Our first rule is an analysis rule:

CHECK EVERY DATA FIELD THAT YOU INTEND TO USE FOR MISSING OR OUT OF RANGE VALUES.

The data tape which we received from DMDC with the ASVAB scores on it is a good case in point. ASVAB composite scores are derived from subtest scores using formulas and tables. When we checked the tape for subtest scores, we found that about 10% of our cases were missing four key subtest scores, and about 5% were out of range. As near as we could tell, composite scores had been coded where subtest scores should have been recorded for this 5%.

This brings us to rule 2 for analysis:

IF YOU HAVE SUFFICIENT CASES, THROW OUT SUSPECT CASES RATHER THAN SUBSTITUTING MISSING VALUE CODES OR ESTIMATING SCORES.

Our reasoning is that if there is an error in one part of the record, there may be errors in another part of the record. We knew that all our cases should have taken 16 subtests so that missing subtest scores either meant miskeying or that the individual had taken another form of the test.

Rule 3 for analysis is:

RECALCULATE ALL AGGREGATE OR COMPOSITE SCORES FROM RAW DATA IF AVAILABLE ON THE TAPE AND COMPARE THESE TO PRECALCULATED SCORES.

The corollary to analysis rule 3 is:

IF YOUR RECALCULATED SCORES AREN'T THE SAME AS THE PRECALCULATED RECORDED SCORES, PROCEED NO FURTHER UNTIL YOU KNOW WHY.

After we received the SQT tapes, we calculated total scores by adding the number of correct answers and taking the percent of total possible points. When we compared these to the precalculated scores which were on the tapes as we received them we found that they didn't match. We called an expert from the SQT Directorate and discovered that total scores were calculated differently. In some cases written items were weighted differently from hands-on items. We also learned that SQT for different MOS often used different scoring strategies. As the result of this exercise we present analysis rule 4:

NEVER ASSUME THAT A TAPE AND RECORD LAYOUT WITH VARIABLE DESCRIPTIONS WILL BE SUFFICIENT FOR UNDERSTANDING THE DATA.

The corollary to this rule is:

ALWAYS IDENTIFY AN EXPERT SOURCE WHO WORKS FOR THE ORGANIZATION WHICH PROVIDES THE DATA. (We call this the Rosie Parks Corollary.)

By this time we had done enough with the data to have violated a number of housekeeping rules. The most important are:

SAVE EVERY RUNSTREAM YOU WRITE -- YOU WILL NEED IT AGAIN.

CLEARLY LABEL, DATE, AND FILE EVERY PRINTOUT.

USE DESCRIPTIVE NAMES WHEN SAVING DATASETS.

THROW OUT PRINTOUTS DETERMINED TO BE IN ERROR.

For the most part these rules are self-explanatory. We especially like to violate the one about labelling printouts, and on occassion have caused ourselves great distress by violating the one about throwing out printouts determined to be in error. We're sure that you also will have favorites among this list.

Our 5th analysis rule is:

AFTER YOU HAVE RECALCULATED AGGREGATE AND COMPOSITE SCORES AS WELL AS ANY OTHER NEW VARIABLES, APPLY RULE 1.

We have been known to have an error in the format statement and as a result slip a decimal place. Just because you calculated the scores yourself doesn't mean you didn't make a mistake.

Our 6th analysis rule is:

WHENEVER YOU WRITE A PROGRAM, NO MATTER HOW SIMPLE MINDED, ALWAYS RUN IT FIRST WITH TEST DATA HAVING KNOWN ANSWERS.

After we had calculated our corrected regressions, a question as raised with regard to their correctness. As a result we ran some test data through our program to insure it was doing what we intended, but until we had done so, there was this nagging doubt that we might have made an error. The test runs erased that doubt.

Our 7th and last analysis rule is one of our favorites:

AS YOU COMPLETE DATA MANAGEMENT TASKS OR ANALYSES, KEEP DETAILED NOTES.

If it has not happened to you yet, if you don't follow this rule, one day it will. You will reach a point in your analysis work where you begin to doubt the accuracy of everything you have done up to that point. You will have the desire to look through all your old printouts and begin all over again -- perhaps to the point of going back to raw unkeyed data. A colleague of ours at ARI coined a term for this. We call it 'data vertigo.'

As we're sure you know, the experiences which we've related so far, and the rules which govern them are for the most part common sense. We did experience other, more difficult, problems. This last set of problems challenged the scientist in us rather than the clerk. These were the data analysis and psychometric issues we faced. We'd like to relate these problems to you and describe briefly how we dealt with them.

The data that we worked with had come from a selected sample. Each individual in our sample had first joined the Army and then been allowed into a training program on the basis of minimum scores on ASVAB. In every case there were at least two scores on which an individual had been selected. At the same time, ASVAB subtests are highly correlated, so that selection on one subtest or composite results in restriction of range on other subtests. To make matters worse, our restriction of range occurs not only at the low end of scores, but at the high end as well. We have restriction of range at both ends of the scale. The only thing that we could be reasonably certain of was that our sample correlations were underestimates of population values.

To provide estimates of the population correlations, we applied a multivariate correction formula (Gulliksen, 1950; Lord & Novick, 1968). At first this appeared to be relatively straightforward; we knew that our estimates would be questionable, but could live with that. Further, this was the formula that had traditionally been applied at ARI. Part of what we were looking at was the validity of our composites for various minority groups. Of course we had our uncorrected correlations to compare, but we had little reason to believe that restriction of range was equally or randomly distributed across population subgroups. This meant that differences between uncorrected correlations might not reflect true differences in the validity of ASVAB for our subgroups. As fate would have it, we only had a single stratified target matrix to use in our correction formula. We did not have target matrices for each of our subgroups. Fully aware of these problems, but pressed for time, we used the stratified population correlation matrix in our correction formula and applied the same population target matrix to each of our subgroups.

We had also been asked to provide regression lines for each of the various minority subgroups. As you know from the theory of correcting for restriction in range, the regression line which is calculated in the sample is theoretically the same as the regression line in the unrestricted or applicant population. This is *necessarily* true in a simple bivariate case or a multivariate case where all variables enter into the regression. I can assure you that it is not true for multivariate correction when the regression of only a subset of variables is calculated. Our multivariate correction was for selection on 16 variables, but our regression lines were calculated for a simple linear unit-weight composite of 3-5 variables. When a colleague of ours reviewed an early draft of the paper which reported these regression lines, he called to tell us that we had obviously made a mistake. We spent a fair amount of time re-analyzing our data, and when we arrived at the same results, another lengthy period of time trying to understand what happened. As of this date, we do not yet have a definitive answer as to which regression line is more accurate, but we know that the lines are different. We presented both sets in our paper (Hanser & Grafton, 1983).

One important question which we attempted to answer had to do with differences in validity between males and females, and blacks and whites. If you remember that vague paragraph in your statistics book that discussed the distinction between statistical and practical significance, you can begin to appreciate some of our problem. We had sample sizes so large that even extremely small differences would have been significant. We chose not to report significance tests in our initial paper because we did not feel that they could be interpreted appropriately. We are currently struggling with what practical significance means to us. Our plans are to determine where, in our data, statistical significance exists, and to further examine those points for practical significance. We plan to consider standard errors, differences in predicted scores at the cutoff point, expectancy tables, and the like. In many ways the Eaton, Wing, and Mitchell paper (Note 1) explores this same region -- an attempt to determine the meaning of practical significance.

Unfortunately, there are no simple rules which can be applied to the problems mentioned above -- only creative minds working toward practical solutions. We submit that this is why personnel psychology is a science rather than a technology. There is one final law of nature which we have discovered in our work. We're sure you have already, or will in the near future also discover it. We are certain that it was first discovered by the author of the world's second statistics textbook and subsequently rediscovered by every author of a statistics book ever since:

In attempting to answer applied questions:

1) either the data you have are not consistent with the examples provided in any statistics book, or

2) the analysis you want to do is neither available in raw score form in any statistics book, nor is it to be found as a standard procedure in any statistical computer package.

Reference Notes

1. Eaton, N. K., Wing, H., & Mitchell, K. (1983, August). Putting the "Dollars" into Utility Analysis. Paper presented at the meeting of the American Psychological Association, Anaheim, CA.

References

Gulliksen, H. (1950). Theory of Mental Tests. New York: John Wiley & Sons, Inc.

Hanser, L. M., & Grafton, F. C. (1983). Predicting Job Proficiency in the Army: Race, Sex, and Education. In press. Washington, D.C.: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley Publishing Company.

X. PREDICTOR SELECTION AND DEVELOPMENT

Norman G. Peterson (PDRI)
Hilda Wing (ARI)

General Purpose

The general purpose of this activity is to identify an efficient and effective set of initial or preinduction predictors of soldier performance. By efficient, we mean that time and money to be expended on operational administration of the predictors are kept as low as possible, and by effective, we mean that the predictors forecast as accurately as possible the degree of success to be expected of recruits in various aspects of soldier performance, e.g., overall adaptation to the Army, completion of training, performance in specific MOS, and reenlistment.

There are two different but related aspects to this general purpose. First, we will evaluate the effectiveness of the present set of initial predictors used by the Army contrasted with a more comprehensive array of criteria than has been used previously. We will identify and develop new predictors that are most likely to be effective and efficient additions to the present set of predictors. The validity or effectiveness of these new predictors will be investigated in the same way as the validity of the present set of predictors. The evaluation of the efficiency of newly developed predictors will require analysis of the improvement in prediction of soldier performance gained by use of the new predictors over that obtained by the sole use of the present set of initial predictors.

Summary of Overall Procedure

There are 15 procedural steps or subtasks that comprise predictor selection and development. The subtasks are summarized briefly below. Subtasks 1, 2, and 3 have already been completed and the outcomes will be discussed in the next section.

(1) Literature search and planning. Civilian and military research about the relative "success" of predictors for purposes of personnel selection and classification was reviewed and summarized.

(2) Selection of preliminary battery and preparation for administration to FY83/84 longitudinal sample. A set of "off-the-shelf" predictors that comprehensively and efficiently cover the predictor space was identified by Task 2 staff and reviewed by ARI. After selection, the predictors were approved and administration procedures have been prepared.

(3) Administration of preliminary battery to FY83/84 longitudinal sample. The preliminary battery will be administered to a sample of 2,100-4,600 soldiers in training for each of four MOS: 05C, 19E/K, 63B, and 71L. On-site administrators have been trained by Task 2 staff and the administration process will be monitored by Task 2 staff.

(4) Technical review of predictor constructs and measures. Experts will be used to make formal judgments about the usefulness of predictor constructs and measures for predicting soldier performance. Analyses of these judgments will identify a set of predictors judged to be "best bets" in terms of validity and efficiency, i.e., minimal overlap between predictors.

(5) Cost administrative/practicality review. A panel of Army personnel knowledgeable about the field operation of recruitment, selection, and classification will be given information about the administrative procedures, costs, item types, etc., of the predictors surviving the technical review. They will make judgments about administrative feasibility, possible operations problems, etc.

(6) Initial development of predictors for the Trial Battery (new predictors). Predictors chosen for development on the basis of Subtasks 4 and 5 will be designed and small-scale tryouts will be held.

(7) Pilot tests of Trial Battery (new predictors) in the field. The new predictors will be administered to a sample of soldiers from the FY81/82 cohort and a sample of applicants. Data will be collected to allow the investigation of practice effects, fakeability, motivational effects, and stability of measures.

(8) Analysis of Trial Battery. The pilot test data will be analyzed to investigate administrative problems, applicant acceptance, psychometric properties, fakeability and practice effects, and covariances of new predictors, current preinduction predictors, and any criterion information available for the soldiers in the sample.

(9) Analyze Preliminary Battery: FY83/84 cohort school and preliminary battery data. The covariance of the predictors in the Preliminary Battery with current preinduction predictors and training success will be analyzed. Also, analyses of differences in MOS scores on constructs measured by the Preliminary Battery will be analyzed.

(10) Prepare revised Trial Battery for FY83/84 cohort predictor/performance data collection. Information from Subtasks 8 and 9 will be integrated and plans formulated for revising the new predictors. After review and approval by ARI, revisions of the new predictors will give rise to the revised Trial Battery. Training will be provided to Project A staff responsible for administering the battery.

(11) Monitor/assist administration of Trial Battery to FY83/84 cohort. Although the major burden of trial battery administration will be borne by other Project A staff, Task 2 staff will administer the battery for test-retest purposes to a sample of new recruits, as well as providing "on-call" assistance for the major administration effort.

(12) Analyze FY83/84 cohort data: trial battery/performance measures. Data will be available for concurrent validity analyses of the Trial Battery and predictive validity analyses of the Preliminary Battery, although on fewer MOS for the latter than for the former. Fairness analyses, generalizability analyses, and other analyses will also be conducted.

(13) Prepare Experimental Battery and prepare for administration to FY86/87 cohort. Based on analyses from Subtask 12, the Experimental Battery, i.e., the final, revised version of the Trial Battery will be prepared. Test administration materials will be prepared, and personnel designated as on-site test administrators will be trained by Task 2 staff.

(14) Monitor administration of Experimental Battery to FY86/87 cohort, and further analyze FY83/84 cohort data. The administration of the Experimental Battery will be carried out by Army personnel on site at training schools. Task 2 staff will make several scheduled inspection visits, as well as any unscheduled visits necessary to respond to problems. During this time, further analyses of the FY83/84 cohort data will be carried out.

(15) Analyze FY86/87 cohort data and prepare final reports. Predictor response distributions, covariances, etc., of the FY83/84 cohort and FY86/87 cohort will be compared to ascertain if substantial differences occur because of attrition (in the concurrent FY83/84 cohort sample) and other factors. Relationships of the predictors to training performance will be analyzed. Draft and final instrument and technical reports will be prepared.

The Literature Search for Selection Predictors

A major activity this year was a comprehensive literature search and review. The search was conduted by three research teams, each responsible for a broadly defined area of human abilities or characteristics. The three areas were cognitive abilities; noncognitive characteristics such as vocational interests, biographical data, and measures of temperament; and psychomotor/physical abilities. These areas or domains proved to be convenient for purposes of organizing and conducting literature search activities, but they were not used as (nor were they intended to be) a final taxonomy of possible predictor measures.

Within each area, the teams carried out essentially the same steps. These were:

(1) To compile an exhaustive list of potentially relevant reports, articles, books, or other sources.

(2) To review each source and determine its relevancy for the project by examining the title and abstract (or other brief review).

(3) To obtain the books, articles, or reports that were identified as relevant in step two above.

(4) For the materials obtained in step 3, review thoroughly each document for relevant information, and to put this information onto special review forms developed for this project.

Relative to step 1, several methods were used to insure as comprehensive a list of references as possible. First we conducted several computerized searches of relevant data bases. Appendix I presents their names and descriptions. Over 10,000 sources were identified via the computer search.

In addition to the computerized searches, we obtained reference lists from recognized experts in each of the three areas, emphasizing the most recent research in the field. We also obtained several annotated bibliographies from military research laboratories. Finally, we scanned the last several years' editions of research journals that are frequently used in each ability area as well as more general sources such as textbooks, handbooks, and appropriate chapters in the Annual Review of Psychology (an annual book that reviews the most recent research in a number of conceptually distinct areas of psychology).

As is usually the case with such exhaustive search techniques, the majority of the sources identified were not directly relevant for our purposes, for example, the identification and development of promising measures for personnel selection in the U.S. Army, and they were screened out in step 2. The relevant sources were reviewed and two record forms were completed for each source: an article review form and a predictor review form (several of the latter could be completed for each source). These forms were designed to capture the essential information in a standard format.

The output of the literature search served as input for: (a) the selection of the preliminary battery, (b) the writing of the literature review report, and (c) the formulation of a comprehensive model of the predictor space in the form of specifying the predictor constructs that seem to best describe the latent variables measured by the available tests, and (d) the development of the formal technical review that will begin in October of FY84.

Considerable staff time was devoted to defining the total array of constructs that seemed to account for the total predictor space. In a very real sense this was an important step in "theory development" as it pertains to the measurement of individual differences of Army applicants. It is also the array of constructs that will be used by the expert judges in the technical review to scale the expected relationships between the predictor constructs and the array of criterion factors that our current model says constitute total performance space.

The array of criterion factors was produced by the MOS job analyses, the critical incident workshop, the review of archival records, and the analysis of the AIT programs of instruction.

Again, considering both predictor space and criterion space in construct terms has been extremely valuable in our development work so far, and will continue to be so as we refine and expand our knowledge about both these domains through the major phases of the project.

Selection of the Preliminary Selection Test Battery

A second major activity conducted during the first year was the identification and development of the Preliminary Battery. This battery is intended to be an efficient, comprehensive set of predictors not covered by present Army preinduction measures. Its administration to trainees will allow an empirical determination of the extent to which additional, conceptually distinct predictor measures actually measure human abilities other than those that are currently measured and, through follow-up research, of the extent to which such measures add precision to the prediction of success in training performance and on-the-job performance.

The content of the Preliminary Battery was carefully chosen in as efficient a manner as possible to be as comprehensive as possible. The research staff first compiled a list of all even remotely appropriate measures identified in the literature search. This was called "List 1"; it was screened by eliminating measures according to several "Knockout" factors. That is, the following factors were used to eliminate potential predictors from further consideration: (a) measures developed for a single research project only; (b) measures designed for a narrowly specified population/ occupational group, e.g., pharmacy students); (c) measures targeted toward younger age groups; (d) measures requiring special apparatus for administration; (e) measures requiring unusually long testing times; (f) measures requiring difficult or subjective scoring; and (g) measures requiring individual administration.

The result of this screening process was a second and more manageable list of candidate measures. Each measure on "List 2" was evaluated on 12 factors, shown in Figure 11, by at least two knowledgeable members of the research staff. (A five-point rating scale of potential usefulness was used to rate each measure on each of the 12 factors.) These ratings were used to guide the selection of the measures for the third list. However, this list ("List 3") still contained too many measures to administer in the time available. Therefore, List 3 was subjected to a final review by Project A researchers with the emphasis placed on "best bets" for prediction of on-the-job performance, given their collective knowledge of the constructs measured by the potential predictors and the factors that make up the criterion space.

The final content of the Preliminary Battery was a set of eight, timed, cognitive ability tests; a biographical questionnaire; a personal opinion inventory; and a vocational interests inventory. These instruments, collectively, measure a large number of human attributes not currently tapped by preinduction testing.

1. Discriminability--extent to which the measure has sufficient score range and variance, i.e., does not suffer from ceiling and floor effects with respect to the applicant population.

2. Reliability--degree of reliability as measured by traditional psychometric methods such as test-retest, internal consistency, or parallel forms reliability.

3. Group Score Differences (Differential Impact)--extent to which there are mean and variance differences in scores across groups defined by age, sex, race, or ethnic groups; a high score indicates little or no mean differences across these groups.

4. Consistency/Robustness of Administration and Scoring--extent to which administration and scoring is standardized, ease of administration and scoring, consistency of administration and scoring across administrators and locations.

5. Generality--extent to which predictor measures a fairly general broad ability or construct.

6. Criterion-Related Validity--the level of correlation of the predictor with measures of job performance, training performance, and turnover/attrition.

7. Construct Validity--the amount of evidence existing to support the predictor as a measure of a distinct construct (correlational research, experimental research, etc.).

8. Face Validity/Applicant Acceptance--extent to which the appearance and administration methods of the predictor enhance or detract from its plausibility or acceptability to laymen as an appropriate test for the Army.

9. Differential Validity--existence of significantly different criterion-related validity coefficients between groups of legal or societal concern (race, sex, age); a high score indicates little or no differences in validity for these groups.

10. Test Fairness--degree to which slopes, intercepts, and standard errors of estimate differ across groups of legal or societal concern (race, sex, age) when predictor scores are regressed on important criteria (job performance, turnover, training); a high score indicates fairness (little or no differences in slopes, intercepts, and standard errors of estimate).

11. Usefulness of Classification--extent to which the measure or predictor will be useful in classifying persons into different specialties.

12. Overall Usefulness for Predicting Army Criteria--extent to which predictor is likely to contribute to the overall or individual prediction of criteria important to the Army (e.g., AWOL, drug use, attrition, unsuitability, job performance, and training).

FIGURE 11. Factors Used To Evaluate Predictor Measures for the Preliminary Battery

Identification of Potential Psychomotor and Perceptual Selection Tests

The development of computerized selection measures in the perceptual and psychomotor domains is a special emphasis of this project. (Computer-adaptive testing, as that term is usually employed, is being amply pursued by other military research projects and is not our primary focus.) Accordingly, we conducted several activities to get an early start on this part of the project. First, we visited four military laboratories or field units where currently active research utilizing such computerized measures was underway. Second, we developed a demonstration battery of computerized measures on a portable microprocessor (an Osborne 1) to become familiar with software and hardware problems. Finally, we reviewed the output of the literature search described above as regards the reliability and validity both of computerized measures for personnel selection and psychomotor/ perceptual tests.

The four site visits were the Air Force Human Resources Laboratory, Brooks Air Force Base, TX; the Naval Aerospace Medical Research Laboratory, Pensacola Naval Station, FL; the Army Research Institute Field Unit at Ft. Rucker, AL; and the Army Research Field Unit at Ft. Knox, KY. Personnel at all these sites were gracious and helpful. During these visits we tried to answer five questions. The questions and the answers we obtained can be summarized as follows:

(1) What computerized measures are in use?

We found over 60 different measures in use across the four sites. A sizable number of these were specialized simulators that are not relevant for this project (e.g., a helicopter simulator weighing several tons that is permanently mounted in an air-conditioned building).

There were, however, many measures in the perceptual, cognitive, and psychomotor areas that were relevant.

(2) What computers were selected for use?

(3) What computer languages are being used?

We observed three different microprocessors in use: the Apple, Terak, and PDP 11; and three different computer languages: PASCAL, BASIC, and FORTRAN. There appears, in fact, to be relatively little in common among the four sites in terms of the hardware/software used.

(4) How reliable are these computerized measures?

(5) What criterion-related validity evidence exists so far for these measures?

Data are currently being collected at all four sites to address the reliability and criterion-related validity questions. The research at AFHRL is at the point of administering computerized measures to fairly large samples of subjects. This is also true of the research at Ft. Rucker where they expect to have validity data collected and analyzed by sometime later this year. Documentation of the results of these efforts will allow estimation of the reliability and criterion-related validity of the measures under examination at these two locations.

A number of the measures have been under research at NAMRL for some time now, but criterion-related validity has not been the primary focus of that research. The prototype information processing measures developed there have been shown to be sensitive to individual differences within chronological age groups as well as to age-related changes across different age groups.

Data on the computerized measures at Ft. Knox are currently being analyzed. While there are potential problems with range restriction in the predictors and the criterion measures, significant, positive correlations between microprocessor measures and their higher fidelity, "hands-on" counterparts are being found.

To summarize, there is little information currently available regarding the reliability or criterion-related validity of the computerized measures in use at these sites. This is not surprising because most of these measures have been developed only recently.

After conducting these site visits, we programmed a short demonstration battery in BASIC on the Osborne 1, a portable microprocessor. The purpose of this activity was to implement some of the techniques and procedures observed during the site visits to determine the degree of difficulty of such programming and to gauge the quality of results to be expected from use of a common portable microprocessor (with a language that is common to many machines), which has some disadvantages in terms of processing power, speed, and flexibility. This short battery was self-administered, recorded the response and time to respond, and contained five tests: simple reaction time (pressing a key when a stimulus appeared); choice reaction time (pressing one of two keys in response to one of two stimuli); perceptual speed and accuracy (comparing two alphanumeric phrases for similarity); verbal comprehension (vocabulary knowledge); and a self-rating form (indicating which of two adjectives "best" describes the test taker, on a relative 7-point scale). We also experimented with the programming of several types of visual tracking tests, but did not include these in the self-administered demonstration battery.

Summary of First Year Activities

In sum, we have accomplished what we think is a landmark survey of potential selection measures for improving selection and classification decisions for U.S. Army enlisted personnel. Based on this survey, additional analyses of expert judgment, and several reviews, we developed the preliminary selection battery, which has been carefully designed to provide comprehensive information about what kinds of measures will provide the most useful supplements to the ASVAB. Finally, we have begun the initial development work for new psychomotor and perceptual tests that could become part of the preinduction test battery .

Next Steps in the Development of Selection Predictors

During the next year the following activities will be carried out for the purpose of developing and validating new selection measures.

1. The Preliminary Battery will be administered to trainees entering classes in the following MOS: 05C (Ft. Gordon), 19E/K (Ft. Knox), 63B (Ft. Dix and Ft. Leonard Wood), and 71L (Ft. Jackson). Testing monitors and administrators at each site were trained during September 1983, and testing will be carried out from October 1983 through June 1984.

2. A technical review of possible predictor measures will occur in October. This will consist of collecting and analyzing expert judgments of the expected relationship between the most promising predictor constructs and the various performance factors in the training, Army wide, and MOS-specific performance domains. The predictor and criterion variables to be used by the raters are shown in Appendix J.

3. In November, a panel of Army personnel responsible for administration of preinduction measures will review the predictor measures that come out of the technical review, in order to identify serious administrative problems with any measures that are scheduled for further development.

4. A draft of the literature review report will be completed in October.

5. Development of computerized measures will continue, including a pretest in November and January at a MEPS station. A preliminary report on computerized measures will be prepared in March.

6. In December, the measures to be included in further development will be decided on (based primarily upon the technical and cost reviews mentioned in points 2 and 3 above). These measures will then be known as the Trial Battery. Item writing will begin and tryouts are scheduled for March, April, and May of 1984.

7. Initial data from the Preliminary Battery administration will be analyzed in January and February, and the results will be used to inform Trial Battery development (described in point 6).

8. In June 1984, the Trial Battery will be put into final form for the pilot test in summer/fall of 1984 on USAREUR.

Associated Reports

The research efforts listed touch on two concerns about enlistment predictors which are frequently raised by the Army: How do the current predictors really work? What exactly are we measuring and is it what we should be measuring?

The practice effects and comparability papers respond to the operational concern about the stability of the score scale of the current cognitive predictors found in ASVAB. The score scales are, in general, quite stable. Retesting leads to noticeable improvement in the average scores for the speeded subtests (Numerical Operations and Coding Speed) but little change in the other subtests. Hence, while there is some improvement in composite scores (AFQT, Aptitude Area composites) for those retested, the amount of improvement is small. An example is that of an initial AFQT Mental Category score of IVB being most likely (.84) to retest to no higher than a IVA. Further, the factor structure of the battery is constant over retesting.

When test scores change over time it is likely that the changes have causes other than practice, or retesting. One other cause is an improvement in the applicant pool in general. Another possible explanation, identified during 1980, for ASVAB 6 & 7 is that the score scales may be miscalibrated; the test battery may become less difficult. The second research project is a detective story. It was noticed that the quality of Army applicants improved dramatically from the late 1970's to the early 1980's, an improvement which coincided with the introduction of new forms of the ASVAB. Was the norming of the new forms accurate? To answer the question required comparing several data bases of Army applicants and accessions, with scores from different forms of the ASVAB administered at different times and under different conditions. The answer was yes, the new ASVAB is correctly normed and yes, the Army is accessioning higher quality enlistees.

It is important and comforting to know that current Army predictors are accurate. On the other hand, a major purpose of Project A is also to investigate new predictors. Ideas for predictors come from theory, but we have to assure that they are what we need. The third paper, on information processing, describes an attempt to apply state-of-the-art concepts of cognitive processing to existing Army measures, to see if and how the theory can be applied. We concluded that the theory can, indeed, be applied successfully: Significant criterion-related validity coefficients can be obtained from scales developed to tap these new theoretical constructs. We were not so successful as to indicate that no further research is necessary, however. Another approach is described in the fourth paper, on ability requirements. In this research, a subset of predictor abilities was presented to military personnel in selected MOS to elicit their judgments of the relevance of these abilities to their MOS. The results demonstrate that such judgments can be made reliably and that they are understandable by the judges. The next step is to determine the validity of this approach, research that is most appropriate to the scope of Project A and which would not be possible in a smaller, less comprehensive effort.

The Impact of Practice Effects on ASVAB Test Scores:*
Some Implications for Initial Selection
and Classification of Military Personnel

David Friedman and Arline H. Streicher
RESEARCH APPLICATIONS, INCORPORATED

Hilda Wing and Frances Grafton
United States Army Research Institute for the
Behavioral and Social Sciences

The Armed Services Vocational Aptitude Battery (ASVAB) is used to screen and initially classify applicants for military service. The ASVAB is a multiple aptitude instrument which consists of eight power and two speeded subtests. The power tests are designed to obtain measures of cognitive abilities, as well as knowledge specific to the areas of mechanics, auto/shop and electronics. The speeded subtests are generally designed to test for perceptual speed skills of the individual.

Applicants are initially placed into mental and job ability categories based on the scores achieved on the ASVAB. Each branch of the service has developed specific job ability composites which consist of various combinations of ASVAB subtest scores. All four of the major branches of the service use a composite, known as the Armed Forces Qualification Test (AFQT), to assign applicants to categories of mental ability to determine, in part, whether or not the applicant is mentally fit for selection.

The AFQT composite of the ASVAB consists of four subtests: (1) Arithmetic Reasoning (AR) containing 30 items; (2) Word Knowledge (WK) containing 35 items; (3) Paragraph Comprehension (PC) containing 15 items; and (4) Numerical Operations (NO) containing 50 items. The formula used to determine the AFQT score is the arithmetic sum of the correct answers for AR, WK, PC, and one-half of the correct answers for the NO subtest – ($AR = WK + PC + \frac{1}{2} NO$). The raw score range for AFQT is 0 to 105 (Maier & Grafton, 1981).

Two typical job abilities composites used by the Army are the Combat (CO) and Electronics (EL) composites. CO is based on the sum of four subtest scores: (1) AR; (2) Coding Speed (CS) which consists of 84 items; (3) Auto/Shop Information (AS) which consists of 25 items; and (4) Mechanical Comprehension (MC) which consists of 25 items. CO ranges from 40 to 155. The EL composite also is based on the arithmetic sum of four subtest scores: (1) General Science (GS) containing 25 items; (2) AR; (3) Mathematics Knowledge (MK) containing 25 items; and (4) Electronics Information (EI) containing 20 items. EL score ranges from 40 to 147.

The AFQT and CO composites each consist of three power and one speeded subtest. The EL composite consists of four power subtests.

* Paper presented at the 24th Annual Conference of the Military Testing Association in San Antonio, Texas, November 1982.

Typically, approximately 500,000 persons apply for service in the Army each year. Of these applicants, at least six percent take the ASVAB on more than one occasion. Those persons who repeat the test do so for a number of reasons. They are requested by MEPS personnel to repeat a portion of the test in order to verify their scores, they are encouraged by recruiters to repeat the test in order to improve their scores, or the individuals themselves ask for a retest.

Verification retesting of individuals can occur immediately after the person takes a full test battery. A complete retest of the individual is not normally allowed until six months following any administration of the battery. However, a waiver can be obtained to repeat the administration after only thirty days. Since the stated purpose of retesting on the ASVAB is to verify or improve test scores, a research question was posed which sought to determine the effect of score changes, if any, on the selection and initial placement of recruits who took the ASVAB on repeated occasions.

The data used to address this question were taken from a study of the reliability of the ASVAB subtest scores reported for FY 1981 Army applicants (Friedman, et. al., 1982). The pool of retested applicants who served as the subjects for this study consisted of all individuals for whom there were recorded two different test forms of the current ASVAB.[1] Individuals who were identified as having taken a verification retest or who had been administered the same form of the ASVAB on two different occasions were not included in this subject pool. In addition, a pool of more than 140,000 one-time-only test takers were identified from the Army applicant file. These individuals served as subjects for purposes of comparative analyses.

A factor analysis of the test scores achieved by the one-time-test takers and the repeaters was conducted. The results showed the same two factor structure of the eight power and two speeded tests for each group. Thus, it was concluded that the test scores achieved by these groups were acting in the same way for the battery as a whole.

In Table 1, the mean subtest scores achieved by one-time-only and repeat test takers on the last previous and most recent test taken are presented. It is possible for some of these applicants to have taken the ASVAB on more than one occasion; a different form of the ASVAB, such as the high school version could have been taken at a previous time. However, for this study only the scores reported for the two most recent test administrations were used. Furthermore, only individuals who took versions of the currently operational ASVAB were used as subjects for this study. As can be seen, the average subtest scores for the one-time-only test takers are greater than the average subtest scores for repeat test takers for all subtests except Coding Speed; and that only for the most recent test taken. This change in test score for a speeded test raises the issue of the possible effect of speeded subtest scores on the selection of applicants. Since these subtest scores are combined into composites, such as the AFQT, CO and EL, changes in subtest scores as a result of retesting could have a significant impact on the selection and initial placement of applicants.

[1] The currently operational ASVAB consists of six parallel forms.

TABLE 1

MEAN SUBTEST SCORE COMPARISON FOR FY 1981 ARMY APPLICANTS

		One Time Test Takers (N = 143,279)	Test Repeaters (N = 27,923)				
			Most Recent Test Taken		Last Previous Test Taken		
ASVAB Subtest Name	Number of Items in Subtest	$\overline{X}$ (1)	$\overline{X}$ (2)	s.d.	$\overline{X}$ (3)	s.d.	z value (2,3)
GS General Science	25	13.893	11.119	3.789	10.768	3.705	10.50
AR Arithmetic Reasoning	30	15.953	11.912	4.216	11.355	3.830	16.34
WK Word Knowledge	35	22.307	17.985	5.737	17.465	5.479	10.95
PC Paragraph Comprehension	15	9.331	7.772	2.846	7.344	2.707	18.21
NO Numerical Operations	50	33.319	32.271	9.322	29.334	8.853	38.18
CS Coding Speed	84	41.695	42.007	13.449	36.352	13.207	50.13
AS Auto/Shop Information	25	14.433	11.885	4.747	11.369	4.645	12.98
MK Mathematics Knowledge	25	10.932	8.326	3.039	7.993	2.972	13.09
MC Mechanical Comprehension	25	13.292	11.078	4.106	10.355	3.884	21.37
EI Electronics Information	20	10.905	9.114	3.209	8.721	3.181	14.53

The effect of test score changes are demonstrated in Table 2. The first set of columns, "ASVAB Subtest," indicates the proportional contribution of each subtest to the composite as directed by the composite structure. The next two sets of columns indicate what the actual proportional contributions were for these two groups of subjects. For each of AFQT, CO, and EL, the data show the average percent of correct items for the two groups for each subtest of the three composites (Row B). As can be seen, the one-time test takers consistently scored higher on all subtests in each composite. The impact of the individual subtest scores on each composite is shown in Row E. For the AFQT, the percent contribution of NO is greater for both one-time-only test takers and repeaters. For the Army Combat (CO) composite, CS makes the greatest percent contribution, even for the ideal score (see ASVAB subtest columns). In the Army Electronics (EL) composite, no one test score contributes more than any other score to the composite. EL contains no speeded subtest.

Scores obtained on the two speeded subtests contribute significantly to the overall score obtained for AFQT and CO. The contribution of these subtest scores does affect the composite category to which the applicant is assigned. On the AFQT, a shift was made by more than 35% of the CAT IVA repeat test takers to CAT IIIB (see Table 3). Since the Army accepts from Category III persons and generally not from Category IV, this shift is dramatic. A similar shift occurs for those repeat test takers in the 95-99 range of the Army CO composite. More than 27% shifted to the next higher category. This large shift does not occur in the Army EL composite.

Both the AFQT and the Army CO composites included speeded subtests. The data show a significant increase in mean scores achieved on the NO and CS subtests of the ASVAB. These, in turn, contribute more to these composites than do the other subtests. The increase in the number of items correct in these speeded subtest scores can result in changes in the category to which the applicant is assigned in each composite. In some cases these shifts affect whether or not the individual is selected and properly placed. Given the fact that an elaborate research program is being undertaken that will consider the use of ASVAB scores to predict performance, more research is needed to evaluate all the effects of speeded tests such as those presently used in the ASVAB.

TABLE 2
FY 1981 ARMY APPLICANTS

COMPOSITE	ASVAB Subtest				One time-only test takers (N=143,279)				Repeaters Most Recent Test Taken (N=27,923)			
Armed Forces Qualification Test (AFQT)	AR	WK	PC	NO	AR	WK	PC	NO	AR	WK	PC	NO
A. Number of Items in Subtest	30	35	15	25	30	35	15	25	30	35	15	25
B. Average Percent Correct	100	100	100	100	53.3	64.3	62.7	66.8	39.7	51.4	51.8	64.5
C. Number of Items (A x B)	30	35	15	25	16	23	9	17	12	18	8	16
D. Standard Score Equivalent	--	--	--	--	--	--	--	--	--	--	--	--
E. Percent Contribution	29	33	14	24	25	35	14	26	22	33	15	30
Army Combat (CO)	AR	CS	AS	MC	AR	CS	AS	MC	AR	CS	AS	MC
A. Number of Items in Subtest	30	84	25	25	30	84	25	25	30	84	25	25
B. Average Percent Correct	100	100	100	100	53.1	49.9	58.0	53.2	39.7	50.0	47.5	44.3
C. Number of Items (A x B)	30	84	25	25	16	42	15	13	12	42	12	11
D. Standard Score Equivalent	67	75	65	67	48	49	48	46	42	49	43	42
E. Percent Contribution Based on Standard Score Equivalent	24	28	24	24	25	26	25	24	24	28	24	24
Army Electronics (EL)	GS	AR	MK	EI	GS	AR	MK	EI	GS	AR	MK	EI
A. Number of Items in Subtest	25	30	25	20	25	30	25	20	25	30	25	20
B. Average Percent Correct	100	100	100	100	55.6	53.3	44.0	55.0	44.5	39.7	33.3	45.6
C. Number of Items (A x B)	25	30	25	20	14	16	11	11	11	12	8	9
D. Standard Score Equivalent	67	67	71	68	44	48	48	47	40	42	43	42
E. Percent Contribution Based on Standard Score Equivalent	24.5	24.5	26.0	25.0	23.5	25.7	25.7	25.1	24.1	25.1	25.7	25.1

TABLE 3

RETEST SHIFTS IN AFQT ASSIGNED
MENTAL CATEGORY IN PERCENTS*
FY81 ARMY APPLICANTS N=27,923

Most Recent Test	CATV	CATIVC	CATIVB	CATIVA	CATIIIB	CATIIIA	CATI&II
I&II	1	0	0	0	2	30	71
IIIA	1	0	0	3	16	34	16
IIIB	2	2	9	36	59	29	6
IVA	3	14	34	41	18	4	3
IVB	7	29	35	15	3	2	0
IVC	36	44	20	4	1	2	1
V	50	11	2	1	0	0	1
Percent* of Total	9	32	23	31	4	1	1

Last Previous Test

*Figures may not add to 100 because of rounding.

REFERENCES

Friedman, D. & Streicher, A. Study of the reliability of scores for fiscal year 1981 Army applicants: Armed Services Vocational Aptitude Battery forms 8, 9 and 10, Final Report, July, 1982.

Maier, M. & Grafton, F. Scaling Armed Services Vocational Aptitude Battery (ASVAB) form 8AX, Research Report 1301, U.S. Army Institute for the Behavioral and Social Sciences, January, 1981.

WORKING PAPER SELECTION AND CLASSIFICATION TECHNICAL AREA 82-7

FINAL STATUS REPORT ON THE COMPARABILITY OF ASVAB 6/7 and 8/9/10 APTITUDE AREA SCORE SCALES

FRANCES C. GRAFTON, KAREN J. MITCHELL AND HILDA WING

1 MARCH 1983

APPROVED BY:
NEWELL K. EATON
Chief, Selection and
Classification
Technical Area

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

*To be published as Working Paper Selection and Classification Technical Area 82-6.

FINAL STATUS REPORT ON
THE COMPARABILITY OF ASVAB 6/7 AND ASVAB 8/9/10
APTITUDE AREA SCORE SCALES

PROBLEM: Preliminary analyses of Defense Manpower Data Center (DMDC) and Military Enlistment Processing Command (MEPCOM) accession data for FY78-FY81 point to discrepancies in the distributions of aptitude area scores for ASVAB 6/7 and ASVAB 8/9/10. Of concern are the (1) equivalence of Armed Forces Qualification Test (AFQT) and aptitude area scores assigned through the recalibrated ASVAB 6/7 and ASVAB 8/9/10 testing program, and (2) quality of applicants accessed during Fiscal Years 1977-1982. These issues were addressed, in part, in the Interim Status Report on the Comparability of ASVAB 6/7 and ASVAB 8/9/10 Aptitude Area Score Scales, by Grafton, Mitchell and Wing (ARI Selection and Classification Technical Area Working Paper 82-5). This final report more completely addresses these concerns.

Data presented in the interim report indicated that AFQT and aptitude area scores for ASVAB 6/7 CY76-78 accessions with potential ineligibles excluded and for FY81 accessions taking ASVAB 8/9/10 were similar. Mean AFQT scores were slightly below the population mean of 50; the mean aptitude area scores clustered around 100. Potential ineligibles were high school graduate accessions entering in FY76-FY80 and scoring below the 16th percentile on the recomputed 1980 AFQT scale or were non-high school graduate accessions scoring below the 31st percentile. Mean AFQT and aptitude area scores for FY82 accessions taking ASVAB 8/9/10 were slightly higher than FY81 8/9/10 means. The mean aptitude area scores for FY81 accessions administered ASVAB 6/7 were higher than the population means, while the AFQT mean for this group was below the population mean. Examination of the FY81 ASVAB 6/7 data revealed that approximately 85% of the aptitude area scores reported on the MEPCOM tapes were computed using the miscalibrated ASVAB 6/7 conversion tables. MEPCOM-reported AFQT scores, however, were the correctly computed scores from the 1980 conversion tables.

In the interim report, operationally administered ASVAB 6/7 aptitude area scores were also compared to experimentally administered ASVAB 8A scores for 2375 male military applicants. For each aptitude area, mean ASVAB 6/7 scores were slightly higher than ASVAB 8A scores. It is hypothesized that this difference may reflect the compromise problem observed on ASVAB 6/7 in its fourth year of operational use or motivational differences for applicants taking operational and experimental forms. Other explanations for observed differences may be uncovered with further investigation.

OBJECTIVE: Specifically, to (1) compare FY81 ASVAB 8/9/10 accession data to data on FY81 accessions with correctly computed ASVAB 6/7 scores, and (2) compare the score distributions for ASVAB 8/9/10 FY81 accessions with the score distributions for ASVAB 6/7 FY81 recruits with correctly computed AFQT and aptitude area scores, excluding data for potential ineligibles.

METHOD/RESULTS: AFQT and aptitude area score cumulative frequency plots were evaluated for CY76-78 ASVAB 6/7 accessions, FY81 ASVAB 6/7 accessions (for whom all aptitude area scores were recomputed by ARI), and FY81 and FY82 ASVAB 8/9/10 accessions. These appear in Figures 1 through 11. Cumulative frequency plots of data of this type are conventionally presented with the frequency index on the ordinate and the ability variable represented on the abscissa. The assistance request that generated this work stipulated that the data be presented in the following manner. Cumulative frequency plots are drawn with the aptitude area score scales on the ordinate. The abscissa indexes the percent of accessions scoring at or above a given score point. Mean AFQT and aptitude area scores for these groups appear in Table 1.

Score means and cumulative frequency plots for the FY82 accessions taking ASVAB 8/9/10 exceeded those of other groups on AFQT and all aptitude areas. Score means and plots for FY81 ASVAB 8/9/10 accessions were slightly below the FY82 examinees. Score means and plots for FY81 ASVAB 6/7 subjects were below the FY81 ASVAB 8/9/10 group. CY76-78 ASVAB 6/7 accessions had the lowest scores.

The higher FY82 AFQT and aptitude area scores reflect the rise in accession standards instituted in March 1982. The difference in FY81 ASVAB 6/7 and ASVAB 8/9/10 examinees can be explained, in part, by the fact that potential ineligibles taking ASVAB 6/7 in FY80 were accessed in FY81 through the Delayed Entry Program. Potential ineligibles were high school graduates scoring below the 16th percentile on the correct AFQT scale or non-high school graduates scoring below the 31st percentile. When these subjects were excluded from the analyses on the FY81 ASVAB 6/7 group, the cumulative frequency plots for these examinees approximated the FY81 ASVAB 8/9/10 plots for AFQT and nine of the aptitude areas (all but CO). These plots point to the equivalence of AFQT and nine of the aptitude area scores assigned through the ASVAB 6/7 and 8/9/10 testing programs for examinees accessing in the same fiscal year. Cumulative frequency plots for CY76-78 accessions administered ASVAB 6/7 were below the plots for the other groups on all indices. Like the ASVAB 6/7 FY81 data, these data include records of potential ineligibles.

Modest differences between the cumulative frequency plots for FY81 ASVAB 6/7 examinees with potential ineligibles excluded and FY81 ASVAB 8/9/10 examinees were noted on the Combat (CO) aptitude area (Figure 8). The FY81 ASVAB 6/7 frequency plot falls below the 8/9/10 FY81 plot at all score points. In comparison to other aptitude areas (97.4-101.3), the Combat mean (X=95.3) is low for all ASVAB 6/7 data reported here. The mean score on Combat for the FY81 ASVAB 8/9/10 group (X=99.4), however, is very close to the population mean of 100; the same is true of the CO mean for the FY82 group (X=101.9). Potential sources of a discrepancy for the ASVAB 6/7 CO composite are discussed below. The ASVAB 8/9/10 CO composite mean is close to the population mean of 100 for both FY81 and FY82 examinees.

The reason that observed scores on FY81 ASVAB 6/7 CO fall below the reference population mean of 100 cannot be determined from these data. The discrepancy may be the result of random measurement or equating error; it may be traced to the combination of subtests summed in the composite, or it may be attributable to a psychometric or operational event not identified here. When ASVAB 6/7 was recalibrated, the composites were referenced to AFQT 7A. For each composite the subtest raw scores were summed and equipercentile equating techniques were used to equate the cumulative frequency distribution of the composite to that of AFQT 7A. Random error may exist in any equating process. Any such error in equating a particular composite would not affect norming efforts for the other composites. Random equating error may have been different for CO than for the other composites. Alternatively, the low mean scores on the ASVAB 6/7 CO composite may be attributable to the particular subtests summed in the composite. On ASVAB 6/7, CO contained the AR, SI, SP, and AD subtests; it also contained the CC classification inventory. It was the only composite containing both a spatial and a perceptual subtest, Space Perception and Attention-to-Detail. It was the only composite using the CC, the Outdoors classification inventory. No spatial or classification inventory score is used in any 8/9/10 composites. It is not possible to test hypotheses about these possible source(s) of the discrepancy using currently available data.

Aptitude area cumulative frequency plots for 2375 male military applicants experimentally administered ASVAB 8A and administered ASVAB 6/7 through the operational testing program in FY80 were also evaluated. These appear at Figures 12 through 21. Frequency plots for the ASVAB 6/7 aptitude areas were above the 8A plots in all cases. The difference may reflect such things as sampling error, differences in motivation for examinees taking operational and experimental batteries, and/or compromise on the operationally administered forms. The plots for the General Technical area crossed at the low score range. The Combat area plots crossed at the high end. All frequency plots were judged to be highly similar across forms. CO was not exceptional in comparison to the other aptitude areas as it was in previous analyses.

CONCLUSIONS: The data show an increase in accession quality on AFQT and the aptitude area composites during Fiscal Years 1977-1982. The FY82 accessions had mean AFQT and aptitude area scores above the population means; this difference reflects the rise in accession standards instituted in March 1982. FY81 accessions taking ASVAB 8/9/10 had slightly lower AFQT and aptitude area score means than those in FY82. The means and cumulative frequency plots for FY81 ASVAB 6/7 accessions, with potential ineligibles excluded, were comparable to the ASVAB 8/9/10 data for AFQT and nine of the aptitude areas. Mean scores for CY76-78 and FY81 ASVAB 6/7 accessions with records of potential ineligibles included in the sample were below the population means.

A modest difference in the Combat aptitude area mean for ASVAB 6/7 examinees with potential ineligibles excluded and FY81 ASVAB 8/9/10 examinees was noted. This may reflect random measurement or equating error; it may be traced to the individual subtests that made up the composite, or it may be attributable to an event not identified here. The source of the difference between the ASVAB 6/7 and 8/9/10 CO composites could not be determined from these data. The CO data for the FY81 and FY82 ASVAB 8/9/10 examinees were consistent with the population data.

Cumulative frequency plots for examinees operationally administered ASVAB 6/7 were somewhat higher than the plots for experimentally administered ASVAB 8A. This difference may reflect such things as sampling error, differences in motivation for operationally and experimentally administered testings, and/or compromise on the operationally administered forms.

Table 1

Mean AFQT & Aptitude Area Scores for CY76-78 and FY81-82 Accessions

Score	CY76-78 ASVAB 6/7 (N=309937) PI Included**	FY81 ASVAB 6/7* (N=24100) PI Included**	FY81 ASVAB 6/7* (N=20416) PI Deleted**	FY81 ASVAB 8/9/10 (N=82441)	FY82 ASVAB 8/9/10 (N=120368)
AFQT	41.4	42.9	47.4	48.3	53.4
GT	95.5	95.1	98.5	99.5	102.9
GM	93.6	95.0	97.6	97.4	100.1
EL	94.3	96.1	98.1	98.5	101.7
CL	94.6	96.0	99.2	100.0	102.4
MM	93.3	95.3	97.0	97.9	100.2
SC	92.8	94.0	97.4	99.3	101.8
CO	90.4	93.1	95.3	99.4	101.9
FA	95.9	98.6	101.3	100.2	103.0
OF	93.4	96.4	98.3	98.0	100.6
ST	96.0	97.7	100.5	97.4	100.7

* FY81 ASVAB 6/7 scores were computed using the August 1980 score conversion table.

** PI are potential ineligibles who were accessed during 76-81. An accession was deemed to be a PI if a high school graduate scoring below the 16th percentile on the recomputed AFQT scale or if a non-high school graduate scoring below the 31 percentile.

APTITUDE AREA AFQT

Figure 1

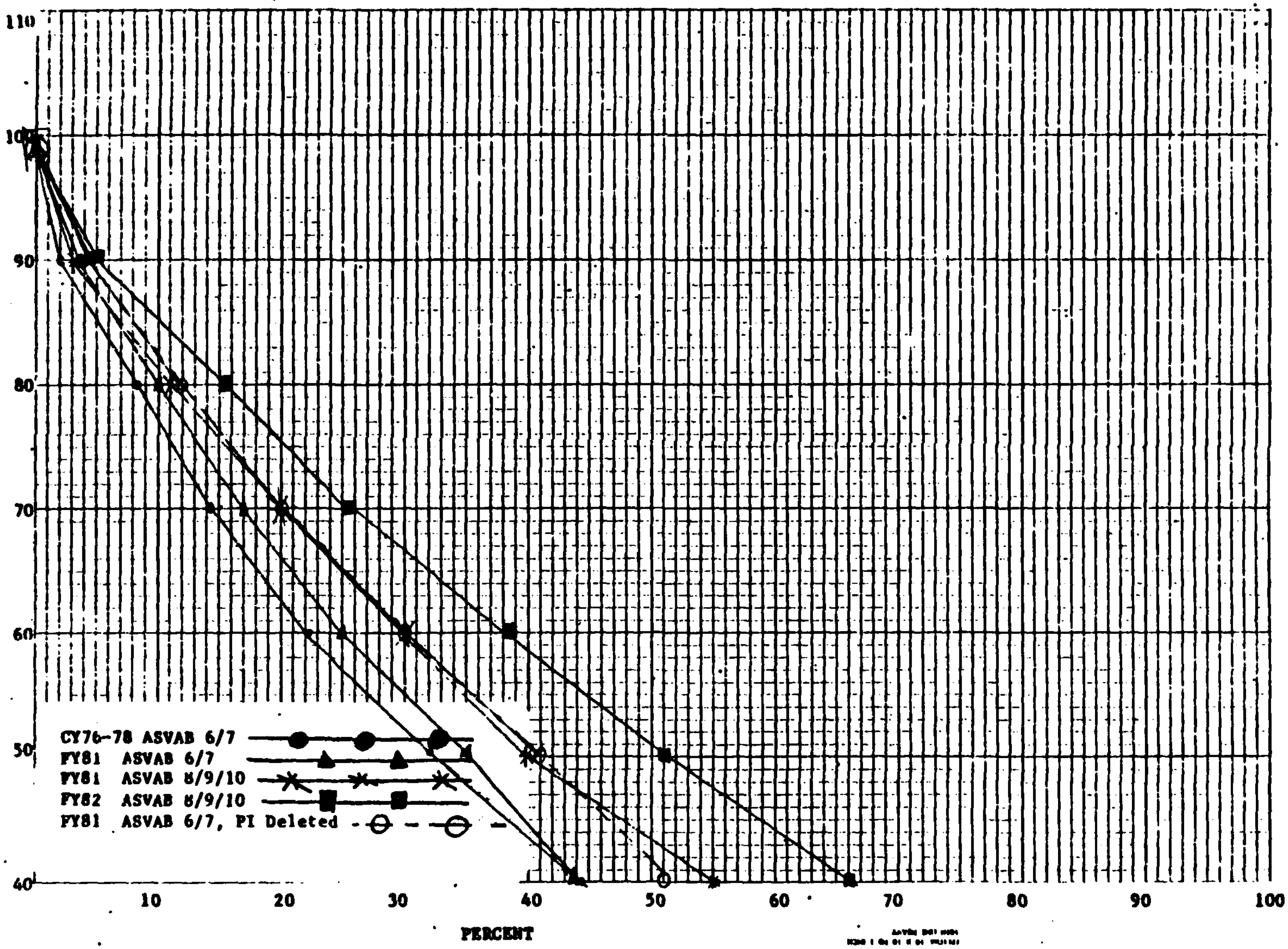
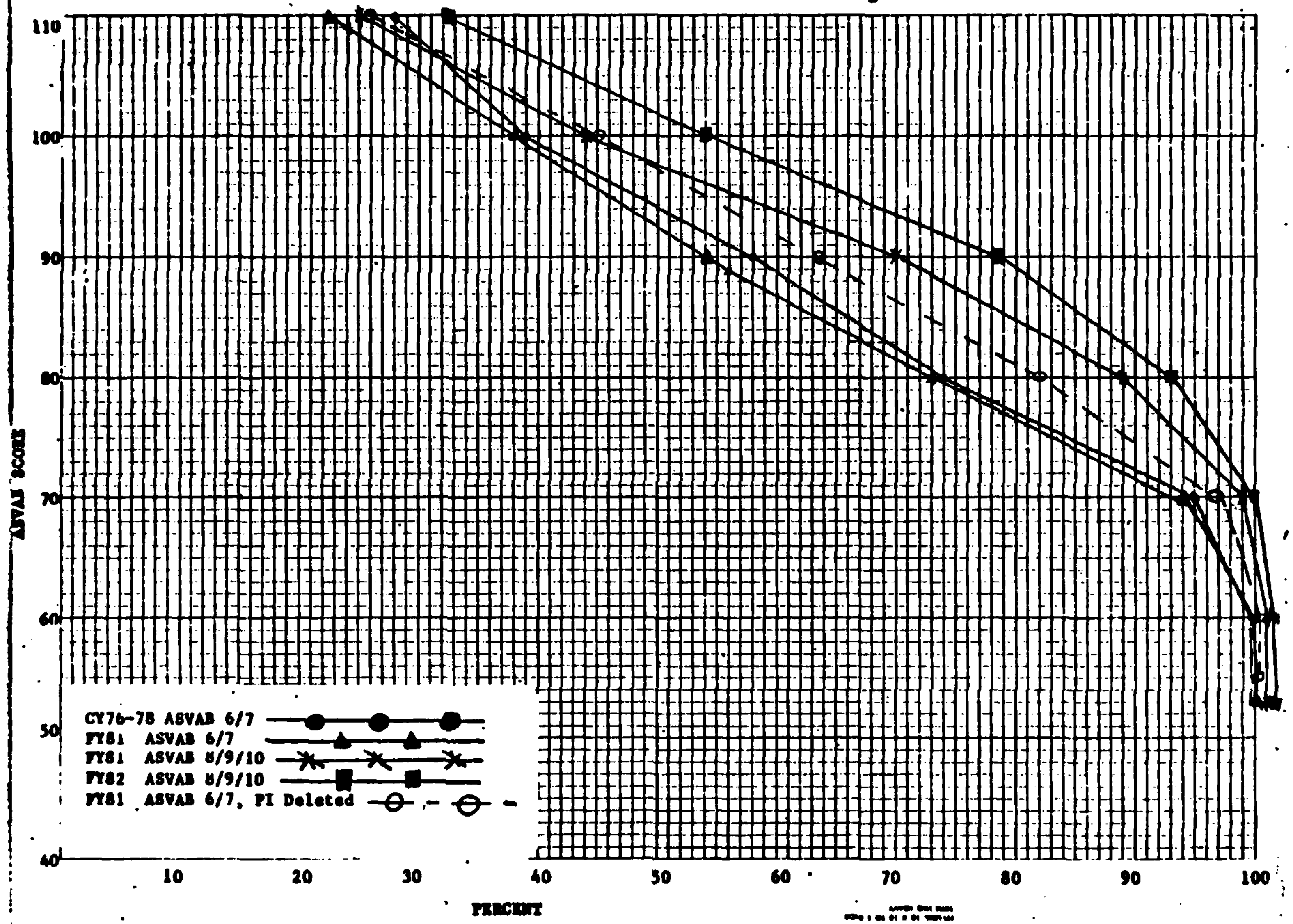
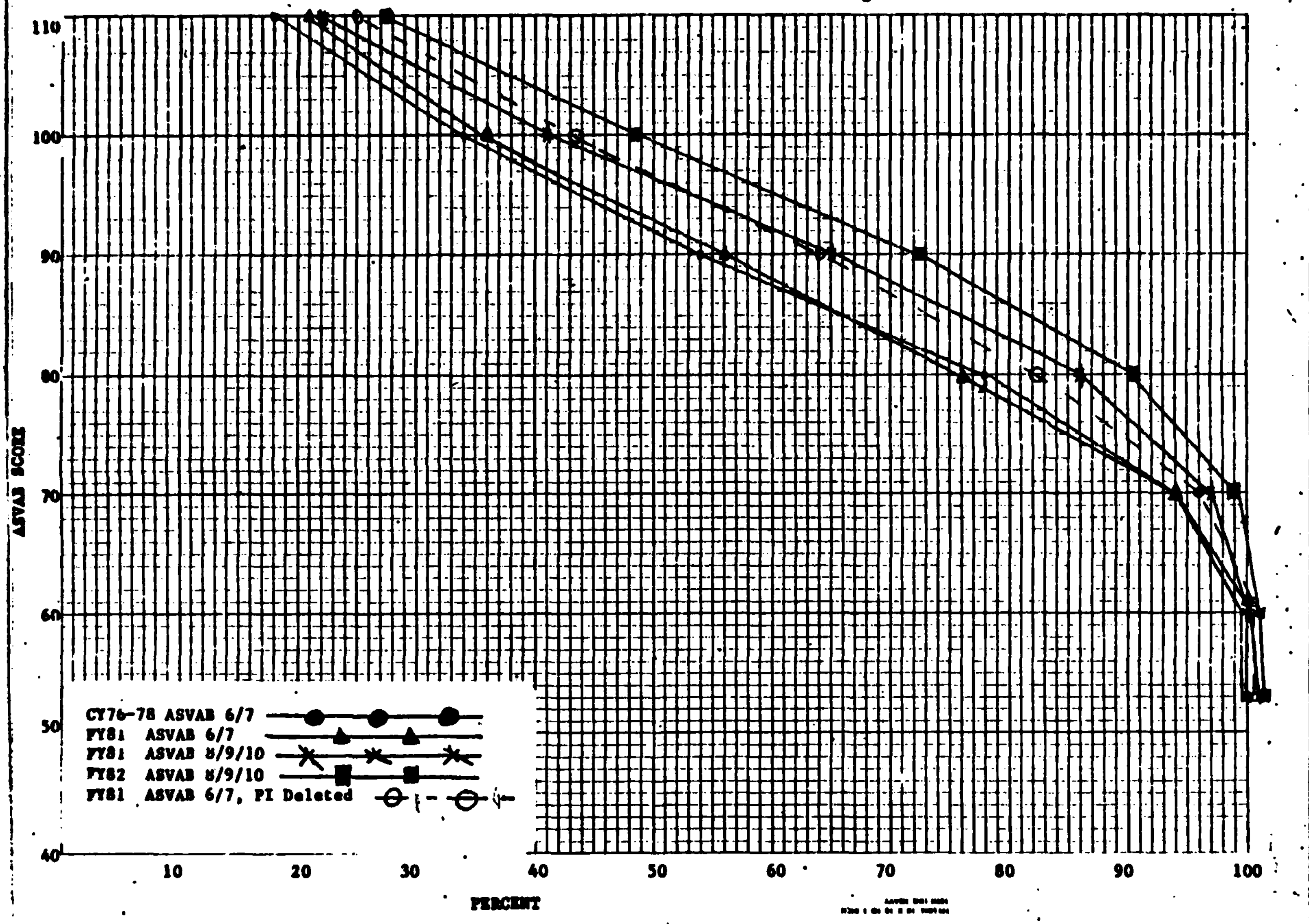


Figure 2


APTITUDE AREA GT
ASVAB SCORE
PERCENT
CY76-78 ASVAB 6/7
FY81 ASVAB 6/7
FY81 ASVAB 8/9/10
FY82 ASVAB 8/9/10
FY81 ASVAB 6/7, PI Deleted
110
100
90
80
70
60
50
40
10
20
30
40
50
60
70
80
90
100

APTITUDE AREA GM

Figure 3



APTITUDE AREA EL

Figure 4

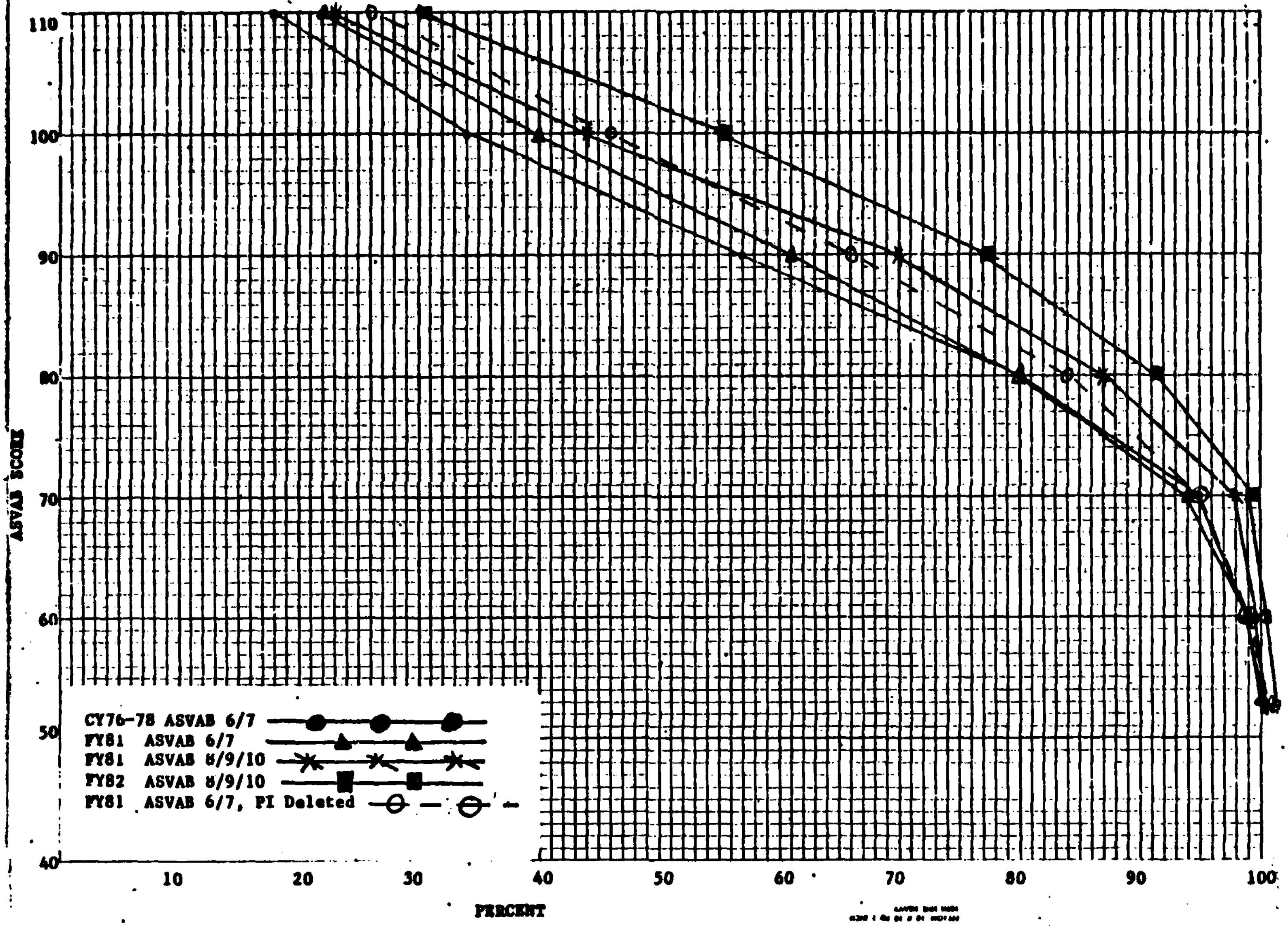
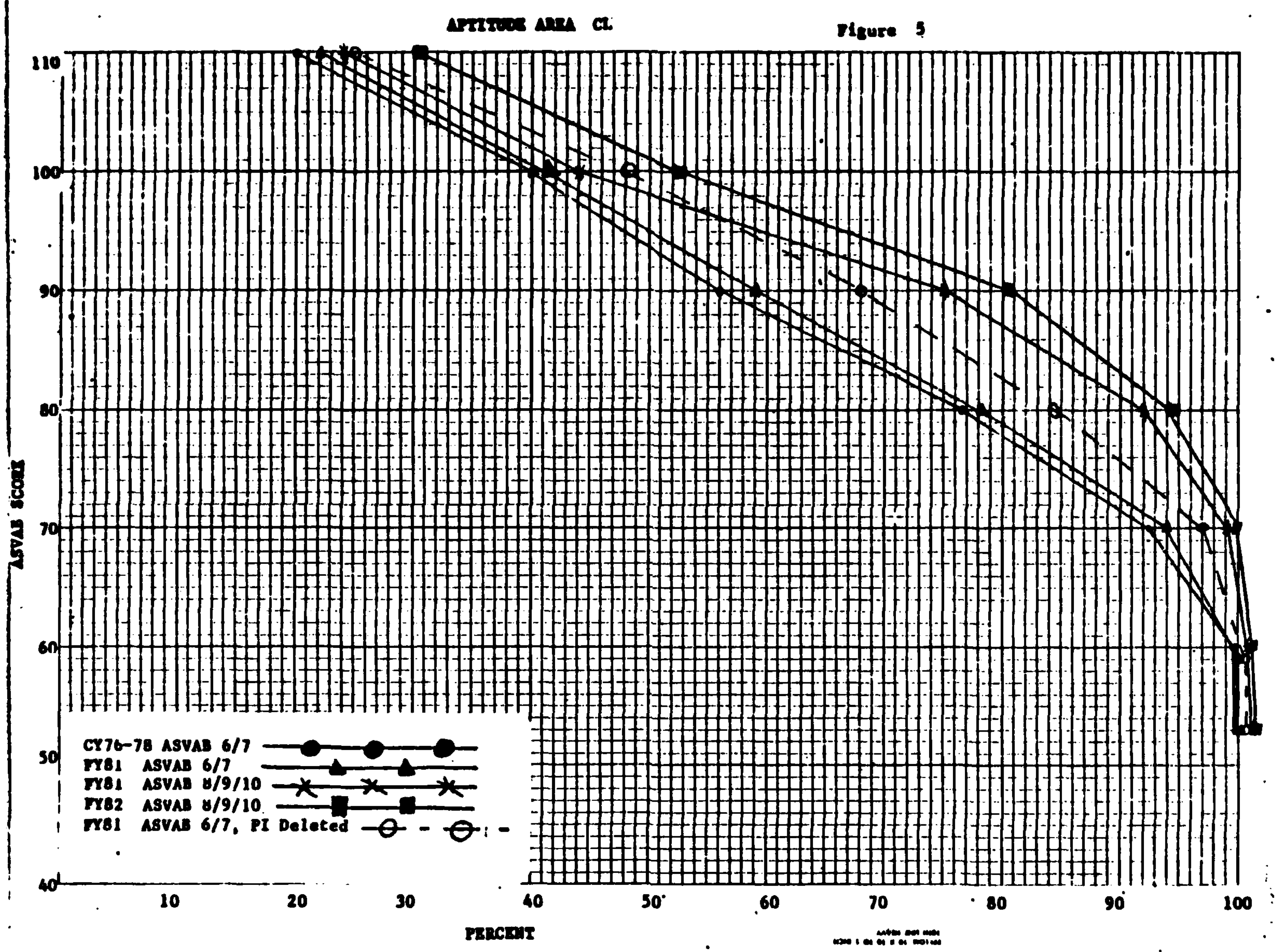
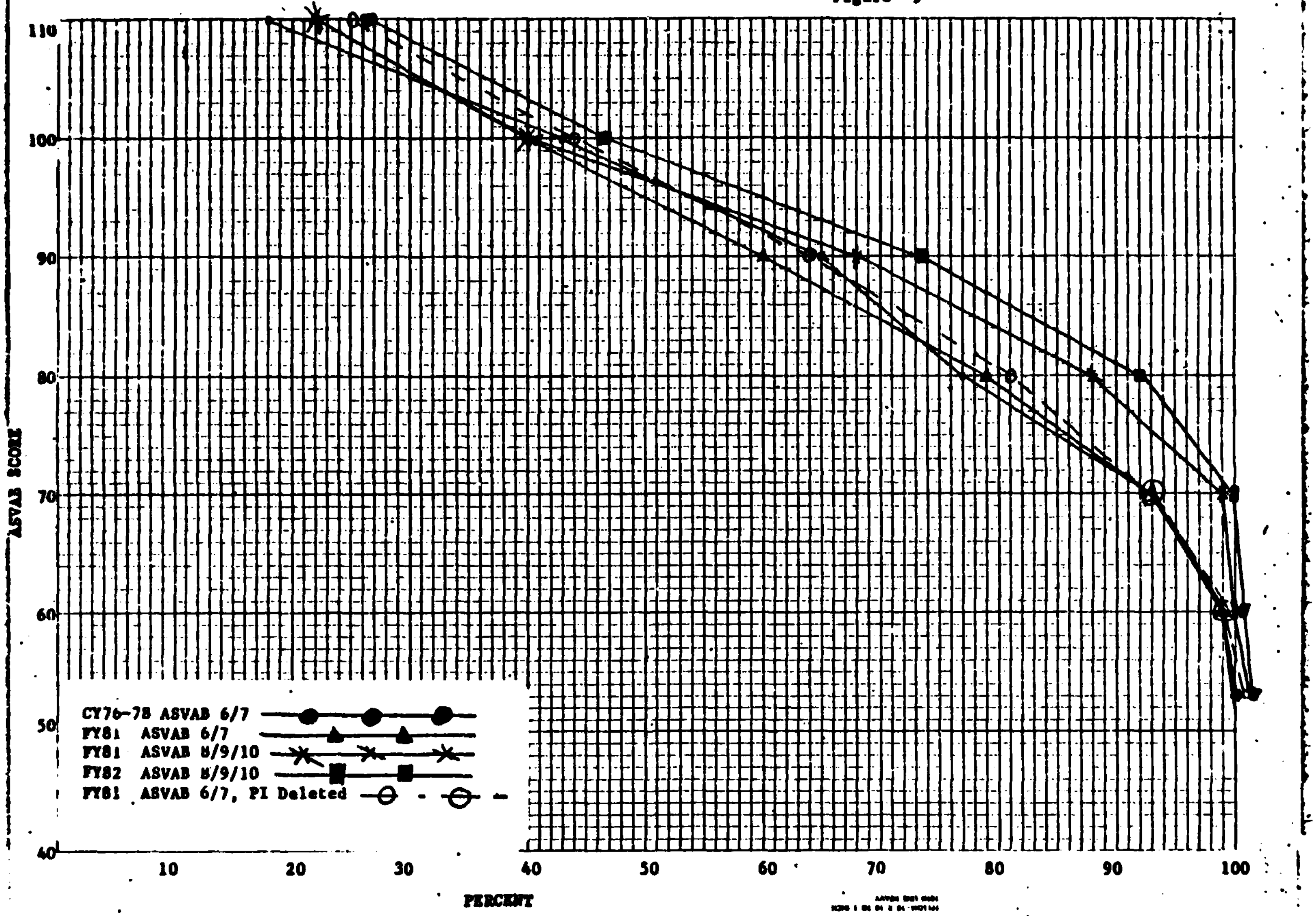


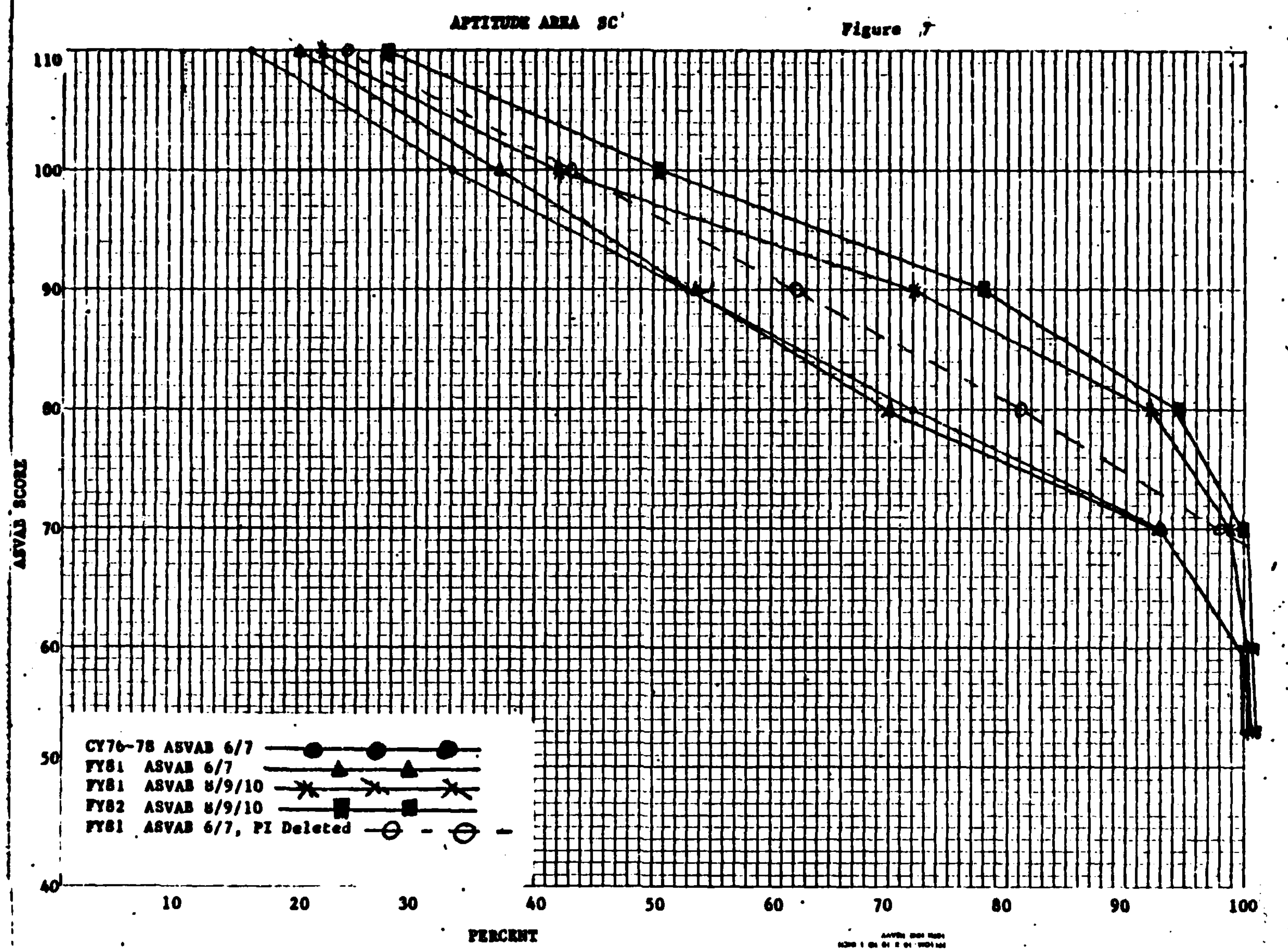
Figure 5


APTITUDE AREA CL
ASVAB SCORE
110
100
90
80
70
60
50
40
PERCENT
10
20
30
40
50
60
70
80
90
100
CY76-78 ASVAB 6/7
FY81 ASVAB 6/7
FY81 ASVAB 8/9/10
FY82 ASVAB 8/9/10
FY81 ASVAB 6/7, PI Deleted

APTITUDE AREA MM

Figure 6




APTITUDE AREA SC
Figure 7
ASVAB SCORE
110
100
90
80
70
60
50
40
10
20
30
40
50
60
70
80
90
100
PERCENT
CY76-78 ASVAB 6/7
FY81 ASVAB 6/7
FY81 ASVAB 8/9/10
FY82 ASVAB 8/9/10
FY81 ASVAB 6/7, PI Deleted

APTITUDE AREA CO

Figure 8

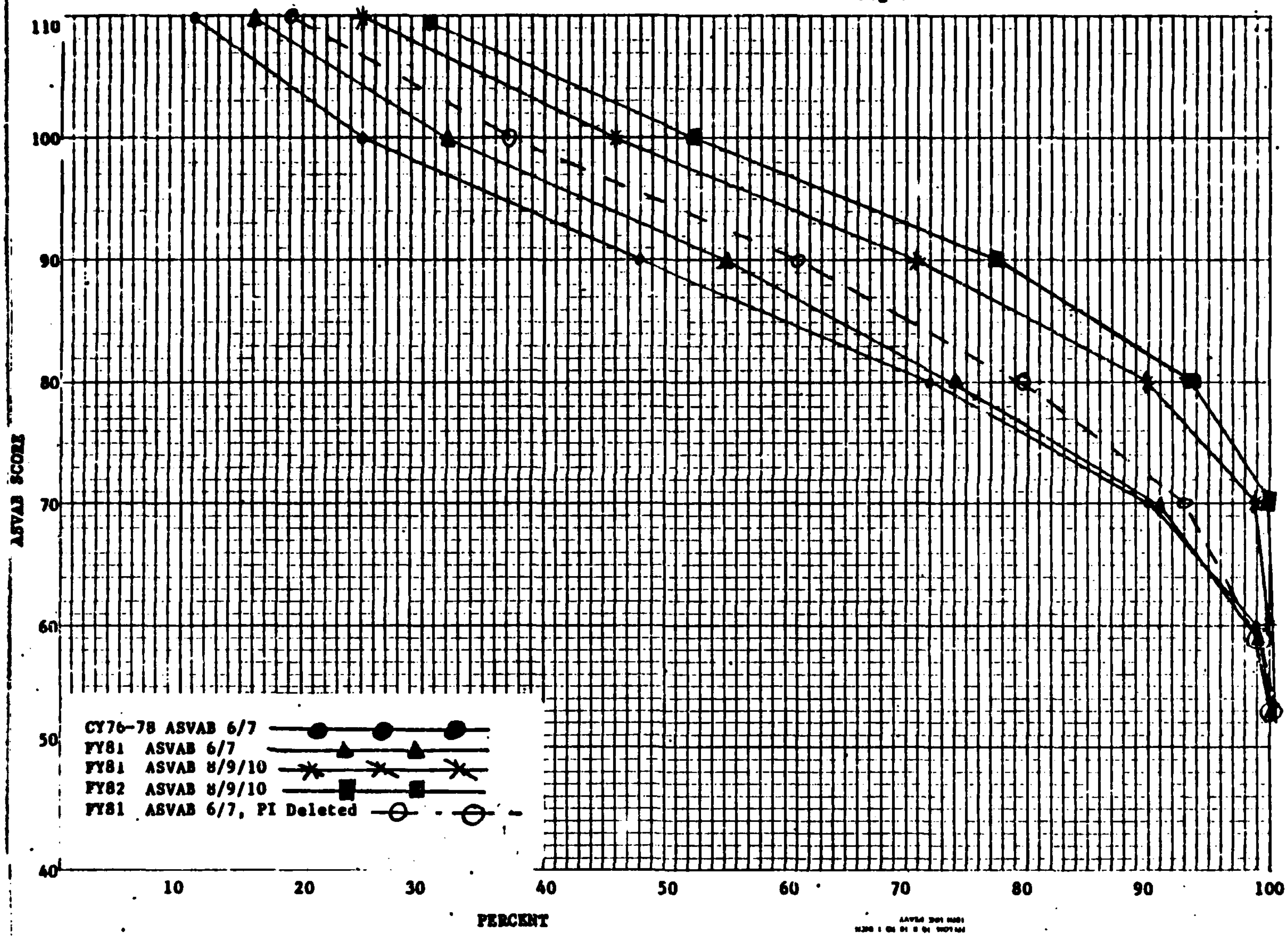
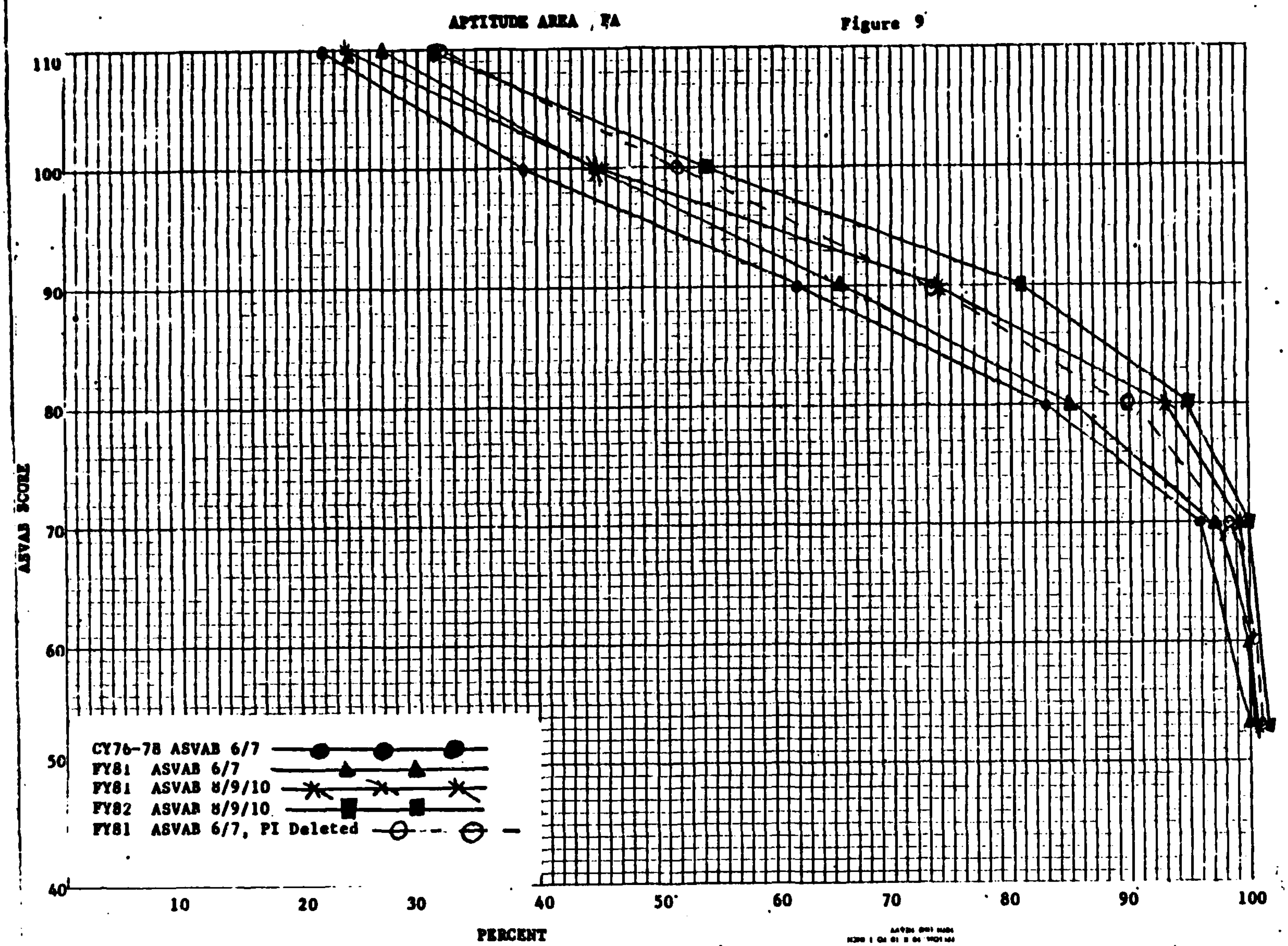
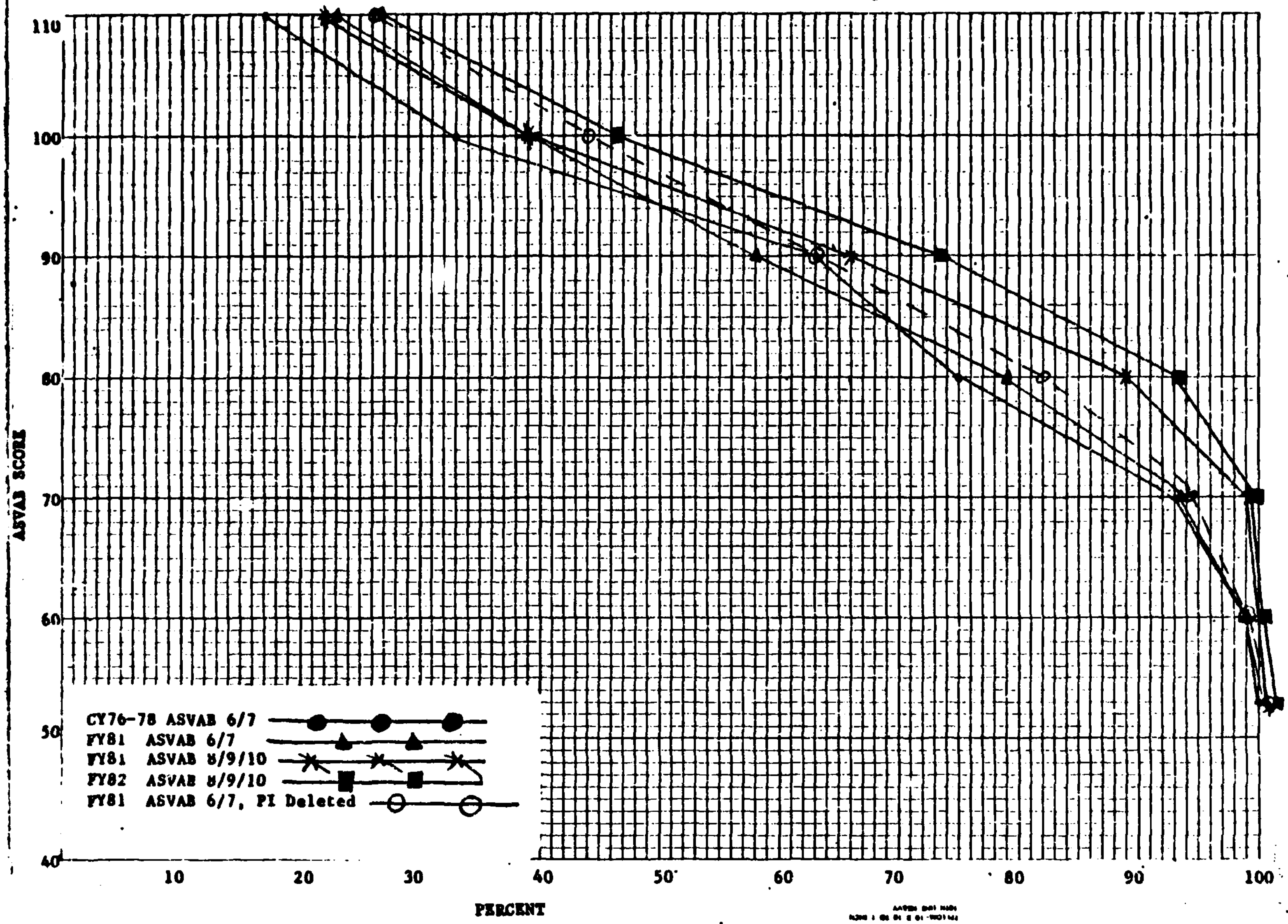


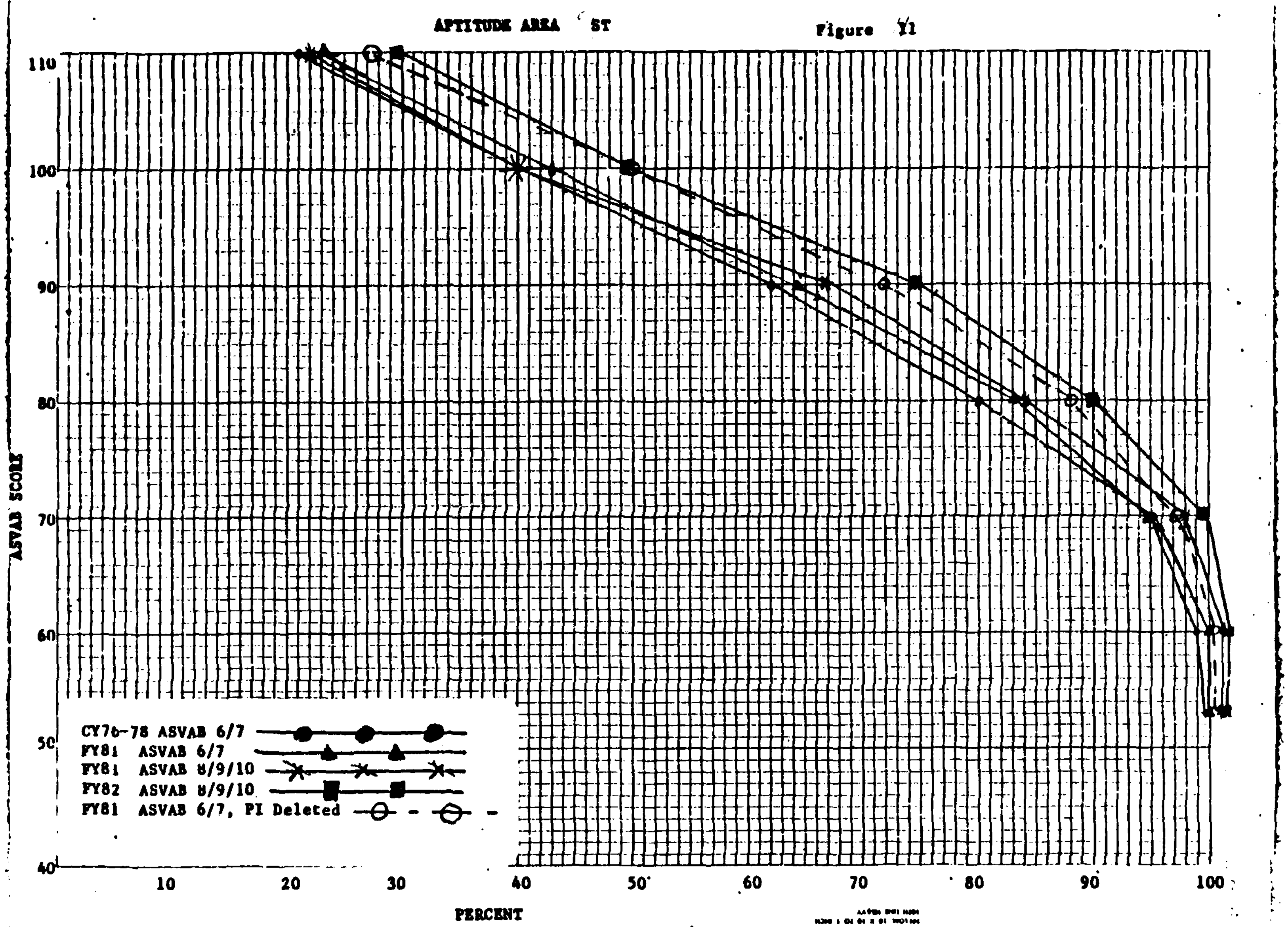
Figure 9


APTITUDE AREA , FA
110
100
90
80
70
60
50
40
ASVAB SCORE
10
20
30
40
50
60
70
80
90
100
PERCENT
CY76-78 ASVAB 6/7
FY81 ASVAB 6/7
FY81 ASVAB 8/9/10
FY82 ASVAB 8/9/10
FY81 ASVAB 6/7, PI Deleted

APTITUDE AREA . OF

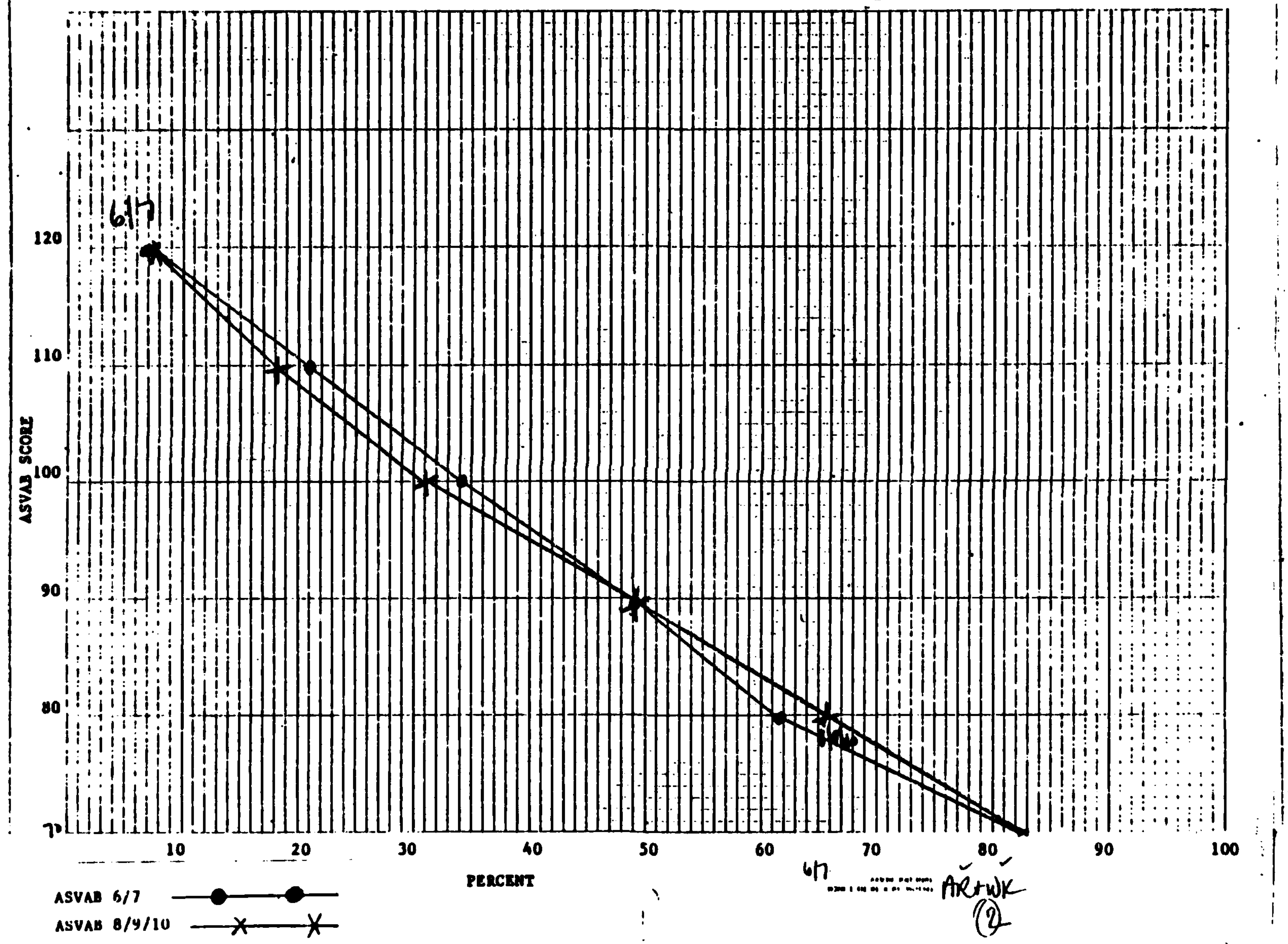
Figure ,10



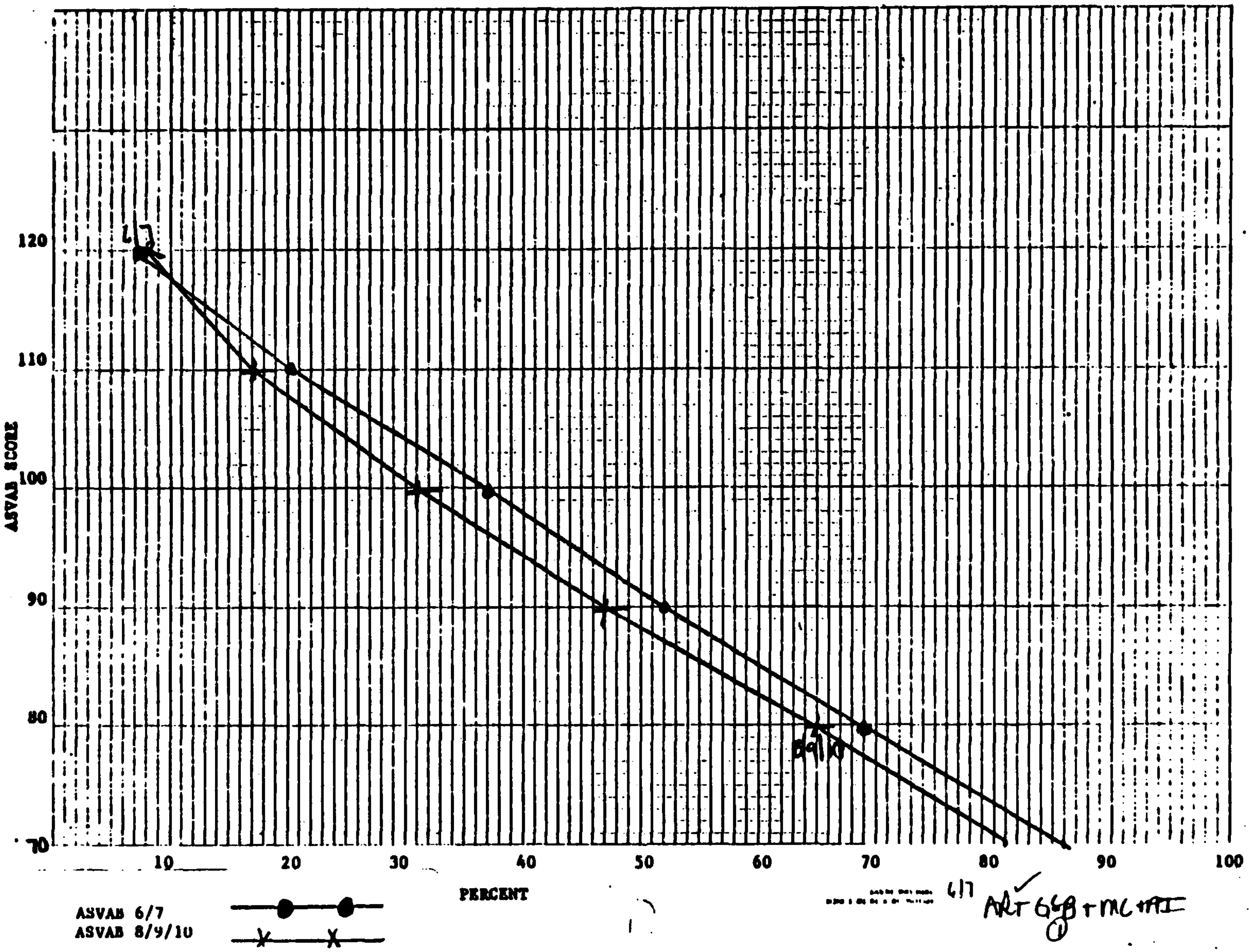

APTITUDE AREA ST
Figure Y1
ASVAB SCORE
PERCENT
110
100
90
80
70
60
50
40
10
20
30
40
50
60
70
80
90
100
CY76-78 ASVAB 6/7
FY81 ASVAB 6/7
FY81 ASVAB 8/9/10
FY82 ASVAB 8/9/10
FY81 ASVAB 6/7, PI Deleted

APTITUDE AREA ST

Figure 12

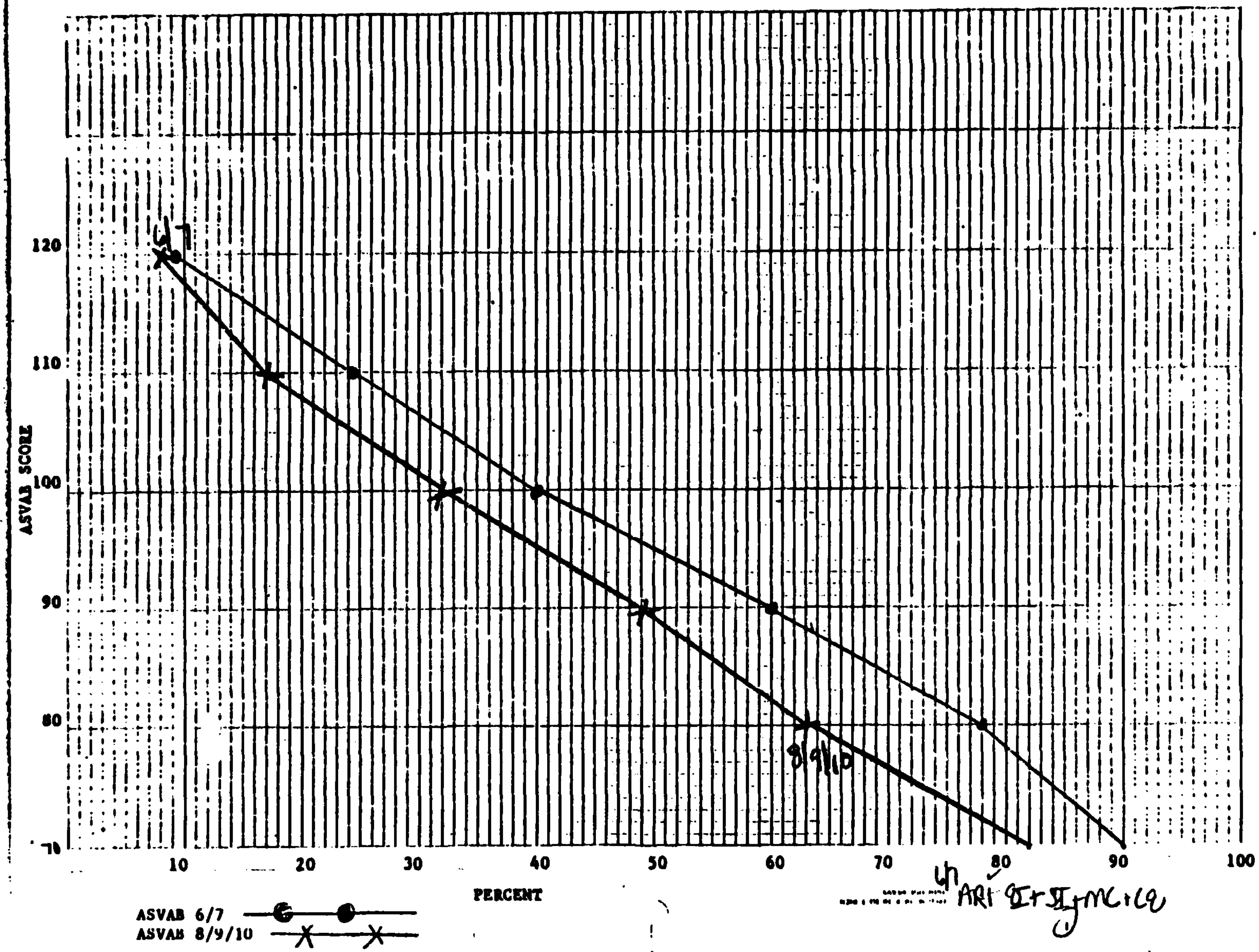


APTITUDE AREA GM Figure 13



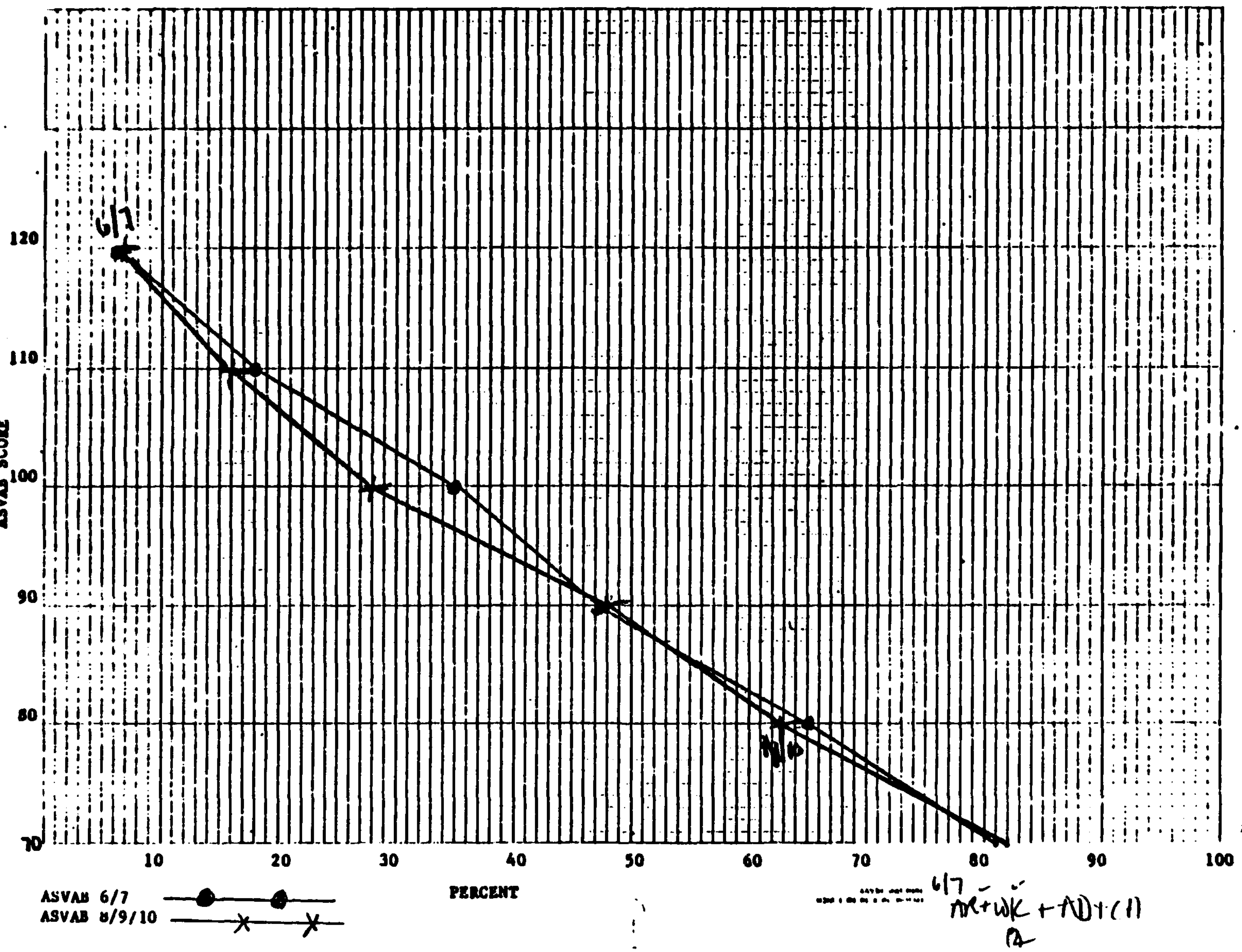
APTITUDE AREA EL

Figure 14



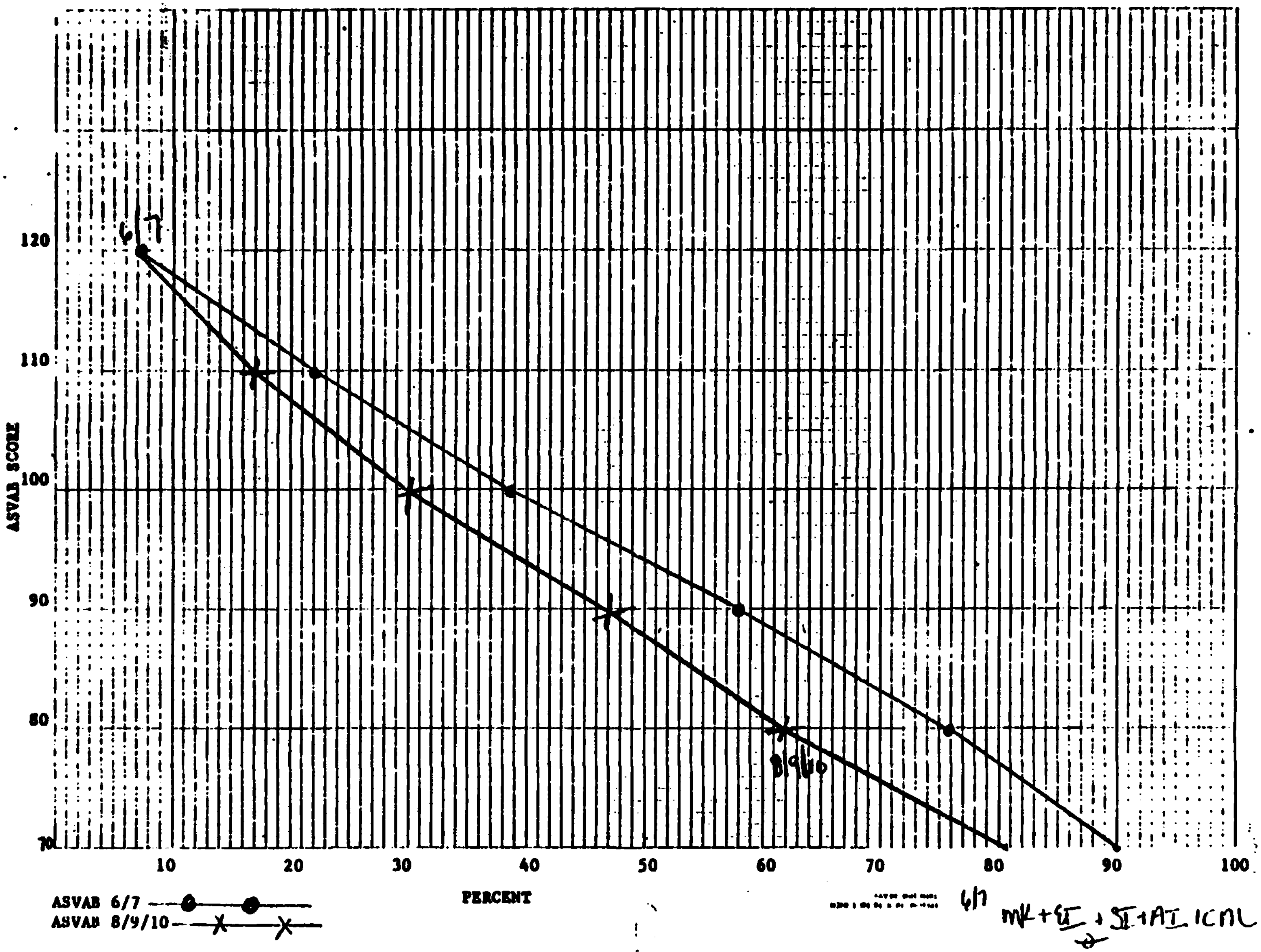
APTITUDE AREA CL

Figure 15



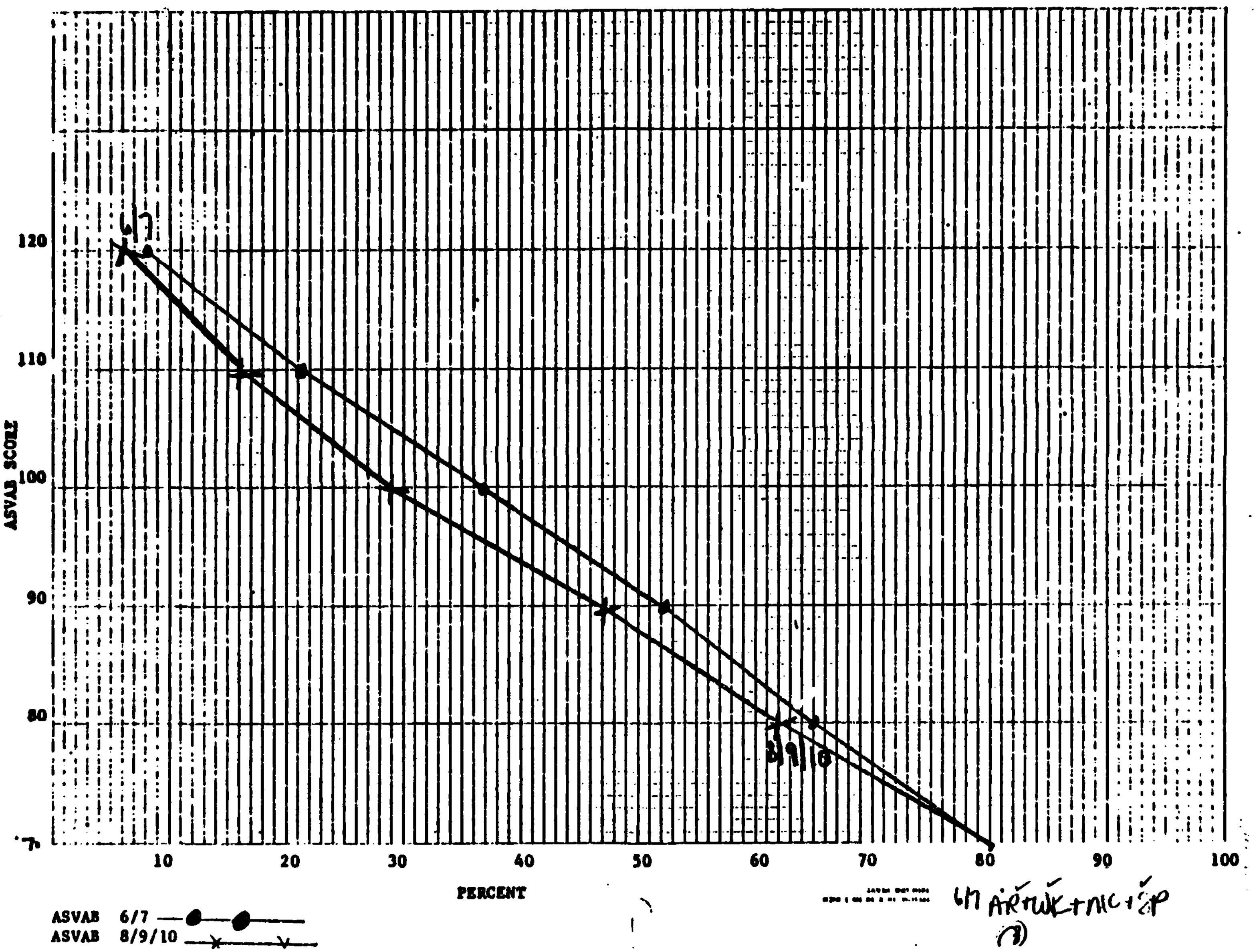
APTITUDE AREA MM

Figure 16



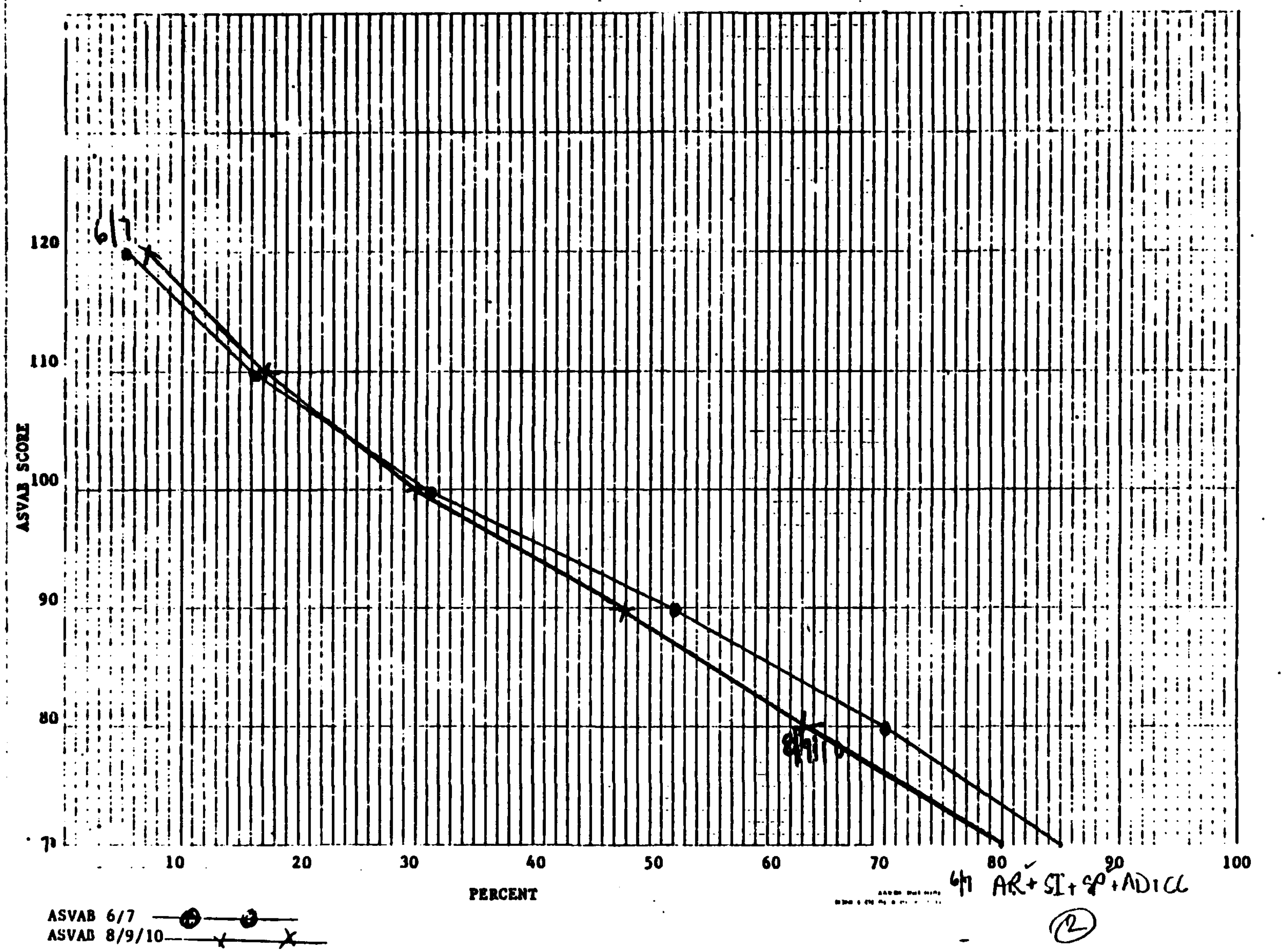
APTITUDE AREA SC

Figure 17



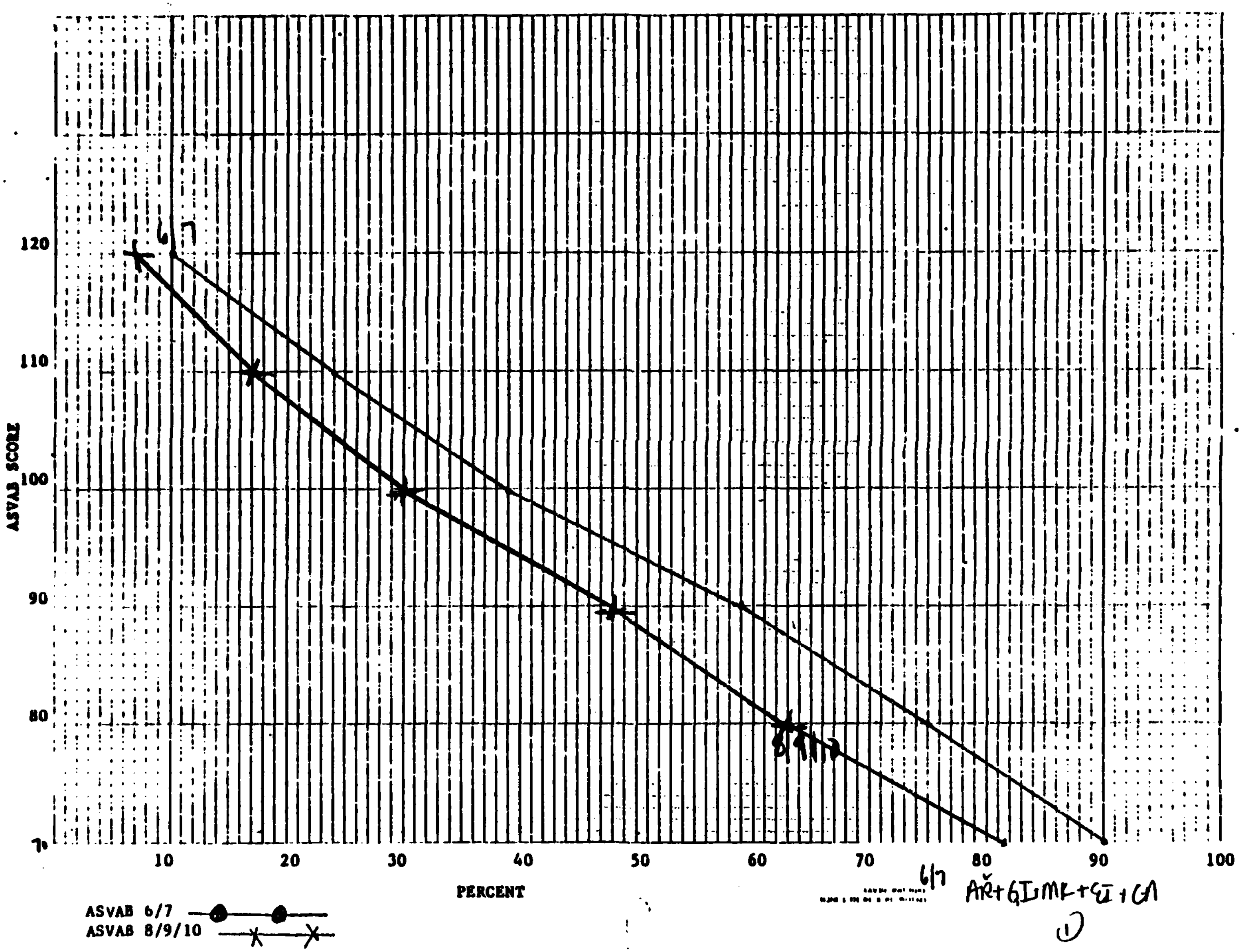
APTITUDE AREA CO

Figure 18



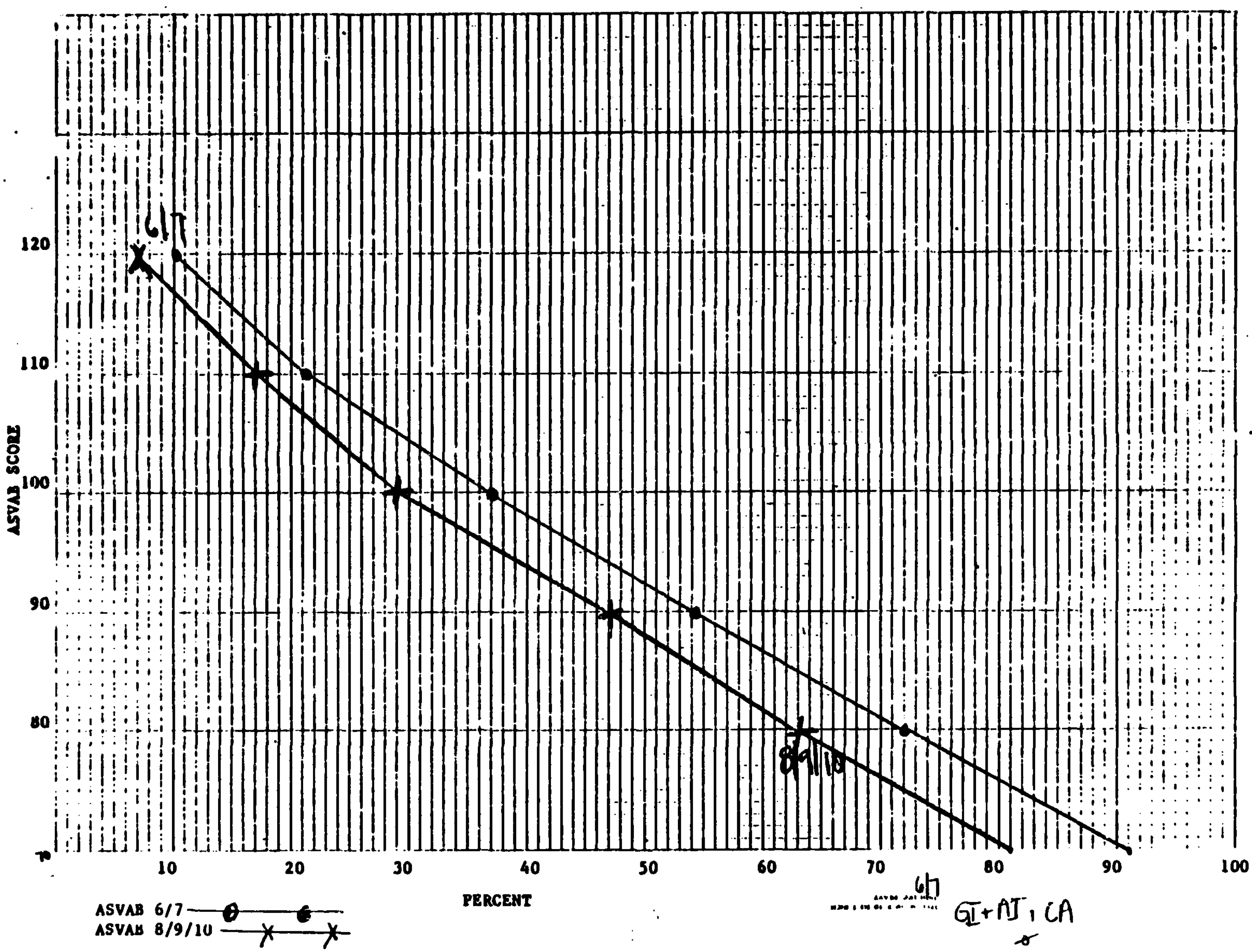
APTITUDE AREA FA

Figure 19



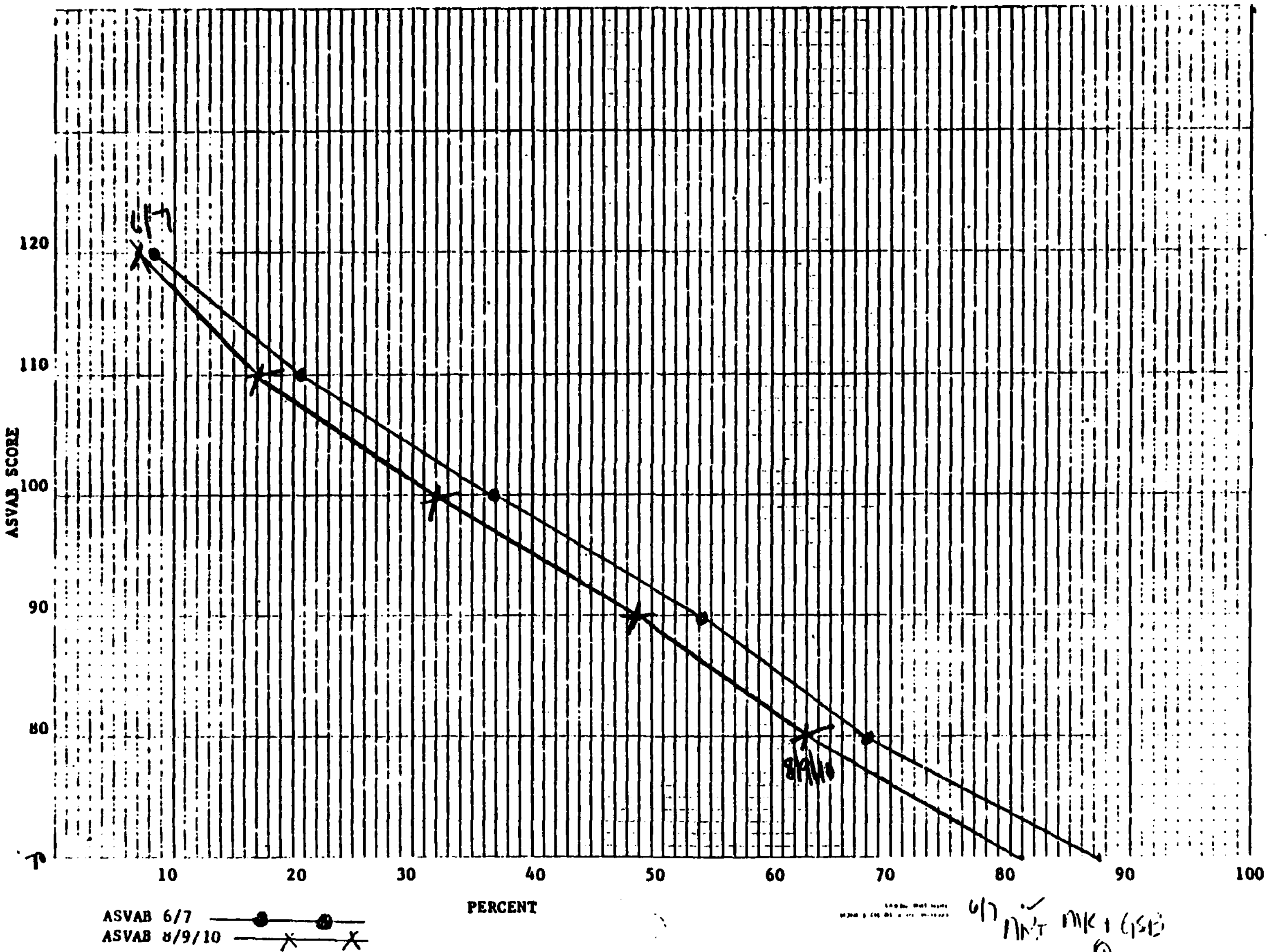
APTITUDE AREA . OF

Figure 20



APTITUDE AREA ST

Figure 21



VERBAL INFORMATION PROCESSING PARADIGMS: *

A REVIEW OF THEORY AND METHODS

Karen J. Mitchell
Army Research Institute for the
Behavioral and Social Sciences*

*5001 Eisenhower Avenue
Alexandria, Virginia 22333

Review of Verbal Processing Theory

* To be published as an ARI Technical Report.

VERBAL INFORMATION PROCESSING PARADIGMS:

A REVIEW OF THEORY AND METHODS

Recent advances in cognitive psychology have resulted in methods for identifying the cognitive processing operations, memory stores, and strategies involved in performance on test items and test-like tasks. The cognitive processing operations, stores, and strategies that cognitive psychologists examine represent psychological constructs which may contribute to both item and subject differences in observed performance.

Current attempts at applying theories of human cognition to analyses of performance on cognitive tasks range from broad analyses of a number of tasks to specific and detailed models for performance on a single task type. The methods for analysis, similarly, range from intuitive analyses of performance to computer simulation of human protocols and mathematical modeling of response time and response accuracy. Investigations of the algorithms and heuristics people use in processing information have focused on very simple cognitive tasks such as, deciding whether or not two visually presented letters are the same or different, to complex cognitive activities like reading text and solving algebra word problems. The ability components employed by these models, likewise, span a wide range.

Recent attempts at applying theories of human cognition to analyses of performance differences on test items suggest general dimensions along which differences are manifested. Investigations of cognitive ability components relevant to performance on test items and test-like tasks have focused on verbal and imaginal encoding; retrieval; code access; categorization; executive control; rule induction; inference; semantic, procedural and strategic knowledge; memory span; spatial visualization; etc. Various test-like tasks have been examined; these include tasks involving verbal analogy processing (Sternberg, 1977a, 1977b, 1980; Gentile et al., 1977; Pellegrino & Glaser, 1979, 1980; Barnes & Whitely, in press; Whitely & Barnes, 1979; Whitely & Schneider, 1980), geometric analogy processing (Mullholland, Pellegrino, & Glaser, 1980; Whitely & Schneider, 1981), series completion (Pellegrino & Glaser, 1980), syllogistic reasoning and transitive inference (Falmagne, 1975; Sternberg, 1979 1980; Sternberg et al., 1980; Sternberg & Weil, 1980), spatial rotation and visual comparison (Egan, 1979; Cooper, 1980), block design problems (Royer, 1977), matrix pattern abstraction (Hunt, 1974), and comprehension of text (Frederiksen, 1978, 1980).

A brief summary of the theory and methods of cognitive processing paradigms relevant to the analysis of performance on verbal tasks follows. Processing operations, strategies, and structures with a history of empirical and theoretical support are presented. Relevant measurement methodologies and analytic techniques are discussed.

THE FACTOR ANALYTIC APPROACH TO THE EXAMINATION OF VERBAL PERFORMANCE

The structure of cognitive ability as it relates to performance on tests and test-like tasks has traditionally been examined using factor analytic and related methodologies. The research pioneers of the factor analytic movement—Spearman, Thomson, Thurstone—have paved the way for the use of solid, empirically based analyses in aptitude test construction and validation. Factorial methods have been developed, according to Thurstone (1947, p. 55), for the purpose of " . . . identifying the principal dimensions or categories of mentality". Guilford (1967, p. 41) describes the goal of factor analysis as the identification of " . . . an underlying latent variable along which individuals differ".

Factor analytic methods can be employed at the stages of both hypothesis formulation and hypothesis testing. In the first instance, factor analysis serves as a useful exploratory technique. It allows analysts to derive a " . . . crude first map of a new domain" (Thurstone, 1947, p. 56). Exploratory factor analytic examinations of items, subtests or intact tests then allow one to proceed beyond initial stages to more direct forms of psychological experimentation in the laboratory. In the second instance, factor analysis can be employed as a " . . . method of comparing, confirming, or refuting alternative hypotheses initially suggested by nonstatistical arguments or evidence" (Burt, 1970, p. 17).

Factor analysis begins with a matrix of intercorrelations among items, subtests, or tests and distributes variance within variables and between factors in such a way that a set of underlying hypothetical performance constructs are suggested. The output from the factor analysis is a factor structure or factor pattern matrix of correlation coefficients between each variable, item or test, and each underlying factor. The factor matrix for a given factor analysis is typically rotated to some mathematically permissable coordinate reference system to facilitate interpretation.

The factor analytic definition and measurement of verbal ability is well documented. Verbal comprehension, defined as " . . . the ability to understand the English language," is referenced in at least 125 published studies (Ekstrom, French & Harman, 1976, p. 163). Tests of vocabulary, similarities, opposites, verbal analogies, proverbs, quotations, grammar, spelling, and paragraph and reading comprehension have loaded highly on the verbal factor in a number of studies. Verbal factors have been variously titled word knowledge, word fluency, verbal reasoning, cognition of semantic units, and cognition of semantic relations or systems. The basic distinctions between factor types may be summarized as follows: word fluency is typically characterized by tests dealing with single and isolated words; facility with grapheme or phoneme relations rather than semantic knowledge is tapped. The word knowledge factor taps semantic knowledge; it seems to reflect experience with and knowledge of the English language. The verbal reasoning factor may be seen as reasoning in reading or the ability to see relationships among ideas, to draw inferences from a paragraph or derive the principal thought or idea from a passage.

For the purposes of studying cognitive ability component differences relevant to performance on standardized tasks, factor analytic methods can play a promising role in theory-oriented research. Factor analyses of item or subtest data can be used to confirm or refute theory-oriented characterizations of processing requirements for items and tests. Alternatively, theory-oriented characterizations of cognitive processing operations, stores, and strategies underlying test item performance can be detailed and " . . . the role of various processes in a total matrix of cognitive operations, " can be identified (Carroll, 1976, p. 41). An example of the application of factor analytic methods to the examination of human information processing is seen in the work of John Carroll; a description of his work is included below.

THE RATIONAL ANALYSIS APPROACH TO THE EXAMINATION OF VERBAL PERFORMANCE

Using Hunt's Distributive Memory Model (discussed more fully in a later section), Carroll has attempted a rational and empirical analysis of the necessary and sufficient cognitive processing operations, stores, and strategies underlying performance on the 74 psychometric tests of the French, Ekstrom and Price (1973) Kit of Reference Tests for Cognitive Factors. Research suggests that the French Kit contains good marker tests for 24 different cognitive ability factors.

The memory model employed by Carroll depicts the processing of language messages in stages. The first stage is a decoding stage, in which arbitrary physical patterns are recognized as representations of language concepts. The second stage is an active memory stage, in which the recognized lexical items are rearranged in memory until they form a coherent linguistic structure. The third stage is a sentence producing stage, in which the semantic meaning of the linguistic structure is extracted and incorporated into our knowledge of the current situation. In the fourth stage, the current situation itself is analyzed with respect to information held in long term memory, and if appropriate, a response is chosen and emitted. A control or executive system directs the flow of information in the processing system and has access to the various levels of memory storage (Hunt, 1971, 1973, 1980).

Carroll has developed a uniform system for classifying the characteristics of the tasks represented by the items of each test in the French Kit. The classificatory scheme addresses the types of stimuli presented, the kinds of overt responses that are required to demonstrate performance, the sequencing of subtasks within the tasks, the cognitive processing operations, stores, and strategies that Carroll, conceiving himself as a subject, would employ in performing the test tasks, and the probable ranges of relevant temporal parameters. The scheme considers the term and contents of memory that would probably be addressed in storage, search, and retrieval operations.

The cognitive processing operations and strategies outlined by the system (Carroll, 1976) are processes that are explicitly specified or implied in task instructions and that are necessary to successful completion of the task. These operations and strategies are of three types: attentional, memorial, and executive. The latter two are further subdivided as follows—there are three kinds of memorial operations storing, searching, retrieving. Executive strategies are exemplified by such things as simple judgements of stimulus attributes such as to reveal identity, similarity, or comparison between two stimuli; manipulations of memorial contents, such as 'mentally rotating' a visiospatial configuration; and information transformations that produce new elements from combinations, reductions, etc. of old elements. In all, twenty different operations and strategies are outlined; they are:

1. Identify, recognize, interpret stimulus
2. Educe identities or similarities between two or more stimuli
3. Retrieve name, description, or instance from memory
4. Store item in memory
5. Retrieve associations, or general information, from memory
6. Retrieve or construct hypotheses
7. Examine different portions of memory
8. Perform serial operations with data from memory
9. Record intermediate result
10. Conduct visual inspection of stimulus
11. Reinterpret possible ambiguous item
12. Image, imagine or form abstract representation of stimulus

13. Mentally rotate spatial configuration
14. Comprehend and analyze language stimulus
15. Judge stimulus with respect to a specified characteristic
16. Ignore irrelevant stimuli
17. Use a special mnemonic aid
18. Rehearse associations
19. Develop a special search strategy
20. Chunk or group stimuli or data from memory (Carroll, 1976, p. 39).

Carroll has used his classification scheme as a basis for specifying the potential sources of individual differences underlying each of the 24 French Kit cognitive ability factors. He postulates that individual differences might arise through: (1) differences in the composition and ordering of processing operations and execution rules incorporated in the system; (2) differences in the temporal parameters associated with those execution rules; (3) differences in the processing capacity of the executive and its associated memory stores; and (4) differences in the contents of long term or permanent memory stores.

Carroll has found that nearly all pairs of tests from the same factor have one or more classification codes in common and that the patterns of the codes are generally distinct from factor to factor. The cognitive processing operations, stores, and strategies identified as being characteristic of the 24 factors and the tests that represent them are quite diverse with respect to type of operation and memory store involved, temporal parameters and other details. A description of the factors defined by verbal tasks and the cognitive processing operations, stores, and strategies characterizing these factors includes:

1. Factor FW (Word Fluency) is the facility to produce words that fit one or more structural, phonetic or orthographic restrictions unrelated to the meaning of words. The cognitive processing operation involved in word fluency is a search of a "lexiographemic" portion of long term memory for instances fitting the orthographic requirements. Strategies may include the use of an alphabetic mnemonic to search the memory systematically. French Kit tests loading on this factor are: (a) Word Endings test, where the task is to write as many words as possible ending with certain given letters, (b) Word Beginnnings test, where the task is to write as many words as possible beginning with given certain letters, and (c) Word Beginnings and Endings test, where the task is to write as many words as possible beginning with one given letter and ending with another.

2. Factor FA (Associational Fluency) is the ability to produce rapidly words which share a given area of meaning or some other common semantic property. Associational fluency entails a search of a major portion of a long term lexicosemantic store, with special attention to semantic or associational aspects. Strategies may involve searching long term memory for different meanings of the stimulus word. Tests loading on this factor are: (a) Controlled Associations test, where the task is to write as many synonyms as possible for each of four words, (b) Opposites test, where the subject is asked to write up to six antonyms for each of four words, and (c) Figures of Speech, where the subject is asked to provide up to three words or phrases to complete each of five figures of speech.

3. Factor FE (Fluency of Expression) is the ability to think rapidly of word groups or phrases. Like associational fluency, expressional fluency involves a search of lexiosemantic memory but with special attention to grammatical features of lexical items and different syntactical patterns of phrases and sentences. Cognitive processing strategies may involve the use of grammatical mnemonics such as considering grammatical classification in the search for words. Tests loading on this factor are: (a) Making Sentences, where the subject is asked to make sentences of a specified length when the initial letter of some of the words is provided, (b) Arranging Words, where the subject is asked to write up to twenty different sentences using the same four words, and (c) Rewriting, where the subject is asked to rewrite each of three sentences in two different ways.

4. Factor V (Verbal Comprehension) is the ability to understand English words. This factor is almost exclusively dependent on the contents of lexicosemantic long term memory store, ie., upon the probability that the subject can retrieve the correct meaning of a word. Tests loading on this factor are: (a) Vocabulary I, a four choice synonym test, (b) Vocabulary II, a five choice synonym test, (c) Extended Range Vocabulary test, a five choice synonym test having items ranging from very easy to very difficult, (d) Advanced Vocabulary I, a five choice synonym test consisting mainly of difficult items, and (e) Advanced Vocabulary II, a four choice vocabulary test consisting mainly of difficult items. Carroll states that a more diversified set of tests of this factor would probably call on other aspects of the lexiosemantic store, particularly on the grammatical feature portions.

THE INFORMATION PROCESSING APPROACH TO THE EXAMINATION OF VERBAL PERFORMANCE

Although each of the paradigms discussed here may be thought of as information processing paradigms, a more general discussion of verbal information processing may be informative. The information processing viewpoint holds that performance on cognitive tasks may be described by the operation of integrated programs for the processing of information available from sensory channels and memory stores. The paradigm poses that the presentation of stimuli initiates a sequence of processing stages. Each stage operates on the information available to it. The operations transform the information in some manner; furthermore, these operations take a measurable amount of time. The output of each processing stage is in the form of transformed information, and this new information is the input to the succeeding stage.

The operations, stores, and strategies of the human information processing system are usually described as analogues to computer system structures. The cognitive ability components are used to formulate information processing models of tasks. The major concern of information processing research is to identify cognitive processing operations, stores, and strategies and to determine how they operate.

Most general models of the human information processing system include a short term sensory storage or buffer component, a memory component consisting of two or three subsystems distinguished by relative time duration of information storage—short term, intermediate term or working memory and long term storage; a response selection or generation component; and a central or executive processor. There is much less unanimity in the literature with respect to the cognitive processing operations and strategies.

Rose (1980) describes a number of cognitive processing operations which have a history of empirical and theoretical support. His compendium includes:

1. Encoding, the operation by which information is input into the processing system, including the initial set of operations that converts the physical stimulus to a form that is appropriate for the task. Different task demands may require different levels of analysis of the stimulus. Posner (1969) has called this dimension 'abstraction,' the operation by which different types of information about the stimulus are extracted.

2. Constructing, the operation by which new information structures are generated from information already in the processing system. This is what Neisser (1967) and others have called 'synthesis.'

3. Transforming, the operation by which a given information structure is converted into an equivalent structure necessary for task performance. In contrast to constructing, transformations do not involve any new information abstraction; rather, this operation requires the application of some stored rules to the information structure already present.

4. Storing, the operation by which new information is incorporated into existing information structures while its entire content is retained.

5. Retrieving, the operation by which previously stored information is made available to the processing system.

6. Searching, the operation by which an information structure is examined for the presence or absence of one or more properties. The information structure examined may be one already in the processing system or one external to it.

7. Comparing, the operation by which two information structures, either internal or external to the processing system, are judged to be the same or different. The structures need not both be physical entities, for example, a physical entity may be compared to a stored representation or description in order to determine identity.

8. Responding, the operation by which the appropriate motor action is selected and executed.

Newell and Simon (1972) and Simon (1978) have shown that systems of cognitive ability components can be depicted by computer simulations of complex problem solving activities, such as the solving of chess or symbolic logic problems. They have determined the cognitive processing operations, stores, and strategies necessary for a computer program to extrapolate sequential material such as number or letter sequences, to translate and solve algebra word problems, the Tower of Hanoi puzzle, perception in chess, to understand task instructions, and to spell English words. A running program serves as a built-in empirical test via computer modeling.

THE COGNITIVE CORRELATES APPROACH TO THE EXAMINATION OF VERBAL PERFORMANCE

What Pellegrino and Glaser (1979) have referred to as the Cognitive Correlates approach to individual differences can be traced back to the work of Hunt, Frost and Lunneborg (1973). This line of research has been continued by Hunt and his associates throughout the decade (Hunt, 1976; Hunt, Lunneborg & Lewis, 1975; Hunt & Lansman, 1975; Hunt, 1976, 1978). The basic premise behind their work is that examinee performance on relatively simple laboratory tasks can be used to identify cognitive ability components underlying performance on complex cognitive tasks. Hunt and his collegues examine tasks that are theoretically related to performance on verbal information processing problems in order to determine how behavior on these tasks is related to performance on verbal aptitude tests. The goal of this approach is to specify the cognitive ability components that are differentially related to high and low levels of verbal competence.

Hunt and his colleagues posit that there are two types of cognitive ability components underlying verbal performance. The first set of components is based on semantic knowledge, on the ability to deal with words and the concepts they represent. The second set of components is based on strategic knowledge, on the exercise of information-free, mechanistic operations. These operations dictate the transformation of both the internal and external physical representation of a symbol; these strategic knowledge operations do not depend on information associated with the symbol. They are the means by which information structures are transformed into equivalent structures necessary for task performance. Hunt and his colleagues propose that effectiveness in verbal information processing depends on the relation of the stimulus information to the information structures stored in semantic memory, on the way the information is organized, and on the manipulative efficiency of the mechanistic processes.

As is typical of most modern theories of cognition, the model employed by Hunt and his colleagues draws a distinction between two types of memory. The first is a relatively small active memory and the second is a theoretically infinite long-term memory. Long term memory may be thought of as a collection of basic memory units or engrams in conjunction with the associations that define them. The engrams, collectively, represent the semantic knowledge information structures and the mechanistic, information-free structures. Verbal information processing takes place when active memory images, aroused partly by the recognition of current input and partly by the recognition of the previous state of active memory, are supplemented by semantic knowledge information structures and transformed by processes controlled both by sensory input and by the arousal of strategic knowledge-based structure rules stored in long-term memory. This model is Hunt's Distributive Memory Model discussed above.

With this model as a frame of reference, Hunt and his colleagues ask questions about differences between examinees representing a wide range of verbal competence. Tests of verbal ability, composites that test knowledge about syntax, spelling, vocabulary, and the ability to comprehend brief statements, are administered to subjects and the data are used to identify ability subgroups. Subgroup performance is then compared on laboratory tasks which have cognitive ability component characteristics defined by prior investigations.

Hunt and his colleagues have conducted laboratory experiments to examine individual cognitive ability component differences in (1) lexical recognition, arousal speed; (2) speed of information manipulation in short term memory; (3) storage differences in short and intermediate term memory; (4) speed of information transmission from place to place in the total system; (5) programming which shifts the burden of information processing from one component of the memory system to another; and (6) attention allocation (Hunt, 1976, 1980). Hunt and his colleagues have observed individual differences in these processes within the population represented by university students and within a population of somewhat lower than average ability. These differences appear to account for a moderate portion of individual variation in verbal competence.

In their examination of the decoding operation, for example, they have found evidence for a clear association between verbal competence and the simple act of identifying highly overlearned symbols. Their research in this area has relied primarily on the letter identification task developed by Posner and Mitchell (1967). In the Posner task, two letters are presented simultaneously on a visual display screen and the subject's task is to indicate whether the letters are the same or different. Under physical identity (PI) instructions, letter pairs are to be identified as 'same' only if the letters are exact duplicates of each other, as in the pair (AA). Under name identity (NI) instructions, letter pairs are to be called 'same' if they are different visual codes for the same letter, as in the pair (Aa). The difference between reaction time to classify an item as same under name identity instructions and the time to classify an item as same under physical identity instructions is assumed to reflect the extra processing operations required to carry the analysis to the same level. Hunt cites moderate negative correlations for the difference measure and verbal aptitude. Low verbal subjects are seen to have high difference scores and high verbal subjects are seen to have low difference indices.

Hunt, Lunneborg and Lewis (1975) have examined the active memory capacities of high verbal and low verbal college students using a version of the Brown-Peterson short term memory paradigm. In this procedure the subject is shown four letters, asked to repeat a string of digits presented visually, and finally to recall the four letters. A positive correlation is observed between examinee behavior on the task and verbal aptitude test performance. High verbal students are seen to code items more rapidly than low verbal students and high verbals have a lower relative error frequency. Hunt and his colleagues postulate that the observed differences are associated with language competence. Greater short term memory capacity, they say, may indicate an increase in the strategies that a high verbal individual can use in verbal problem solving..

Hunt (1978) has also examined the relationship between sentence verification reaction time and measures of verbal aptitude. The task, developed by Clark and Chase, is designed to assess how subjects compare information from various sources in order to verify sentences. Subjects are presented with a display containing a sentence and a picture. The subject is asked to determine whether the sentence is an accurate or inaccurate representation of the picture. The display sentences are of the form:

'Star (plus) is (is not) above (below) plus (star).'

Sentences can be either positively worded or negatively worded. They can be either true or false representations of the displays:

+ *

* +

There are four possible sentence combinations: A true affirmative description, a true negative description, a false affirmative description, and a false negative description. The dependent variable is the latency of the subject's judgements.

It is hypothesized that subjects will first encode the sentence and picture, then perform transformations based on the modifiers in the presented sentences, e.g., transform 'below' to 'not above,' and finally compare the sentence and picture. The differences in processing time for the four conditions are assumed to reflect the number of transformations that must be executed to process the sentence as well as the complexity of the comparison of the verbal and pictorial representations. Findings are seen to be consistent with this hypothesis. A negative correlation is evidenced for response latency and verbal competence.

THE COMPONENTIAL ANALYSIS APPROACH TO THE EXAMINATION OF VERBAL PERFORMANCE

Sternberg says that his (1977a, 1977b, 1979) componential theory of cognitive performance is directed at the analysis of performance differences in the elementary operations involved in task performance, in the strategies for task performance into which processes combine, and in the representations of information upon which the operations and strategies act. A component is defined as an elementary information process that operates upon internal representations of objects or symbols (Newell & Simon, 1972). A component may translate sensory input into a conceptual representation, transform one conceptual representation into another, or translate a conceptual representation into motor input. Components are classified by level of generality and function.

Sternberg describes general components, for example, encoding and response components, as prerequisite to successful performance of all tasks of a global task type, e.g., reasoning tasks. Class components, such as inference, mapping relations, or applying relations, are components common to a particular class of tasks, for example, inductive reasoning tasks. Specific components are required for performance of single tasks within a task universe.

Components perform five different kinds of functions. Metacomponents are higher-order control processes used for planning how a problem should be solved, for making decisions regarding alternate courses of action during problem solving, and for monitoring solution processes. These are analygous to the executive or control subsystems discussed above. Performance components are processes that are used in the execution of a problem solving strategy. Acquisition components are processes used in learning new information. Retention components are processes used in retrieving previously stored knowledge. Transfer components are used in carrying knowledge over from one task or task context to another.

Componential analysis defines information processing models of performance that specify: (1) the nature and order of component process execution, and (2) the mode of component execution, that is, whether components are executed serially or in parallel, as self terminating or exhaustive processes, holistically or particularistically. Cognitive tasks may be decomposed using the methods of partial tasks, stem splitting, systematically varied booklets, and the method of complete tasks. The method of stem splitting is discussed for illustrative purposes.

The method of stem splitting involves items requiring the same number and type of information processing components, but with different numbers of executions of the various components. The method of stem splitting applied to a verbal analogy task, for example, might take the following form:

1. red : blood :: white : (a) color
   (b) snow

2. red : blood :: (a) white : snow
   (b) brown : color

3. red : (a) blood :: white : snow
   (b) brick :: brown : color

The first item involves the encoding of five terms—red, blood, white, color, and snow; the inference of one relation, the color/substance relation; the mapping of one relation, the color/substance relation onto white and its alternatives; the application of two relations, the color/substance relation onto color and snow; and one response, b. The second item requires the encoding of six terms—red, blood, white, snow, brown and color; the inference of one relation, again, the color/substance relation; the mapping of two relations, the color/substance relation onto white and brown; the application of two relations, the color/substance relation onto white/snow and brown/color; and one response, a. The third item requires encoding of seven terms, inference of two relations, mapping of two relations, application of two relations, and one response. The primary dependent variable for the analogy task might be solution latency or response choice. Controls are introduced for requirement differences in the encoding process. The cognitive processing contributions of the inference, mapping, and application of relation components may then be individually examined. Experimental results suggest the psychological reality of each of the three components in verbal analogy processing. Solution latency increases with additional executions of the various components; response accuracy also decreases for the more complex items.

Sternberg has examined such task types as linear, categorical and conditional syllogisms, and verbal and schematic-picture analogies. Componential models accounting for as much as 96% of the variance in solution latency and response choice data have been constructed.

THE COGNITIVE PROCESS OUTCOME MODELING APPROACH TO THE EXAMINATION OF VERBAL PERFORMANCE

Also working in the area of componential analysis is Whitely (1980, 1981; Barnes & Whitely, in press; Whitely & Barnes, 1979; Whitely & Schneider, 1980, 1981) and her colleagues at the University of Kansas. These researchers have also examined cognitive processing operations, stores, and strategies in terms of performance on test items and test-like tasks. These researchers charge test developers to begin the test construction process by elaborating theories of item tasks. The theories can then be used as item specifications in test development. Cognitive process outcome models can be used to factor item and examinee response variance in accordance with componential theories of task performance. Multicomponent latent trait and linear logistic latent models can be employed to relate cognitive ability component performance to ability test outcomes.

These latent trait models assume that aptitude test items can be decomposed into subtasks that reflect an exhaustive set of cognitive processing components. Cognitive ability components are defined by item subtasks and/or stimulus information measures that is, by records of the cognitive processing operations, stores, and strategies purportedly involved in performance on a test item. The models specify both a mathematical model of item performance and a latent trait model for cognitive ability components. The latent trait models express the probability of success on each subtask as a logistic function of item difficulty and person ability on the underlying cognitive ability component. The mathematical model expresses the probability of success on the total item as the joint probability of passing the subtasks for each cognitive ability component. Models have been developed to estimate joint, conditional and marginal maximum likelihoods for the multicomponent and linear logistic latent trait models.

The linear logistic latent trait model will be discussed for the purposes of illustration. This technique models item difficulty from stimulus information. Maximum liklihood estimates of person ability and cognitive ability component contributions to item difficulty are generated. The linear logistic latent trait model is similar to the Rasch model in that only item difficulty is examined; no discrimination or guessing parameters are postulated. The following equation is the Rasch model for the probability that person j passes item i:

$$1. \quad P(x_{ij}=1) = \frac{\exp(\xi_{ij} - \sigma_i)}{1 + \exp(\xi_j - \sigma_i)}$$

where ξ_j =the ability level for person j, and

σ_i =the difficulty for item i.

The linear logistic latent trait model factors item difficulty into a subset of m stimulus complexity components according to the following linear model.

$$2. \quad \sigma'_i = \sum_m (f_{im}\,\eta_m) + \alpha$$

where f_{im} =the number of executions of component m that are involved in solving item i,

η_m =the difficulty of processing component m, and

α =a normalization constant.

The complete linear logistic model, then, is given:

$$3. \quad P_{(x_{ij}=1)} = \frac{\exp(\xi_{ij} - (\sum_m f_{im}\eta_m + \alpha))}{1 + \exp(\xi_{ij} - (\sum_m f_{im}\eta_m + \alpha))}$$

Maximum likelihood estimates are computed for the parameters; an estimate for the difficulty of each cognitive ability component is obtained along with standard errors and error correlations. Examinee item and component likelihoods are computed.

Cognitive process outcome analyses involve: (1) testing alternative component models of response accuracy, (2) decomposing item difficulty into cognitive ability component contributions, and (3) assessing the cognitive processing operation, store, and strategy person parameters as individual difference measures. Whitely and her colleagues have examined verbal and geometrical analogy tasks.

Performance on verbal analogy and verbal classification test items has been modeled by Whitely and her colleagues in terms of such cognitive processing operations as image construction and response evaluation. The image construction operation involves defining the attributes of the ideal solution to an item; response evaluation involves selecting the response alternative that best fulfills a given set of ideal solution attributes. Image construction is probably best regarded as an inductive reasoning operation since it involves constructing a general rule from particular stimuli. Response evaluation, on the other hand, involves deductive reasoning, since it depends on the evaluation of specifics according to a general rule.

In cognitive process outcome analyses, subjects might be given a verbal analogy item such as:

STEM tree : sap :: man :

ALTERNATIVES
1. axe
2. woman
3. maple
4. blood
5. arm.

For the image construction subtask the subject is asked to specify the rule or set of attributes for the ideal completion of the analogy item. For the response evaluation subtask, the subject is given an analogy 'image' and asked to select the response alternative that best fulfills the image. Verbal classification subtasks might be constructed in the same way as the analogy subtasks for the various cognitive ability components. Whitely's data support the feasibility of modeling response accuracy on verbal aptitude items from image construction and response evaluation operations. The inclusion of these subtasks account for from 70 to 83 percent of the variance in item performance.

THE CHRONOMETRIC ANALYSIS APPROACH TO THE EXAMINATION OF TEST ITEM PERFORMANCE

Frederiksen (1980) and his colleagues have sought to develop a series of cognitive ability component measures that are representative of the verbal information processing components involved in reading text. Their measures are designed to assess skills involved in the translation of letter patterns into sound patterns, in the recognition and encoding of patterns, in the retrieval of semantic information, and in the formulation of representations of text. The theoretical model that guides the selection of cognitive ability component measures for Frederiksen and his colleagues is defined by four levels of verbal information processing: (1) visual feature extraction, (2) perceptual encoding, (3) decoding and (4) lexical analysis. Visual feature extraction is the operation by which different types of information about the stimulus display are extracted. Perceptual encoding is the operation by which information is input into the system, and decoding is the operation by which arbitrary physical patterns are recognized as representations of grapheme and phoneme concepts in the lexicon.

In lexical analysis an attempt is made to match letter strings input in the preceeding stages to appropriate semantic referents. For phrase and sentence units, analysis is also directed at organizing these meaning elements into coherent text representations. Lexical, semantic, and syntactic knowledge is called upon in the identification of lexical items and in phrase and sentential analysis. The lexical analysis process may be either data driven or hypothesis driven. When lexical analysis is data driven, grapheme and phoneme data alone drive the analysis process. When lexical analysis is hypothesis driven, information available from the analysis of previous text supplements the data driven analysis process. Contextual information is encoded by the reader and serves to generate hypotheses about subsequent text. The reader may engage in an iterative process of discourse representation and revision.

Frederiksen and his colleagues propose that skilled readers are better able to execute cognitive processing operations, gain access to and search memory stores, and define processing strategies at all levels. An advantage in visual feature extraction, perceptual encoding, decoding, and data driven lexical access is hypothesized for skilled readers. It is hypothesized that skilled readers are better able to integrate information from perceptual and contextual sources in generating hypotheses about text and in gaining access to memory stores.

The chronometric analysis approach to the assessment of cognitive processing operations, stores, and strategies holds that the monitoring of processing time provides an important tool for the measurement of cognitive ability components. Chronometric analysis looks at reaction time differences for experimental conditions that vary the processing load placed on a single cognitive ability subsystem. The reaction time contrasts provide a measure of relative processing difficulty under the contrasted conditions.

Frederiksen and his coworkers have examined verbal performance using such tasks as the 'pseudoword decoding' task and the 'reading in context' task. In Frederiksen's pseudoword decoding task, subjects are asked to pronounce letter strings bearing a close resemblance to English forms. The letter strings represent a number of different variations, including variations in length, number of syllables, and type of vowel. The subject's reaction time from the presentation of the display to the onset of vocalization is the dependent variable. Increases in reaction time have been observed for each added letter in a letter string, for each added syllable, and for letter strings containing digraph rather than single vowels. The reaction time differences are assumed to be indicative of the additional processing time required to handle the more complex letter string forms.

The reading in context task centers on the use a subject makes of prior context in generating perceptual hypotheses in reading. The task presents the subject with a series of displays in three frames. The first frame contains an incomplete paragraph. The second frame is blank, and the third frame presents the final phrase of the passage. Subjects are presented with the context paragraph, they are instructed to read it at their own pace and press a response button when they have finished. The blank frame is then presented for 200 msc. The final passage phrase frame follows and is projected for 200 msc. The dependent variable is the number of words or word fragments reported correctly for the third frame. Subjects are presented with all three frames in one condition. In a second condition, subjects are presented only the second and third frames. Experimenters are able to assess the subject's use of context in generating and testing hypothesized word sequences by looking at visual span measurements, defined as the number of letter spaces from the leftmost correct reported letter to the rightmost correct letter. Increases in visual span have been observed for the condition where frame one provides prior context. The increase in visual span is assumed to be indicative of the use of information structures provided by prior context to construct hypotheses about subsequent text.

THE KINTSCH AND VAN DIJK PROSE PROCESSING APPROACH TO THE EXAMINATION OF VERBAL PERFORMANCE

The Kintsch and van Dijk (1978) prose processing model attempts to describe the system of mental operations that underlie text comprehension. The model is based on the premise that the comprehension act can be decomposed into component processess. The Kintsch and van Dijk prose processing model has its roots in the propositional theory outlined by Kintsch in The Representation of Meaning in Memory (1974). The scheme is further explained by Turner and Greene (1978); these authors also provide a step-by-step guide to propositionalizing text.

The Kintsch and van Dijk prose processing model is concerned primarily with semantic structures. A full grammar, necessary for both the interpretation and production of text, is not specified. The model operates at the level of assumed underlying semantic structures. The theory posits that comprehension involves knowledge use and inference construction. The model does not, however, specify the knowledge bases necessary for comprehension, nor does it discuss the processes involved in inference construction.

The Kintsch and van Dijk prose processing model represents textual information in terms of a text base. A text base is an ordered set of interrelated propositions depicting the underlying meaning of prose. Propositions are idea units; each proposition represents a single idea. A proposition consists of a relation (previously called a predicate) and one or more arguments. The relation connects sets of arguments to form an idea unit. The arguments are either concepts or propositions themselves. A concept is realized in language by a word or phrase. The words themselves are inconsequential. It is the abstract concepts they represent that are of interest.

Kintsch and van Dijk have adopted the convention of writing a proposition as follows:

(TRACK, ROCKET, RADAR)

The relation is TRACK. The first argument is ROCKET which functions in the semantic role of an object; and the second argument is RADAR which serves the semantic role of instrument of the action defined by the relation. The actual English text for this proposition might be expressed as: "The radar tracked the rocket," or The rocket was tracked by radar "

Propositions can be classified into three classes: Prediction, Modification and Connection. These classes are defined by the types of relations propositions contain. Relation types impose constraints on the classes of arguments that can be taken.

Predicate propositions express ideas of action or being. The relations are usually verbs. Arguments serve such semantic roles as agent, experiencer, instrument, object, source, or goal of the stated action. Nominal propositions, expressing set membership, and references may also be predicate propositions. A referential proposition is one which states that the referent of one argument is the same as that of a second argument. Propositions of reference are frequently implied.

Modifier propositions change a concept by restricting or limiting it by means of another concept. Four different types of modifiers are discussed: Qualifiers, Quantifiers, Partitives and Negations. These classes indicate the specific type of modification that is involved. Qualifier propositions limit or restric. the scope of an argument or proposition by expressing a quality or attribute of it. Quantifier propositions express definite or indefinite quantities. Partitive propositions indicate a part of a col'ective whole. Propositions of negation express the complement of a proposition.

Connective propositions relate text facts or propositions to each other. Connective propositions may be expressed in the text or they may be inferred. They are important to providing text cohesion. The arguments of connective propositions are often propositions themselves. Eight categories of connectives are specified:

1. Conjunction, expressing union, association, or combination.
2. Disjunction, expressing opposition or alternatives.
3. Causality, expressing cause-and-effect or correlated events.
4. Purpose, expressing reason, purpose or intent.
5. Concession, expressing admission of a point or yielding.
6. Contrast, expressing divergence or comparison.
7. Condition, expressing prerequisite states, restriction, or qualification.
8. Circumstance, expressing time, location or mode of action.

A text base, then, is a cohesive, interrelated set of predicate, modifier, and connective propositions. These propositions represent the meaning of text. The target text may be coherent, connected discourse united by a common theme or topic or it may be incomplete and characterized by missing logical links, facts, references, etc. The propositions suggested by the text itself may not be sufficient to form a connected and coherent text base. The reader may be called upon to supply prerequisite general or contextual knowledge and to make inferences about possible, likely or necessary bridging propositions in order to establish semantic coherence. The incidence of inference construction is recognized by the Kintsch and van Dijk prose processing model; the model does not address itself to the nature of processing inherent in inference construction.

Turner and Greene (1978) state that the Kintsch and van Dijk model of prose processing can be used as a tool for research into the cognitive processes involved in the comprehension of text. Kintsch and van Dijk have examined the relationship between meaning as represented using propositional analysis and behavioral indices of processing difficulty.

They have demonstrated a relation between number of propositions expressed in a text base and processing difficulty. Kinsch and Keenan (1973) systematically varied the number of propositions in a text base while holding constant the number of words in the selection. They observed that reading time increased and recall decreased as a function of number of propositions expressed. Kintsch et al. (1975) looked at processing difficulty as a function of the number of different arguments used in a text base. Short texts controlled for number of words and propositions and differing in number of different arguments were read and recalled by groups of subjects. Reading times were longer and recall poorer for texts with many different arguments. Texts with fewer arguments had shorter reading times and higher levels of recall. Kintsch and van Dijk conclude that comprehension difficulty is positively related to the number of propositions that must be processed and the number of different arguments that need be encoded.

Miller and Kintsch (1980) propose that, in addition to propositional density and number of different arguments, comprehension difficulty is related to the incidence of inference construction. Using a computer program written in two parts, a chunking program to perform the initial segmentation of text and a coherence program to simulate processes involved in maintaining semantic coherence, Miller and Kintsch examined processing difficulty and inference construction. Miller and Kintsch modeled twenty texts of varying readability and used these data to predict empirically generated recall and readability statistics. They found significant relations between number of connecting or bridging inferences necessary to connect segments of text and reading time and recall data. They summarize that the processing necessary to generate inferences implied by or implying stated propositions and necessary to semantic coherence is psychologically relevant and related to comprehension difficulty.

A MODEL OF VERBAL PERFORMANCE

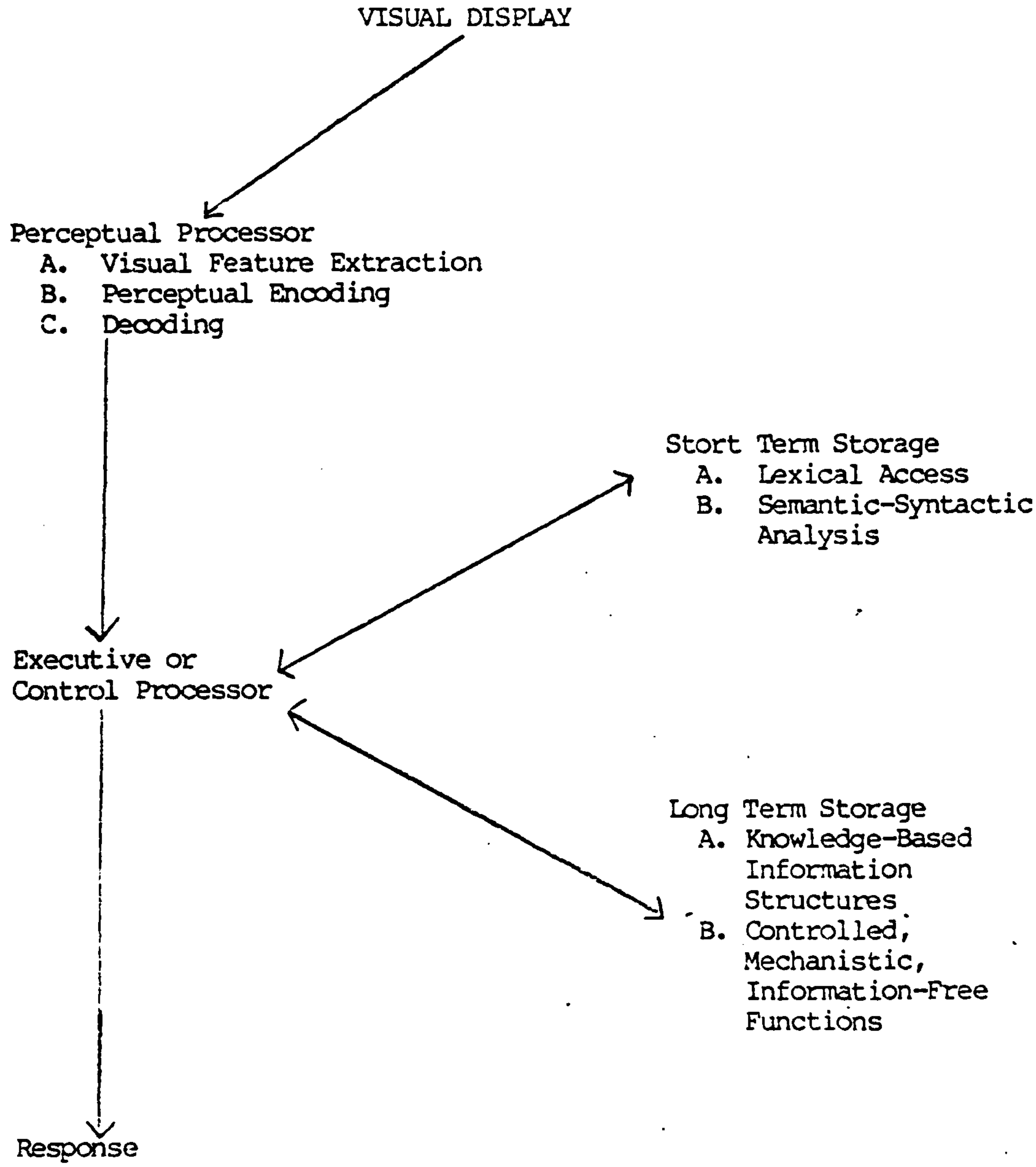
The theory and methods of factor analytic, information processing, chronometric analysis, cognitive correlates and componential analysis approaches to the study of individual differences are summarized above. This review of the definition and measurement of cognitive processing operations, stores, and strategies involved in performance on verbal tasks provides a framework for the following general model of verbal performance.

The model of verbal performance outlined below brings together portions of the paradigms outlined by Frederiksen, Hunt, Carroll, Pellegrino, Kintsch, and others. The model describes verbal performance in terms of the subset of processing skills associated with text analysis.

The model depicts verbal performance by five processing or storage structures. The first structure might be thought of as a perceptual processor, the second as an executive or control processor, the third as the locus of lexical access and semantic-syntactic analysis, the fourth as knowledge-based information and mechanistic information-free storage, and the fifth as a response processor. Each structure is discussed below. The structures are not strictly serially or hierarchically ordered. The flow of information within the system is not necessarily sequential or parallel. A schematic of the model follows.

Figure 1

Model of Verbal Performance



The PERCEPTUAL PROCESSOR is the structure that inputs stimulus information to the processing system. It includes the set of operations that converts the physical stimulus to a form that is appropriate for the task; it includes the operations that match stimuli to appropriate grapheme and phonemic representations. The perceptual processor is characterized by three operations described by Frederiksen as: visual feature extraction, perceptual encoding, and decoding. Visual feature extraction is the operation by which different types of information about the stimulus display are extracted. Perceptual encoding is the operation by which information is input into the system, and decoding is the operation by which arbitrary physical patterns are recognized as representations of grapheme and phoneme concepts in the lexicon. These operations may be thought of as automated, mechanistic processes. The processor may be thought of as a short term sensory storage or buffer component.

The EXECUTIVE OR CONTROL PROCESSOR is the structure that controls the flow of information in the system and has access to the various levels of memory storage. This structure determines the nature of a problem, selects processes for solving a problem, decides on a strategy for combining these processes, decides how to allocate processing resources, decides how to represent the information upon which processes act, and monitors solution processes. This structure is analogous to Sternberg's metacomponent and to the executive processor described by Snow, Whitely, and others.

The LEXICAL ACCESS/SEMANTIC-SYNTACTIC ANALYSIS STRUCTURE is a short term storage or working memory structure. In lexical access/semantic-syntactic analysis, an attempt is made to match letter strings input at the perceptual processing stage to appropriate semantic referents. Analysis is directed at attaching meaning to perceptual patterns. For phrase and sentence units, analysis is also directed at organizing these meaning elements into coherent text representations. Lexical, semantic, and syntactic knowledge is called upon in the identification of words and in phrase and sentential analysis.

Lexical Access is defined as the retrieval of information about individual words from long term memory. In lexical access, grapheme and phoneme data drive the retrieval of semantic information.

Semantic-syntactic analysis takes place in short term memory; it is defined by the retrieval of knowledge-based structures and information-free functions. These structures are discussed by Hunt (1978); they are described in greater detail in the following paragraphs. In semantic-syntactic analysis, the knowledge-based and information-free long term memory structures are accessed and, in the case of the information-free functions, executed in short term memory to form a semantically coherent representation of prose. Information about individual words stored in long term memory is retrieved and arranged to form a semantically coherent structure with respect to current sensory input and recognition of the previous state of active memory. Kintsch and van Dijk (1978) have developed a prose processing model which references the types of knowledge-based structures and information-free functions involved in semantic-syntactic analysis.

The fourth structure is a long term storage structure. This structure is the locus of KNOWLEDGE-BASED INFORMATION STRUCTURES and CONTROLLED, MECHANISTIC, INFORMATION-FREE FUNCTIONS. The knowledge-based information structures represent semantic and syntactic knowledge. These structures represent the ability to deal with words and the concepts they represent. They reflect experience with and cognizance of the English language. The knowledge-based information structures are also associated with knowledge of the world and world events. These knowledge structures are mediated by verbal knowledge but represent information about the world ancillary to mastery of the English language.

The controlled, mechanistic, information-free functions are the operations by which information structures are transformed to equivalent structures necessary for task performance. No semantic or syntactic information is associated with these strategic knowledge structures. These operators are defined by learned, stored transformation rules. These transformations do not involve new information abstraction; rather, they require the application of stored rules to information structures already present. Examples of controlled, mechanistic information-free operators are the processes of inferring, comparing, generalizing, reinstating, and image construction.

The comparison operator, for example, is the structure by which two or more information structures are examined and judged to be the same or different. The inferencing operator is a structure used to establish links between propositions when semantic coherence for a text is not maintained via shared arguments.

Inferencing strategies generate the missing or non-derivable connecting or bridging information necessary to maintain semantic coherence. Inference processes may be used to determine reference or define enabling conditions. They may also be used to specify resultant events, that is, events not entailed by stated conditions, but bearing a high probability of occurrence given stated conditions.

The final structure is the RESPONSE OPERATOR. This is the structure through which appropriate motor actions are selected and executed. The response operator is the structure by which the examinee makes an observable response, such as selecting one response from a set of multiple alternatives.

This verbal comprehension or verbal information processing model characterizes performance with respect to the subset of processes which underly text comprehension. The model synthesizes portions of processing paradigms described by Frederiksen, Hunt, Carroll, Pellegrino, Kintsch, and others. It provides a useful conceptual framework for the examination of cognitive processing operations, stores, and strategies involved in performance on verbal tasks.

REFERENCES

Barnes, G. M. & Whitely, S. E. Problem Restructuring Processes for Ill Structured Verbal Analogies. Memory and Cognition, in press.

Burt, C. The Genetics of Intelligence. In W. B. Dockrell (Ed.), On Intelligence. Toronto: The Ontario Institute for Studies in Education, 1970.

Carroll, J. B. Psychometric Tests as Cognitive Tests: A New Structure of Intellect. In L. B. Resnik (Ed.), The Nature of Intelligence. Hillsdale, Jersey: Lawrence Erlbaum Associates, 1976.

Cooper, L. A. Spatial Information Processing: Strategies for Research. In R. E. Snow, P. A. Federico & W. E. Montague (Eds.), Aptitude, Learning and Instruction, Volume 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Egan, D. E. Testing Based on Understanding: Implications from Studies of Spatial Ability. Intelligence, 1979, 3, 1-15.

Ekstrom, R. B., French, J. W., and Harman, H. H. Manual for Kit of Factor Referenced Cognitive Tests. Princeton, New Jersey: Educational Testing Service, 1976.

Falmagne, R. J. (Ed.). Reasoning: Representation and Process. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Frederiksen, J. A. Chronometric Study of Component Skills in Reading. Cambridge, Massachusetts: Bolt, Beranek and Newman, 1978.

Frederiksen, J. A. Component Skills in Reading: Measurement of Individual Differences Through Chronometric Analysis. In R. E. Snow, P. A. Federico, & W. E. Montague, (Eds.), Aptitude, Learning and Instruction, Volume 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Gentile, J. R., Tedesco-Stratton, L., Davis, E., Lund, N. J. & Agunane, B. C. Associative Responding Versus Analogical Reasoning by Children. Intelligence, 1977, 1, 4, 369-380.

Guilford, J. P. The Nature of Human Intelligence. New York: McGraw Hill, 1967.

Hunt, E. What Kind of Computer is Man? Cognitive Psychology, 1971, 2, 57-98.

Hunt, E. The Memory We Must Have. In R. Schank & K. Colby (Eds.), Computer Models of Thought and Language. San Francisco: Freeman, 1973.

Hunt, E. Quote the Raven? Nevermore! In L. Gregg (Ed.), Knowledge and Cognition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Hunt, E. Varieties of Cognitive Power. In L. B. Resnik (Ed.), The Nature of Intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976.

Hunt, E. Mechanics of Verbal Ability. Psychological Review, 1978, 85, 109-130.

Hunt, E. Intelligence as an Information-Processing Concept. British Journal of Psychology, 1980, 71, 449-474.

Hunt, E., Frost, N., & Lunneborg, D. Individual Differences in Cognition: A New Approach to Intelligence. In G. H. Bower, (Ed.), The Psychology of Learning and Motivation, Volume 7. New York: Academic Press, 1973.

Hunt, E. & Lansman, M. Cognitive Theory Applied to Individual Differences. In W. K. Estes (Ed.), Handbook of Learning and Cognitive Processes: Introduction to Concepts and Issues, Volume 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Hunt, E., Lunneborg, C., & Lewis, J. What Does it Mean to be High Verbal? Cognitive Psychology, 1975, 7, 194-222.

Kintsch, W. The Representation of Meaning in Memory. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Kintsch, W. & Keenan J. Reading Rate and Retention as a Function of the Number of Propositions in the Base Structure of Sentences. Cognitive Psychology, 1973, 5, 3, 257-274.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. Comprehension & Recall of Text as a Function of Content Variables. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 196-214.

Kintsch, W. & van Dijk, T. Toward a Model of Text Comprehension and Production. Psychological Review, 1978, 85, 363-394.

Miller, J. R. & Kintsch, W. Readability and Recall of Short Prose Passages: A Theoretical Analysis. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 4, 335-354.

Mulholland, T., Pellegrino, J. W., & Glaser, R. Components of Geometric Analogy Solution. Cognitive Psychology 1980, 12, 252-284.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts 1967.

Newell, A. & Simon, H. A. Human Problem Solving. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

Pellegrino, J. W. & Glaser, R. Cognitive Correlates and Components in the Analysis of Individual Differences. Intelligence, 1979, 3, 187-214.

Pellegrino, J. W. & Glaser, R. Components of Inductive Reasoning. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.) Aptitude, Learning and Instruction, Volume 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Posner, M. I. Abstraction and the Process of Recognition. In G. H. Bower & T. J. Spence (Eds.), The Psychology of Learning and Motivation: Advances in Research and Theory. New York: Academic Press, 1969.

Posner, M. I. & Mitchell, R. F. Chronometric Analysis of Classification. Psychological Review, 1967, 74, 392-409.

Resnik, L. B. (ed). The Nature of Intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976.

Rose, A. M. Information-Processing Abilities. In R. E. Snow, P. A. Federico & W. E. Montague (Eds.), Aptitude, Learning and Instruction, Volume 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Royer, F. L. Information Processing in the Block Design Task. Intelligence, 1977, 1, 32-50.

Simon, H. A. Identifying Basic Abilities Underlying Intelligent Performance of Complex Tasks. In L. B. Resnik (Ed.), The Nature of Intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976.

Sternberg, R. J. Component Processes in Analogical Reasoning. Psychological Review, 1977, 31, 356-378 (a).

Sternberg, R. J. Intelligence, Information Processing and AnalogicalReasoning: The ComponentialAnalysis of Human [illegible]ities. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977 (b).

Sternberg, R. J. Toward a Unified Componential Theory of Human Reasoning. In M. Friedman, J. Das, & N. O'Connor (Eds.), Intelligence and Learning. New York: Plenum Press, 1980.

Sternberg, R. J. & Weil, E. M. An Aptitude-Strategy Interaction in Linear Syllogistic Reasoning. Journal of Educational Psychology, 1980, 72, 226-234.

Thurstone, L. L. Multiple Factor Analysis. Chicago, Illinois: University of Chicago Press, 1947.

Turner, A. & Greene E. Construction and use of a Propositional Text Base. JSAS Catalogue of Selected Documents in Psychology, 1978, 3, 58, MS 1713.

Whitely, S. E. & Barnes, G. M. The Implications of Processing Event Sequences for Theories of Analogical Reasoning. Memory and Cognition, 1979, 1, 323-331.

Whitely, S. E. & Schneider, L. M. Process Outcome Models for Verbal Aptitude. Technical Report, NIE-80-1, for the National Institute of Education, 1980.

Whitely, S. E. & Schneider, L. M. Information Structure for Geometric Analogies: A Test Theory Approach. Applied Psychological Measurement, 1981, 5, 3, 383-397.

Examination of Ability Requirements

For the Infantry Career Management Field*

Darlene M. Olson
Lawrence M. Hanser

U.S. Army Research Institute
Alexandria, Virginia 22333

Research (Fleishman, 1976, 1982; Dunnette, 1976; Peterson & Bownas, 1981; and Rossmeissl & Dohme, 1982) has shown that ability requirements are important in military organizations for: 1) selection and classification, 2) weapon system design and development, 3) training and 4) performance evaluation. The cognitive and psychomotor domains have been extensively taxonomized by Fleishman (1972, 1975), who defined the relationship between patterns of human abilities and performance on diverse tasks. In addition, researchers at the Educational Testing Service (ETS) (Ekstrom, French, Harman, & Dermen, 1976) have examined the structure of cognitive abilities and have developed measures of these aptitudes. Although these taxonomies are conceptually useful, more research is needed to 1) describe the ability requirements of various military jobs, 2) establish linkages between ability taxonomies and task/job requirements of specific MOS, and 3) describe the relationship between patterns of abilities and variability in individual performance across tasks and over time. Research in these areas could provide empirical information that might better optimize the match between enlisted personnel with certain abilities and specific requirements of an MOS.

This paper examines the pattern of abilities required for effective performance in the Infantry Career Management Field (CMF 11). The purposes of this exploratory research were: 1) to examine whether specific cognitive abilities could be identified as requirements for effective performance in Infantry Military Occupational Specialties (MOS) and 2) to assess these ability requirements through an experimental computerized rating methodology.

---

*Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983.

Design of the Research

Characteristics of the Sample

The data collection sample consisted of a total of nine Subject Matter Experts (SMEs), who were on active duty in various training units at the Infantry School at Ft. Benning, Georgia. These soldiers were chosen as SMEs because of their knowledge of the job tasks and ability requirements for Infantry MOS. Six of the SMEs were Company Commanders and three were Non-Commissioned Officers (NCOs). All the SMEs were White males and had an average of ten years of military service.

Description of the Rating Instrument

This research used a computerized job assessment system, which was developed under a contract to the Systems Research Laboratory of the Army Research Institute (ARI) (Rossmeissl, Tillman, Riggs, & Best, 1983). The job assessment system operationalized in this study used the following equipment: 1) Apple II computer, 2) monitor, 3) two disc drive units, and 4) the Job Assessment Software System (JASS). The rating materials developed in the JASS combined a branching network decision flowchart procedure with rating scales for each ability. The computerized rating scales allow estimation of the level of each ability required for a specific job. The JASS contained 39 cognitive, perceptual and psychomotor abilities that were rated for each Infantry MOS. These abilities included such dimensions as Stamina, Memory, Spatial Visualization and Manual Dexterity.

Description of the Infantry MOS

The Infantry Career Management Field (CMF 11) is comprised of the following four MOS: Basic Infantryman (11B MOS), Indirect Fire Infantryman (11C MOS), Heavy Anti-armor Weapons Crewman (11H MOS),and the Fighting Vehicle Infantryman (11M MOS). Although all of the Infantry MOS require proficiency in basic soldiering and combat skills, three of the Infantry MOS have more specialized duties. In the 11C MOS, soldiers have such duties as emplacement of mortar aiming posts, maintenance of correct sight picture and computation of mortar firing directions. As members of an armored tank crew, soldiers in the 11H MOS have such responsibilities as identifying and destroying enemy armor, evaluating terrain conditions, and firing the TOW weapon system. In the 11M MOS, soldiers in the Infantry Fighting Vehicle (IFV) are involved in the operation of a broad group of weapons (e.g., 25mm cannon and coax 7.62 machine gun) under varied tactical conditions (e.g., tactical operation at night with the IFV in motion when employing the weapon systems).

Procedure

The SMEs in the data collection sample were assembled in groups of three in a conference room at the ARI Field Unit at Ft. Benning. The SMEs provided ability ratings for the Infantry MOS by interfacing with an Apple II computer via a keyboard connected to a CRT display. Specifically, the JASS rating format presented the SMEs with individual sequential definitions of each cognitive ability through a visual display on the CRT. The rater responded on the keyboard terminal either 'Yes' (Y) or 'No' (N) to indicate whether the ability was required for the MOS. A 'No' response from the rater advanced the next aptitude for rating. When the rater responded 'Yes' a scale block appeared on the CRT; the rater then estimated the required level (magnitude) of the ability using a 7-point scale, that did not contain descriptive behavioral anchors. Scaling of the actual ability was accomplished by moving the keyboard symbol (>) to the desired numeric position on the rating scale. Each ability rating was stored in computer memory by pressing the 'Enter' key on the terminal keyboard. After a 'Yes' response the next ability was displayed for rating by pressing the carriage return key. All rating data were stored on disc files and aggregated separately for the four Infantry MOS. In order to determine the average ratings (profile) of required abilities for a specific Infantry MOS, data were collapsed across type of rater (Company Commander vs. NCO). When individual SMEs had experience in more than one Infantry MOS, the JASS rating procedure was repeated for those MOS.

Results

Scale Interpretation

In order to examine the ability requirements for specific Infantry MOS and compare profiles of abilities across MOS, a post-hoc analysis of the data was conducted. This rational approach to scale interpretation provided a conceptual framework for meaningful discussion of rating results. The scale of possible ratings ranged from 0 to 21. A rating above 18 was interpreted to mean that the ability was most highly required for successful performance in a specified MOS. A rating in the range of 15-18 was considered as highly required. A rating below 6 was interpreted to mean that the ability was not highly required for successful performance. Since the majority of the obtained ratings were in the scaled range of 6-15, this scaled area was considered average. Since it was deviations from the average that were of interest for examining differences in ability requirements between MOS, this scaled area did not provide particularly useful information for comparative purposes. The profile of average ratings for each MOS was obtained by averaging the ratings for all SMEs who rated the same MOS.

Reliability of the Ratings

The nine SMEs provided a total of 21 sets of ratings. Since there were an unequal number of raters between Company Commanders and NCOs, and each SME evaluated different MOS, it was necessary to estimate the reliability of incomplete sets of ratings (i.e., not all raters rated every MOS). Hence, reliability estimates for these ratings were derived from computation of the Intraclass Correlation Coefficient for each Infantry MOS (Ebel, 1951). Reliability estimates of .55, .62, and .44 were obtained respectively for ratings of the 11B, 11C, and 11H MOS. Since only one set of ratings was obtained for the 11M MOS, the reliability of these data could not be determined.

Ability Requirements for Infantry MOS

An examination of the profiles of average ratings of required abilities for the Infantry shows which abilities were judged critical for effective performance in a specific MOS and permit general comparisons of ability requirements among Infantry MOS.

Basic Infantryman (11B MOS). In the 11B MOS, in which physical qualifications are among the most demanding in the Army, the profile of average ratings of required abilities confirms the importance of such physical dimensions as Stamina, which involves the maintenance of physical activities for prolonged periods of time and Extent Flexibility, Reaction Time, and Speed of Limb Movement, which involve flexibility and speed of coordinated muscle groups (i.e., arms, trunk and legs). Besides the need for agility and endurance, raters judged the ability to memorize and retain large amounts of information and maintain a correct visual directional sense with respect to interaction with the physical environment (Spatial Orientation) as highly required for successful performance. For the Basic Infantryman, a cluster of abilities which included Explosive Strength, Multi-limb Coordination, Static Strength, and Rate Control, were not judged as highly required for effective performance.

Indirect Fire Infantryman (11C MOS). In the 11C MOS, soldiers serve as members of a mortar squad. In order to perform effectively in this MOS, the profile of average ratings indicates that not only are the basic physical endurance and memorization abilities required, but that fine coordinated movements of arms, wrists, hands, and fingers (Wrist-Finger Speed, Manual Dexterity and Control Precision) such as those which might be necessary in positioning/adjusting equipment, are also crucial. The abilities of Spatial Orientation, which could impact on the accurate placement of mortar aiming posts, maintenance of proper site picture, and firing of the mortar; and Number Facility, which is relevant to the computation of firing directions, were judged as highly required for successful performance in the 11C MOS. As in the 11B MOS, such abilities as Explosive and Static Strength and Multi-Limb Coordination were not rated as highly required.

Heavy Anti-Armor Weapons Crewman (11H MOS). In the 11H MOS, soldiers are required to perform all the duties and tactical operations that are necessary to assault, destroy and defend against enemy tanks and armor vehicles. Ratings indicate that the ability to make controlled muscular movements such as those necessary to adjust a machine or equipment control mechanism (Control Precision) was most highly required for successful performance. This rating for Control Precision distinguishes the ability requirements for the 11H MOS from both the 11B and the 11C MOS, in which Stamina received the highest rating. A cluster of abilities also judged as important for effective performance as a Heavy Anti-Armor Weapons Crewman were: 1) to orient one's body in space (Spatial Orientation), 2) to respond quickly to oral commands and visual signals (Reaction Time), 3) to accurately distinguish similarities and differences in the comparison of visual patterns/objects (Perceptual Speed) and 4) to maintain arm-hand steadiness. Ability requirements such as these could conceivably play a role in the performance of such 11H tasks/duties as loading and firing the TOW weapon system, identifying enemy armor and other targets, and performing land navigation functions.

Fighting Vehicle Infantryman (11M MOS). The one set of ratings obtained for the 11M MOS reveals that nearly two-thirds of the rated abilities fell at the extremes of the scale. For example, at the highest end of the scale the following abilities serve to differentiate the requirements of the 11M MOS from other Infantry MOS: 1) to evaluate and organize information in accordance with planned activities (Information Ordering), 2) to allocate time among competing tasks (Time Sharing), 3) to selectively attend to and disembed relevant information in the presence of distracting stimuli, and 4) to expend the maximum amount of energy in a series of explosive muscular acts. In particular, an interesting finding was that Explosive Strength was rated as much more important for success as a crewman on the IFV than for the other Infantry MOS. Since the 11M MOS involves various positions (i.e., Commander, Gunner, Dismount Crew), perhaps the SME changed rating perspective as he judged the various abilities.

Discussion and Conclusions

The major conclusions of this exploratory research were that 1) profiles of cognitive, perceptual, and psychomotor abilities could be identified as requirements for effective performance in the various Infantry MOS, and 2) these ability requirements could be reliably assessed through application of a computerized rating methodology.

Specifically, comparisons of ability requirements across Infantry MOS indicate that the abilities of Memorization and Spatial Orientation were judged to be highly required for successful performance in the Infantry. In contrast, the abilities of Speed of Closure, Static Strength, Trunk Strength, and Rate Control were not judged as highly required for effective performance as an Infantryman.

In specific Infantry MOS, certain abilities emerged as highly required for effective performance. In the 11B MOS, physical abilities that were related to agility and endurance were important. For success as an Indirect Fire Infantryman, the ability to compute basic arithmetic operations and make fine coordinated limb movements were highly rated. In the 11H MOS, a combination of perceptual and psychomotor abilities such as those required to accurately identify similarities/differences among visual patterns and objects and to maintain steady controlled movements of arms/hands were judged as important for performance. A diverse group of physical, cognitive and perceptual requirements which included the ability to perform explosive muscular acts, organize and evaluate information, and selectively attend to and extract salient information from distracting stimuli were judged as important for the 11M MOS.

Although this research identified some abilities required for effective performance in the Infantry, there were limitations associated with the data collection and methodology. First, the testing procedures lacked adequate standardization and computer equipment malfunctions caused the loss of several sets of ratings. Secondly, the JASS contained cognitive and perceptual motor abilities. Whether the inclusion of other more affective/interpersonal dimensions in the JASS would have more comprehensively defined the ability requirements for the Infantry MOS was not examined. Finally, the graphic rating format used in this research did not include behavioral dimensions or task descriptions as anchor points to assist the raters in their evaluations of ability requirements. Future research which uses the JASS methodology should investigate the role of different rating formats, the use of general versus specific anchors in scale development, and the inclusion of non-cognitive rating dimensions.

REFERENCES

Dunnette, M.D. (1976). Basic attributes of individuals in relation to behavior in organizations. In M.D. Dunnette (Ed.), _Handbook of industrial and organizational psychology_ (pp. 469-520). Chicago: Rand McNally.

Ebel, R.L. (1951). Estimation of the reliability of ratings. _Psychometrika_, _16_ (4), 407-424.

Ekstrom, R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). _Manual for Kit of Factor-Referenced Cognitive Tests._ Princeton, New Jersey: Educational Testing Service.

Fleishman, E.A. (1972). On the relation between abilities, learning, and human performance. _American Psychologist_, _27_, 1017-1032.

Fleishman, E.A. (1975). Toward a taxonomy of human performance. _American Psychologist_, _30_, 1127-1149.

Fleishman, E.A. (1982). Systems for describing human tasks. _American Psychologist_, _37_ (7), 821-834.

Peterson, N.G., & Bownas, D.A. (1982). Skill, task structure, and performance acquisition. In E.A. Fleishman (Ed.), _Human performance and productivity: Human capability assessment_ (pp. 49-105). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rossmeissl, P.G., & Dohme, J.A. (1982). Using rating scales to determine aptitude requirements. _Proceedings of the 24th Annual Conference of the Military Testing Association_, _2_, 350-361.

Rossmeissl, P.G., Tillman, B.W., Rigg, K.E., & Best, P.R. (1982). _Job assessment software system (JASS) for analysis of weapon systems personnel requirements: Research report_ (ARI Research Report MGA 4581-MRO). Alexandria, VA: U.S. Army Research Institute.

XI. SUMMARY AND CONCLUSIONS

John P. Campbell (HumRRO)

Project A was designed and is being executed within a different framework than that of previous R&D projects in the behavioral and social sciences. Historically, past practice has been to allocate funds in relatively small amounts to one investigator or to one research firm for a relatively circumscribed piece of work. Project A was conceived differently. Its aim is to address an integrated set of R&D questions and problems within one project, to develop a complete personnel system for selecting and classifying all entry-level personnel in a large organization. Much of the information required to develop such a system could not be produced by a set of piecemeal projects. Consequently, while the magnitude of Project A is large in terms of total funding, time frame, size of the research staff, and number of research participants, it is expected to produce much more information in a shorter time than would have been the case if the usual framework for allocating R&D funds had been allowed.

Executing and managing such a large integrated project place heavy responsibilities both on the contractor staff and on the professional staff of the Army Research Institute. However, concomitant with the burden of responsibility is the expectation that the resulting classification procedure will be grounded in the most complete data base ever developed for a large personnel system and that many of the most vexing research questions in the field will be addressed comprehensively and directly. After 1 year's experience on the project, the weight of responsibility and the realization that the payoffs will far exceed anything that has gone before, are felt more intensely than ever.

Planning Activities

In general, the first year's activities have been taken up by an intensive period of detailed planning, briefing the advisory groups, preparing the initial troop requests, and beginning the comprehensive predictor and criterion development that will be the basis for the later validation work. The requirement for a detailed research plan to be produced during the first 6 months of the contract was included in the RFP; hindsight judges it to be an even more valuable step than the authors of the RFP might have had in mind. The research staff devoted a great deal of effort to the writing of the research plan, it was carefully reviewed by the advisory groups and by the ARI professional staff, revisions were made, and the completed plan was published in May 1983 under the joint authorship of the contractor and ARI staffs. The Research Plan and the accompanying Master Plan carefully lay out, in detailed fashion, the specific steps to be taken by each subtask in the project, the schedule which will be followed, and the budget allocations that will be made to each subtask during each contract period. These two documents have become the guiding blueprint for the project. They have also proven invaluable as a mechanism for developing a consensus and facilitating communication among contractor staff and between the contractor and ARI.

The detailed planning and review that went into the development of the Research Plan and Master Plan made it possible to lay out clearly and precisely the troop support the project would need during its first 2 years.

Consequently, the project has experienced relatively little difficulty in communicating its needs to the appropriate Army organizations and in gaining their support. For the outstanding cooperation we have received so far, we are most appreciative.

Substantive Activities

The previous chapters in this report have outlined and discussed the substantive activities that have taken place during the first contract year. The major points from this discussion are summarized below.

The LRDB and the FY81/82 Data File

As noted by Hanser and Grafton (1983), no one should expect easy going when attempting to use large-scale computer files of archival data for personnel research. Computerized information systems in large organizations are designed to serve purposes other than personnel research. Consequently, it came as no surprise that the predictor and criterion data for 81/82 accessions were not as neat and clean as we would have liked. In fact, a tremendous amount of effort was devoted to obtaining and merging computer files, editing records, and filling gaps in documentation. However, the result has been the creation of the most extensive file of archival records that has ever been generated for purposes of personnel research. The files encompass 2 years of Army accessions (approximately 200,000 people drawn from approximately 500,000 applicants and subsequently placed in over 300 different skilled entry-level positions). While the available edited records fall far short of containing complete data for everyone, the magnitude of the data base is considerable.

Work on the analyses of these data has just begun and it is too early to make definitive statements about empirical findings. However, one principal objective is to use 81/82 data files to investigate the validity of using new or revised composites of ASVAB subtests to predict success in a wide variety of job training schools. Regardless of whether the analyses point toward new composites, revisions, or no change in those currently being used, the analyses will be based on a far larger data base than ever before.

One obvious, but extremely important, finding from the preliminary analyses is that although the Army is a large organization it is not so large that every MOS (job specialty) contains a sufficient number of incumbents to permit statistical validation analyses. It simply will not be possible to estimate prediction equations empirically for every MOS. Also any validation analysis must deal with differing criterion metrics across MOS, restriction of range due to selection, and considerable skewness in the criterion distributions. Consequently, one extremely significant outcome resulting from having the 81/82 data file is that alternative analytic methods for dealing with these problems can be tried out and evaluated, so that the analyses of the 83/84 and 86/87 data can proceed efficiently and appropriately. The analyses of the 81/82 data base will serve as the benchmark with which the subsequent results to be produced by Project A can be compared. That is, we now have enough information in the 81/82 file to provide a reasonably clear picture of how much selection validity and classification efficiency can be produced within the current system using the current data base.

MOS Task Descriptions

Because the information was not generated for personnel research purposes, the Army's MOS job analysis data needed considerable modification before it could be used by Project A for criterion development. Consequently, a great deal of effort was devoted to refining and integrating task descriptions from the Soldier's Manual and the CODAP occupational survey questionnaires. For each MOS, a data bank of task statements was accumulated from all available sources, and the individual task statements were edited to determine if they indeed focused on observable job tasks, if they were redundant or overlapped with other tasks, and if they were at the same level of generality. Subject matter experts were used to determine if the edited pool of task descriptions provided a complete picture of the content of the MOS. The SME also judged the relative criticality of each task. These steps are currently being carried out for focal MOS so that there will be a precise and thoroughly developed task description for the MOS being considered in Project A. The task descriptions will provide the principal basis for the development of hands-on performance measures and job knowledge tests. As such, they should provide a much better foundation for the subsequent criterion development than has been available in the past.

Assessment of Training Performance

A major objective of Project A is to use a comprehensive, valid, standardized test construction procedure to develop a measure of training success for each focal MOS so that the item content represents both the content of training and the content of the job. That is, the items will sample the job content representatively and will be further identified as being covered in training vs. not being covered in training. When this is accomplished, a measure of direct learning in training (items that match training content) and a measure of indirect learning (scores on the items <u>not</u> directly related to training content) can be related to a variety of job performance criteria with and without ability (as measured by predictor tests) controlled.

To meet these objectives the project staff have spent the past several months visiting key training schools and developing job task descriptions for each MOS. What has been produced is a thorough analysis of the objectives, curriculum, and assessment procedures for the key schools. The process of describing the job content and matching it with training content has just begun and will be completed during FY84. When the matching of training content and job content is completed and the knowledge tests are constructed, we will have achieved the capability for determining how training performance is, or is not, related to job performance.

Job Performance Criterion Development

Our initial model of soldier effectiveness was a bit crude and not well explicated. We said essentially that both specific task performance and the general factors of commitment, morale, and organizational socialization comprised the total domain.

During this year the task descriptions for the four MOS in Batch A have been completed and Batch B is in progress. Further, virtually all the critical incident workshops for MOS-specific task performance factors have been completed. This easily has been the most massive effort ever undertaken to apply these methods to criterion development. There now exist hundreds of critical incidents of specific task performance within each focal MOS, and thousands of critical incidents describing performance behaviors that have a general, not MOS-specific, referent. These large samples of job behaviors are being used to identify MOS-specific and MOS-general performance factors and to develop rating scales (during FY84) to assess individual performance on these factors. This process has produced a revised and expanded model of the criterion space that will be used to generate further criterion development work and to guide predictor selection.

An additional important outcome of the interaction between model development and task/behavior description is the identification of an array of MOS-specific task performance factors that are intended to encompass the unique task content of all MOS in the enlisted personnel job structure. Although it is only a first cut, it will be the basis for the further development of a standardized set of task descriptors that can be applied to any MOS so as to describe thoroughly its content. Such a standardized measure will make it possible to answer a number of important questions that could not have been addressed previously. For example, how similar precisely are any two MOS in terms of their job content? Should they have a common selection algorithm? How different should their training schools be? Etc.

Predictor Selection

A major objective that had to be accomplished during the first contract year was to select the preliminary predictor battery for administration to the 83/84 longitudinal sample and to lay the groundwork for the development of the trial predictor battery. To do this, the project staff carried out what was perhaps the most massive literature search ever done in personnel psychology. The result has been: (a) a very thorough and precise description of the specific measures that might be useful in any selection or classification effort, (b) a summmary of the empirical evidence attendant to each one, and (c) an explication of the latent variables, or constructs, that seem to best represent the content of the operational measures or tests.

The value of this information, while it is extremely high for this particular project, goes far beyond the boundaries of Project A. It will be of crucial importance for almost any personnel selection project that comes after, regardless of the specific jobs or organizations in question. There is now a wealth of valuable and well-organized information that is available for use in future work.

In Conclusion

During its first contract year Project A has stayed on schedule and within its budget. More attention was devoted to detailed planning and outside review than the Army's research staffs had originally envisioned. However, these very thorough and careful preparatory steps were well worthwhile in terms of facilitating communication among everyone associated with the project and uncovering all the unresolved issues that would have plagued us at some later time. Most importantly, it served to coalesce all of the diverse organizational elements whose informed cooperation was essential to the successful execution of the research program. Project A has indeed become a unified and integrated effort.

Also, although much of the research activity during the first year was designed as essentially prepatory, some valuable first year products include the 81/82 data file, the task banks, the critical incident banks, and the literature review of the predictor domain.

We look forward to a productive second year.

REFERENCES

Campbell, J. P., Dunnette, M. D., Arvey, R., & Hellervik, L. The Development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.

Christal, R. E. (Chm.) Collecting, analyzing, and reporting information describing jobs and occupations. Symposium presented at the meeting of the American Psychological Association, Washington, D.C., September 1969.

Christal, R. E. The United States Air Force occupational research project (AFHRL-TR-73-75). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, January 1974.

Hanser, L. & Grafton F. Dusting Off Old Data: Problems Encountered With Archival Records. Paper presented at the American Psychological Association, Anaheim, CA., August 1983.

James, L. R., Muliak, S. A., & Brett, J. M. Causal analysis: Assumptions, models, and data. Beverly Hills, CA. Sage, 1982.

James, L. R., Wolf, G., & Demaree, R. Estimating interrater reliability in incomplete designs (IBR-81-14). Arlington, VA: Organizational Effectiveness Research Programs, August 1981.

Lawley, D. A note on Karl Perason's selection formulas. Royal Society of Edinburgh Proceedings, Section A, 1943, 62, 28-30.

March, J. E., & Archer, W. B. Procedural guide for conducting occupational surveys in the United States Air Force (Technical Report No. 67-11).

Miller, R. B. A method for man-machine task analysis (WADC Technical Report 53-137). Wright-Patterson Air Force Base, OH, June 1953.

Motowidlo, S. J., Dowell, B. E., Hopp, M. S., Borman, W. C., Johnson, P. D., & Dunnette, M. D. Motivation, satisfaction, and morale in Army careers: A review of theory and measurement. (ARI Tech. Rep. TR-76-A7). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences, December 1976 (NTIS No. AD-A036390).

Porter, L. W., Steers, R. M., Mowday, R. T., & Boulian, P. V. Organizational commitment, job satisfaction, and turnover among psychiatric technicians. Journal of Applied Psychology, 1974, 59, 603-609.

Profile of American Youth: 1980 nationwide administration of the ASVAB. Department of Defense (Manpower, Reserve Affairs, and Logistics) Washington, D.C., March 1982.

Rupe, J. C. Research into basic methods and techniques of Air Force job analysis - IV (AFPTRC-TN-56-51). Lackland Air Force Tase, TX: Air Force Personnel and Training Research Center, April 1956.

Shrout, P., & Fleiss, J. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 1979, 86, 420-428.

Smith, P. C. & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for ratings scales. Journal of Applied Psychology, 1963, 47, 149-155.

Steers, R. M. Antecedents and outcomes of organizational commitment. Administrative Science Quarterly, 1977, 22, 46-56.

Van Cott, H. P., Berkun, M. M., & Purifoy, G. R., Jr. How to determine job knowledge content: The application of a procedure to several Air Force jobs. Pittsburgh, PA: American Institutes for Research, December 1955.

Van Maanen, J., & Schein, E. H. Toward a theory of organizational socialization. In B. M. Staw (ed.), Research in organizational behaviors (Volume I). Greenwich, CT: JAI Press, 1979.

APPENDIX A

ELEMENTS IN THE FY81/82 DATABASE

APPENDIX A

DATA ELEMENTS IN THE FY81/82 DATABASE

1. Applicant/Accession Data

Information received from MEPCOM includes data on over 400,000 applicants in each of the two years. About 75,000 of the FY82 applicants are also on the FY81 applicant file. These were either individuals going into the Delayed Entry Program in one year and coming out the next, individuals whose processing was incomplete at the end of the year and completed the next, or individuals who made one or more separate applications in each of the two years. For all applicants, whether accepted or rejected, we have the following items of information:

A. Control Data Block

A1AFEES - MEPS Where Processing Took Place

A1ACTDTE - Action Date

A1UPSTAT - Entry Status of This Update

B. Mini-Personal Data Block

A1NAME

A1SSN

A1SEX

A1RACE - (Does not identify Hispanics separately)

A1DOB - Date of Birth

A1EDYRS - Years of Education

A1EDCERT - Highest Degree Earned

A1HOMLOC - State and County of Home Residence

A1HOMZIP - Zip Code of Home Residence

A1PRISRV - Prior Military Service

C. Aptitude Data Block

A1ASVFM - ASVAB Form Code

A1AFQTPC - AFQT Percentile Score

A1MCAT - AFQT Mental Category

A1ASVBxx - ASVAB Raw Subtest Scores

(Standardized subtest scores and Area Composites have been recomputed and added for applicants taking forms 8, 9, or 10)

A1ASVCyy - ASVAB Area Composite Scores

A1WHOTST - Military/OPM Tester

D. Medical Data Block

A1PULHES - 6 Dimensional Physical Profile

A1HGT - Height

A1WGT - Weight

A1SYSBLD - Systolic Blood Pressure

A1DIABLD - Diastolic Blood Pressure

A1MEDFLx - Medical Failure Codes

Most of the applicants who entered the service did so through the Delayed Entry Program which allows them to reserve a training slot for some future entrance. For all individuals entering or exiting the DEP in a given year, the following information is available (in addition to the data for all applicants).

E. DEP Enlistment Block

A1DEPDTE - Date of Signing DEP Contract

A1PADDTE - Projected Active Duty Date

A1ENTRST - Entry Status (3 for entry into DEP)

A1DEPPG - Program Enlisted For

A1TNGMOS - Training MOS

While most individuals signing DEP contracts do subsequently enter the Army, some do not. These individuals are then "discharged" from the DEP program without ever becoming accessions. For such cases, the following additional information is available:

F. DEP Discharge Block

A1DISDTE - Date of Discharge from the DEP

A1NODEPR - Reason for Discharge from DEP

Finally, for individuals actually entering the Army (accessions), the following information is entered in addition to the variables listed above for all applicants and for DEP entrants:

G. Maxi-Personal Data Block

A1PRSADD - State and County of Present Residence

A1PRSZIP - Zip Code of Present Residence

A1ETHNIC - Ethnic Group Affiliation (includes information necessary to identify Hispanics)

A1CITIZ - Citizenship Status

A1NRDEP - Number of Dependents

A1MARST - Marital Status

A1RELIG - Religious Preference

H. Accession Data Block

A1ENTDTE - Date of Accession to Active Service

A1SVCDTE - Service Computation Date (equal to date of enlistment except for recruits with prior service where the Active Duty Service Date is earlier)

A1DISDTE - Prior Discharge Date (Prior service cases only)

A1ENLTRM - Term of Enlistment (years, 1-6)

A1WAIVER - Indicates Specific Waivers

A1PAYGRD - Pay Grade at Time of Accession

A1GRDDTE - Date of Pay Grade

A1ENTRST - Entry Status (immediate, from DEP, from Reserves)

A1ENTPRG - Program Enlisted For

A1TRNMOS - Enlistment/Training MOS

A1PRMOS - Primary MOS (less consistent than A1TRNMOS)

A1YTHPG - Youth Program (ROTC, JROTC)

A1PGOPT - Program Option

A1TRNSF - Indicates Initial Processing Station

I. Service Required Data (Army)

A1ASI - Additional Skill Identifier

A1MDC - Movement Designator Code

A1GRDAB - Grade Abbreviation

A1BRKSV - Break in Service

A1USARN - USAR/ARNG Personal Code

A1REFCD - Referral Code

A1HRAP - Hometown Recruiter Aide Program

A1EDINC - Education Incentive Program

A1AITDT - Expected AIT Graduation Date (mostly missing)

A1NENL - Number of Times Enlisted/Reenlisted

2. Training Data

ARI has expended considerable effort to collect training information on 1981 and some 1982 accessions. These data indicate the timing and duration of training, the course(s) taken, the overall outcome, and some measure of performance in the course. It is important to note that the nature of these performance measures varies widely by school and sometimes by course or class within school.

A. <u>School Identification Information</u>

T1SCHOOL - School/ATC Code

T1COURSE - Name of Course

T1CLASS - Class ID Number Within Course

T1MOSAWD - MOS Award Upon Completion

T1SKLLVL - MOS Skill Level Awarded

B. <u>Students' Progress Through the Training Program</u>

T1ENRDTE - Enrollment Date

T1GRDDTE - Date of Recycle, Transfer, or Graduation

T1ATTRIT - Type of Attrition

T1DISP - Diposition (Pass, Recycle, Transfer or Drop)

T1SCORE1 - Students' Course Grade or Test Score

T1STYPE1 - Type of Score

T1SCORE2 - Secondary Performance Measure (for some MOS)

T1STYPE2 - Type of Additional Score

T1SELECT - Was Specific MOS Guaranteed for Basic Infantrymen

T1MORSE - Morse Code Taken for 05B and 05C

3. Data from the Enlisted Masterfile (EMF)

The Army Enlisted Masterfile (EMF) contains a significant amount of information that is essential to Project A. In particular, information on each soldier's progress through his or her Army career is captured by the EMF. The EMF also contains important information on the individual background and enlistment conditions of eac soldier that are important checks against similar information obtained from the accession files.

The FY81/82 database currently includes information from the FY82 year-end EMF file. Data from the FY83 year-end EMF and information on separations from the FY81, FY82, and FY83 "loss" files will be made available to us shortly. (The FY82 file does not include recruits who came in after July of 1982 or who separated prior to about June 1982.)

A. Individual Background Data

Note that most of this information has also been obtained from the accession files. The corresponding EMF variables are being used to check or verify the accession data. After completing this check, only one copy of this information will be retained. The background data elements from the EMF that are needed to either add or verify essential information include:

E1SEX - Sex

E1RACE - Population Group

E1REDCAT - Racial/Ethnic Descent

E1CLANG - Language Identify

E1CITIZ - Citizenship Status

E1DOB - Date of Birth

F1MARST - Marital Status

E1NRDEP - Number of Dependents

E1CIVED - Academic Education Level

E1MADCD - College Major

E1STRD - State of Residence at Enlistment

B. Enlistment Conditions

As with background data, much of the enlistment information has also been obtained from accession files. Again, only one copy of this information will be retained after any inconsistencies are resolved. The required enlistment variables include:

E1ASVCxx - All ASVAB Data Area Composite Scores

E1AFQSC - Armed Forces Qualification Test Score

E1AFQG - AFQT Group

E1DLAB - Defense Language Battery Score

E1PHYPR - Physical Profile

E1PHYCA - Physical Limitation Category

E1XFACT - Weight-Lifting Capacity

E1COMPT - Service Component

E1ENLOP - Enlistment Option Code

E1MORWA - Enlisted/Reenlistment Waiver

E1TERMS - Terms of Service or Enlistment

E1BASD - Basic Active Service Date

E1BONIN - Bonus Indicator

E1RPFLG - Recruiter Flag (Promoted or Separated)

E1RCRCD - Recruiter Code

E1PLOEN - State of Enlistment

E1TYPLA - Type of Last Accession

E1DATLA - Date of Last Accession

E1ETSDT - Date of Expiration of Last Term of Service

C. Basic Progress in the Army

E1GRTIT - Grade in Which Serving

E1DOR - Date of Rank

E1PAYGR - Paygrade

E1PAYSX - Paygrade & Sex

E1GRDDT - Date of Last Grade change

E1BPEDT - Basic Pay Entry Date

E1GRDTT - Type of Last Grade Change

E1NCOES - NCO Education System (Level Attained)

E1PROPT - Current Promotion Points

E1PROPDT - Current Promotion Point Date

E1PRVPT - Previous Promotion Points

E1PRVPDT - Previous Promotion Points Date

E1PROPA - Proficiency Pay Status

E1AITDT - AIT Graduation Date

E1PACE - Self-Paced AIT Flag

E1EERWA - EER Weighted Average

E1TUREL - Tour Eligibility

E1SECCLR - Personnel Security Clearance

E1SGTID - Drill Sergeant Qualificaiton

E1ADPAY - Eligibility for Additional Pay

E1VEAP - Veterans Education Assistance Program Code

D. Performance in a Particular MOS

E1CMF - Career Management Field

E1PRMOS - Primary MOS

E1DMOS - Duty MOS

E1SMOS3 - Secondary MOS Current (3-POS)

E1PMOTT - Type of Last PMOS Change

E1PMODT - Date of Last Change to PMOS

E1PGMOS - Primary Progression MOS

EiBOMOS - MOS of Bonus

E1DDSID - Additional Skill Indicator, Duty MOS

E1ADSID2 - Additional Skill Indicator, Previous

E1ADSID3 - Additional Skill Indicator, 2nd Previous

E1PQDES - Primary MOS, in which Tested, SQT Designator

E1PQSCR - Primary SQT Score (For PQDES)

E1PQPER - Skill Qualification Percentile (for PQDES)

E1PMOST - Primary MOST in Which Tested

E1PSQDT - Date of Last Change on PMOS Tested (SQT)

E1PMOST1 - Primary MOS in Which Tested, First Prior

E1PMOST2 - Primary MOS in Which Tested, Second Prior

E1PRQDT - Date of Previous Change in PMOS Tested

E1PRDES - Previous Primary MOS in Which Tested

E1PRQSC - SQT Score for Previous MOS (PQDES)

E1PRPER - Previous SQT Percentile (For PQDES)

E1SQDES - Secondary MOS SQT

E1SSQDT - SMOS SQT Date

E1SQSCR - SMOS SQT Score

E. Indicators of Attrition and Related Problems

E1CHSEP - Character of Separation

E1SPNIS - Separation Program Designator

APPENDIX B

DOCUMENTATION OF FY81 TRAINING DATA EDITING

PREVIOUS PAGE IS BLANK

DOCUMENTATION OF FY81 TRAINING DATA EDITING 8/2/83

These comments on the editing process for the training data apply to the original file of 89,889 records which yielded, after editing, 75,184 training records in the principal training data file. Exact duplicates have been eliminated, as well as many records that appeared to pertain to the same training "incident". In this file, there is more than one record for indiviuals who recycled or attrited in various other ways and for those who graduated from more than one course.

Before beginning the editing of the training data file, 1138 Social Security numbers (SSNs) in the training data were changed to conform to the SSN for that person in the accession file. The initial assumption was that the SSN in the accession file was the correct one. Subsequent investigation revealed that 56 of these SSN changes were in error, and the original SSN was restored for these cases.

The first task undertaken in the editing of the training data was the identification and elimination of duplicate records, where duplicate refers both to records that had the same values on every variable and records that represented the same training incident. The original file of 89,889 records was divided into two files: one file of 49,905 records contained those with only one training record per SSN, and the other contained all other records. Exact duplicates identical on every variable except name, which was not checked) were eliminated from the latter file, yielding a further 12,536 SSNs with only one training record. his left 14,912 records which had more than one occurrence of their SSN. Having been sorted by graduation date, this file of records was combined into 8096 pairs of records in graduation date order for visual examination, where the pairs were the "previous/next" training records for the individual. These 8096 pairs were further divided into three files for ease of handling: 5188 pairs that appeared to be genuine recycles (criteria: had a disposition of A, B, C, or E on the chronologically first record and had a second enrollment date that was later than the first graduation date); 1963 duplicates (criterion: no difference between the two graduation dates); and 945 pairs that did not fall into either of the two other categories.

The records that had initially been judged to be recycles in fact, upon closer examination, were recycles, and both of the records were retained in the file. Senior staff examined the records in the "DUPLICATES" file and indicated whether only one of the two records (and which one) or both records were to be kept. Typically, the differences between the two alleged duplicate records were small; for example, values for score may have appeared on only one of two records for the individual. The 945 pairs in the "OTHER" file were also inspected one by one. These intensive efforts to resolve cases in which an individual had more than one training record resulted in a file of 12,747 records which was added to the file of 62,441 with unique SSNs to yield 75,188 training data records records for further editing and analysis. Four cases were subsequently dropped since they lacked SSNs.

The 75,184 records in the edited training data file represent a total of 69,176 unique individuals. The table below indicates that 5 people have seven records in the training file, 17 have six records, and so forth. Note that 64,265 individuals are represented with only one training record.

NUM OF RECORDS	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	64265	64265	92.901	92.901
2	4130	68395	5.970	98.871
3	563	68958	0.814	99.685
4	147	69105	0.213	99.897
5	49	69154	0.071	99.968
6	17	69171	0.025	99.993
7	5	69176	0.007	100.000

EDITING OF DISPOSITION AND ATTRITION CODES

The disposition and attrition codes (variable names T1DISP and T1ATTRIT) were edited together since looking at them in tandem often suggested what the correct value should have been when one or the other was invalid. The procedure followed was to print records which had inconsistent values (for example, a disposition of H (graduate) used with a non-blank attrition code) as well as invalid codes for either attrition or disposition. Staff then inspected each of these records to resolve problems with the two variables. Information such as whether the individual had a course grade was thus used in determining, for example, whether an ambiguous record represented an attrition or a graduation.

We used documents supplied by ARI to determine which disposition codes were valid, and we then printed out all cases with other dispositions (blank, comma, J, K, N, R, S, W). Similarly, we relied on the documents supplied by ARI to determine what the valid attrition codes were, and we printed out for inspection all others (slash, G, H, I, S, T, U).

In general, invalid disposition codes used in combination with attrition codes were considered to be attritions, and therefore, the dispositions were changed to C (relief). Attrition codes were usually reset to blank when they appeared with disposition codes indicating successful graduation, especially if the individual had a test score. Dispositions of recycle (A,B) or relief (C) should have had non-blank attrition codes; however, when records with these dispositions lacked attrition codes, the attrition codes were left blank since no additional information was available to determine what the appropriate attrition code should have been.

The data from two schools (113 and 906) presented special problems. Although attrition codes of S, T, and U did not appear in documents supplied by ARI, they have been left on the file for school 113 since we verified with the school that these three attrition codes were used in combination with a disposition of C (relief). The values of these three attrition codes used at school 113 are:

- S = TDP (Trainee Discharge Program)
- T = EDP (Expeditious Discharge Program)
- U = AWOL (W was used for AWOL at other schools)

In addition, it appeared that school 906 put the code for disposition in the place reserved for attrition and used the disposition place to describe the next step in training. The editing rules for school 906 can be found in the table below.

Two new disposition codes were created for 12 unusual cases: a value of X when disposition, attrition, and score were all missing; and a value of Y when disposition and attrition were missing, but score was non-missing.

Here are the specific changes that resulted from the editing process:

# OF	OLD VALUES		NEW VALUES		
CASES	DISP	ATTRIT	DISP	ATTRIT	COMMENTS
4	blank	blank	X	blank	if score was missing
7	blank	blank	Y	blank	if score not missing
1	,	G	Y	blank	if score not missing
1	,	H	X	blank	if score missing
1	A	/	A	blank	
1	A	G	G	blank	special rule: school 906
1	A	S	A	S	special rule: school 113
20	A	S	A	S	special rule: all school 906
2	C	H	C	blank	
1	C	I	C	blank	
1	C	S	C	blank	school 804
422	C	S	C	S	special rule: all school 113
41	C	T	C	T	special rule: all school 113
18	C	U	C	U	special rule: all school 113
8	E	A,Q	C	A,Q	
69	E	H	H	blank	special rule: all school 906
1	E	O	E	blank	
	E	R	C	R	all school 811 cases with no score
	E	R	H	blank	all school 811 cases with valid scores
2	H	A	H	blank	score non-missing
1	H	A	C	A	score missing
1	H	C	H	blank	score non-missing
2	H	C	C	C	score missing
3	H	H	H	blank	score non-missing
1	H	J	H	blank	score non-missing
1	H	O	H	blank	score non-missing
1	H	R	H	blank	score non-missing
1	H	S	H	blank	score non-missing
2	J	A	C	A	score missing
1	J	J	C	J	score missing
2	K	A	C	A	score missing
1	K	K	C	K	score missing
1	N	blank	blank	blank	
1	N	blank	H	blank	score non-missing
2	N	A	C	A	score missing
2	R	A	C	A	score missing
3	S	A	C	A	score missing
1	W	W	C	W	score missing

After this editing was completed, it was decided that we needed to differentiate between those records which lacked attrition codes but should have had them and records which properly did not have attrition codes. An attrition code of Z was assigned to those with dispositions of E, F, G, and H (graduates, all of whom had blank attrition codes, N=62077), and Y was assigned to all others who had blank attrition codes (N=62). This left 13,045 records with other attrition codes.

EDITING OF THE CODES FOR SCHOOL, MOS, AND COURSE

The school, MOS, and course codes (variable names T1SCHOOL, T1MOSAWD, AND T1COURSE, respectively) were edited as a group since determining the correct value for one of the three variables usually depended on knowing the values for the other two variables.

The changes to these three variables were minor, and most could be characterized as corrections to typos.

The procedure followed was to generate a crosstabulation of MOS by course for each school separately. Staff then compared the distribution of MOS/course combinations with the list of MOS/course combinations appearing for each school in documents supplied by ARI of the training data. Distributions of the classes and their modal enrollment and graduation dates within each school/MOS/course combination were used to provide additional information when necessary.

Here are the specific changes that resulted from the editing process:

# OF	OLD VALUES			NEW VALUES			
CASES	SCHL	MOS	CRSE	SCHL	MOS	CRSE	COMMENTS
1	029	91E	05	929	91E	05	school 029 invalid
2	061	12B	7B	061	17B	7B	12B/7B invalid 061/12B invalid
50	061	45D	ID	061	45D	1D	typo: ID
31	113	O5B	2A	113	05B	2A	typo: O5B
176	113	O5C	2D	113	05C	2D	typo: O5C
7	121	blnk	5M	121	74F	5M	typo: blnk
14	121	71D	6A	121	74D	6A	71D/6A invalid
1	803	09J	4C	803	64C	4C	09J invalid for school
1	805	63V	3B	805	63B	3B	63V invalid for school
1	805	76	6Y	805	76Y	6Y	typo: 76
1	805	76Y	6I	805	76Y	6Y	typo: 6I
10	807	62H	C6	807	62H	CG	typo: C6 verified with dates
8	809	51N	BN	807	51N	BN	51N at school 807
93	810	13F	34	810	13F	3F	typo: 34
1	810	13F	3E	810	13F	3F	typo: 3E verified with dates
2	811	16E	HA	811	16H	HA	typo: 16E verified with dates
5	835	75E	5E	805	75E	5E	school 835 invalid
6	906	O5D	HD	906	05D	HD	typo: O5D

The resulting school/MOS/course combinations in the file diverged somewhat from the original documents supplied by ARI. Data for the following schools conformed to the MOS/course lists in documents supplied by ARI: 031, 061, 121, 161, 171, 301, 441, 809, 810, 906, and 929.

Differences are summarized below:

Documents supplied by ARI			Training data file		
SCHOOL	MOS	COURSE	SCHOOL	MOS	COURSE
011	67N	7N	011	67N	65
011	67V	7V	011	67V	18
011	71P	1P	011	71P	82
011	93H	3H	011	93H	78
011	93J	3J	011	93J	79
051			no data in training file		
091	41C	G7	091	41C	G7 & G8
091	45K	K9	091	45K	K8 & K9
091	45L	L3	091	45L	L2 & L3
093	24H		no data in training file		
093	24J		no data in training file		
093	24L		no data in training file		
101	43E	ED	101	43E	AR & ED
101	92C	DA	101	92C	DA & PC
113	26V	8D	113	26V	5E
551	67V		no data in training file		
803	63B	3B	803	63B	3B & 6B
804	19K		no data in training file		
804	19L		no data in training file		
805	71L		no data in training file		
not in ARI doc			807	51K	BB & BK
not in ARI doc			807	51M	BM
807	62F	CF	807	62F	CF & CH
807	62H	CH	807	62H	CG, CH, & EC
811	16J	JA/JB	811	16J	JA only

EDITING OF CODES FOR SKILL LEVEL

The four editing changes to the codes for skill level (variable name T1SKLLVL) can all be characterized as corrections to typos:

Old Value	New Value	
1O	10	
YO	10	4 cases
Y1	blank	1 case
11	10	1 case

EDITING OF ENROLLMENT AND GRADUATION DATES

The enrollment and graduation dates in the training data file should be approached with caution by researchers. In almost all cases, there continues to be variation in the enrollment and graduation dates of soldiers who graduated from the same class within a course. Further, for non-graduates, it is not clear whether the graduation date given is the date on which the soldier left the class or the date on which normal graduation would have taken place.

The editing changes to the enrollment and graduation dates (variable names T1ENRDTE and T1GRDDTE) were fairly extensive. A list of modal (most frequently occurring) enrollment and graduation dates for each school/MOS/course/class combination was generated; this list also indicated the first and third quartiles for the dates. The modal dates were used extensively to determine what the correct dates should have been whenever editing procedures indicated that there was a problem.

First, typos for nine cases with obviously incorrect dates were corrected. These records had enrollment dates that were later than 1 July 1982 or graduation dates that were later than 1 January 1983. The changes are here summarized:

# of Cases	Schl/MOS/Crse	Old Value (year) T1ENRDTE	T1GRDDTE	CHANGE
1	807/51R/BR	81	86	86 changed to 81
1	805/63B/3B	84	81	84 changed to 81
2	805/63B/3B	81	87	87 changed to 81
1	161/71M/71	89	81	89 changed to 81
2	011/71P/82	Dec 82	Apr 82	12/82 changed to 12/81
1	929/91C/02	91	81	91 changed to 81
1	929/91Q/15	91	81	91 changed to 81

In addition, schools 807 and 803 had a large number of enrollment and graduation dates that were in the early 1900's, which were obvious typos. So for school 807, the 772 enrollment dates of 1901, the 1 enrollment date of 1918, the 1 graduation date of 1910, and the 752 graduation dates of 1901 were all changed to 1981. The choice of the year 1981 was dictated by the other variables on these records. Similarly, the 25 graduation dates at school 803 in the year 1901 were changed to 1981.

Other schools had a few similar typos in their dates and these too were corrected as follows:

Number	School/MOS/Course	ORIGINAL T1ENRDTE	T1GRDDTE	CHANGE
N=1	805/63B		20MAR10	20MAR81
all Class=10	805/63B			T1ENRDTE=10JAN81
N=1	805/63B		1908	1980
N=1	810/82C		1911	1981
N=1	811/16B	1918		1981
N=1	929/91C	1901		1981
N=1	929/91G		1971	1981

In addition, for MOS 31M, a number of entries had graduation dates that came chronologically before the enrollment dates; these all were non-graduates with an attrition code of 0 (administrative error). The graduation dates of these 22 cases were all set to missing. Fifteen graduates who were in class 291 of course 3E (MOS 13E at school 810) had an erroneous enrollment date of 7 August 1981, which was later than their graduation date of 24 July 1981. The enrollment date was changed to 26 May 1981, which was judged to be correct after looking at the modal values for the classes. Similarly, 30 graduates of class 17 of course EC (MOS 64C at school 807) had an enrollment date of 20 March 1981, which was after their graduation date. The enrollment date was changed to 20 February 1981.

During the next phase of the editing of the enrollment and graduation dates, the data were screened based upon the difference (in number of days) between the graduation date and the enrollment date. We then printed any records that had a negative difference (graduation date before enrollment date) or a difference greater than 180 days (training lasted longer than 6 months). Many of the long training events appeared to be legitimate, but a total of 84 changes to individual records were made. These changes were decided upon by taking into consideration the modal dates for the School/MOS/course/class combination, by noting the disposition code for the soldier, and by comparing the length of training time with that for other classes in the same course/MOS/school in the training data.

The list below indicates these individual changes (ID information has not been printed here):

	Old Values		New Values	
SCHOOL/MOS/COURSE/CLASS	T1ENRDTE	T1GRDDTE	T1ENRDTE	T1GRDDTE
011/67V/18/024	26MAR80	31JUL81	26MAR81	31JUL81
011/71P/82/020	19FEB82	28JAN82	19FEB82	22APR82
011/71P/82/020	04FEB82	19FEB82	19FEB82	22APR82
061/17B/7B/171	10FEB81	23JAN81	10FEB81	24MAR81
061/17C/7C/461	01SEP81	29AUG81	01SEP81	02OCT81
061/17C/7C/461	01SEP81	28AUG81	01SEP81	02OCT81
091/63H/H1/999	04JAN81	18MAR82	04JAN81	18MAR81
093/27E/7E/026	05AUG81	03DEC82	05AUG81	03DEC81
093/27E/7E/026	05AUG81	03DEC82	05AUG81	03DEC81
093/27E/7E/008	27JUN80	05JUN81	07JAN82	04MAY82
113/32D/80/021	21SEP81	18SEP81	21SEP81	16NOV81
113/32D/80/021	21SEP81	18SEP81	21SEP81	16NOV81
113/32D/80/021	21SEP81	18SEP81	21SEP81	16NOV81
113/32D/80/021	21SEP81	18SEP81	21SEP81	16NOV81
113/32D/80/021	21SEP81	18SEP81	21SEP81	16NOV81
113/32D/80/016	13JUL81	05MAY81	13JUL81	28OCT81
113/32D/80/021	21SEP81	18SEP81	21SEP81	16NOV81
161/71M/71/006	20JAN81	19JAN81	20JAN81	23FEB81
161/71M/71/010	11JUL81	19MAY81	02APR81	19MAY81
161/71M/71/502	16APR81	08APR81	16APR81	15MAY81
301/17K/GA/009	06DEC81	18MAR81	06FEB81	18MAR81
301/96B/CE/505	24APR81	16APR81	24APR81	26JUN81
301/96B/CE/503	29APR81	27FEB81	06FEB81	27FEB81

SCHOOL/MOS/COURSE/CLASS	T1ENRDTE	T1GRDDTE	T1ENRDTE	T1GRDDTE
803/94B/4B/	16MAR81	05MAR81	16MAR81	05JUN81
803/94B/4B/	28SEP81	16SEP81	28SEP81	16DEC81
803/94B/4B/	16MAR81	05MAR81	16MAR81	05JUN81
803/94B/4B/	16MAR81	04MAR81	16MAR81	05JUN81
803/94B/4B/	16MAR81	01MAR81	16MAR81	05JUN81
803/94B/4B/	16MAR82	06MAR81	16MAR81	05JUN81
803/94B/4B/	16MAR81	05MAR81	16MAR81	05JUN81
803/94B/4B/	16MAR81	05MAR81	16MAR81	05JUN81
805/63B/3B/010	10JAN81	13MAR82	10JAN81	13MAR81
805/63B/3B/026	15APR81	12APR81	15APR81	06JUL81
805/63B/3B/005	05FEB80	24FEB81	01DEC80	24FEB81
805/76Y/6Y/050	29SEP81	16NOV80	29SEP81	16NOV81
805/76Y/6Y/041	09SEP81	23JUL81	09SEP81	20OCT81
805/94B/4B/037	29JUN81	06JUN81	29JUN81	04AUG81
805/94B/4B/024	06APR81	03SEP82	06APR81	03SEP81
807/12F/AF/014	26APR80	01MAY81	16MAR81	01MAY81
807/12F/AF/016	26APR80	12MAY81	30MAY81	12MAY81
807/64C/EC/029	04MAY81	12JUN82	04MAY81	12JUN81
807/64C/EC/017	20FEB81	13FEB81	20MAR81	.
807/64C/EC/017	20FEB81	11FEB81	20MAR81	.
810/13B/3B/131	09JAN81	03JAN81	09JAN81	25MAR81
810/13B/3B/203	16OCT81	27OCT81	16OCT81	21JAN82
810/13B/3B/203	16OCT81	19JUL81	16OCT81	21JAN82
810/13B/3B/204	23OCT81	27FEB81	23OCT81	27FEB82
810/13B/3B/202	09OCT81	13JAN81	09OCT81	13JAN82
810/13E/3E/181	06MAR81	07APR80	06MAR81	07APR81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
810/82C/2C/291	01AUG81	16JUL81	25MAY81	16JUL81
810/82C/2C/291	01AUG81	08JUN81	25MAY81	08JUN81
811/16D/DA/019	30JUL80	16OCT81	30JUL81	16OCT81
811/16D/DB/017	12JUN81	30JUL80	12JUN81	30JUL81
813/95B/SB/027	26JUN81	08JUN81	24APR81	08JUN81
906/05G/DP/006	04MAY81	06APR81	04MAY81	03SEP81
906/05G/DP/006	04MAY81	06APR81	04MAY81	03SEP81
906/05H/DT/501	20NOV81	21OCT81	11SEP81	21OCT81
906/05H/DT/019	23JUN81	01JUN81	20JUL81	04AUG81
906/05H/DT/003	11FEB81	17MAR82	19OCT81	17MAR82
906/05H/DT/502	11JAN82	11FEB81	05OCT81	11FEB82
906/05H/DT/501	26JUL81	15JUL81	11SEP81	09DEC81
906/05K/PF/031	02MAR81	13JAN82	17AUG81	13JAN82
906/05K/PF/031	02MAR81	14JAN82	17AUG81	14JAN82
906/05K/PF/035	06APR81	16FEB82	30AUG81	16FEB82
906/33S/CY/004	16DEC81	18NOV81	16OCT81	18NOV81
906/98C/GF/078	12MAR82	28APR81	12MAR81	28APR81
929/91B/01/002	25OCT81	08DEC82	25OCT81	08DEC81
929/91B/01/001	13OCT81	23NOV80	13OCT81	23NOV81
929/91B/01/002	25OCT81	30AUG81	25OCT81	08DEC81
929/91D/04/004	29MAR80	25JUN81	29MAR81	25JUN81
929/91F/07/001	02NOV81	17FEB81	02NOV81	17FEB82
929/91P/14/004	13FEB81	01FEB81	13FEB81	16MAY81
929/91R/16/009	26JUL80	18AUG81	25JUL81	18AUG81
929/94F/27/029	24AUG81	10JUL81	24AUG81	10NOV81

There were 45 non-graduates who had graduation dates earlier than their enrollment dates, typically differing by about a week. For these cases, the graduation and enrollment dates were reversed since inspection revealed that this procedure would still give dates within the range of dates for the class.

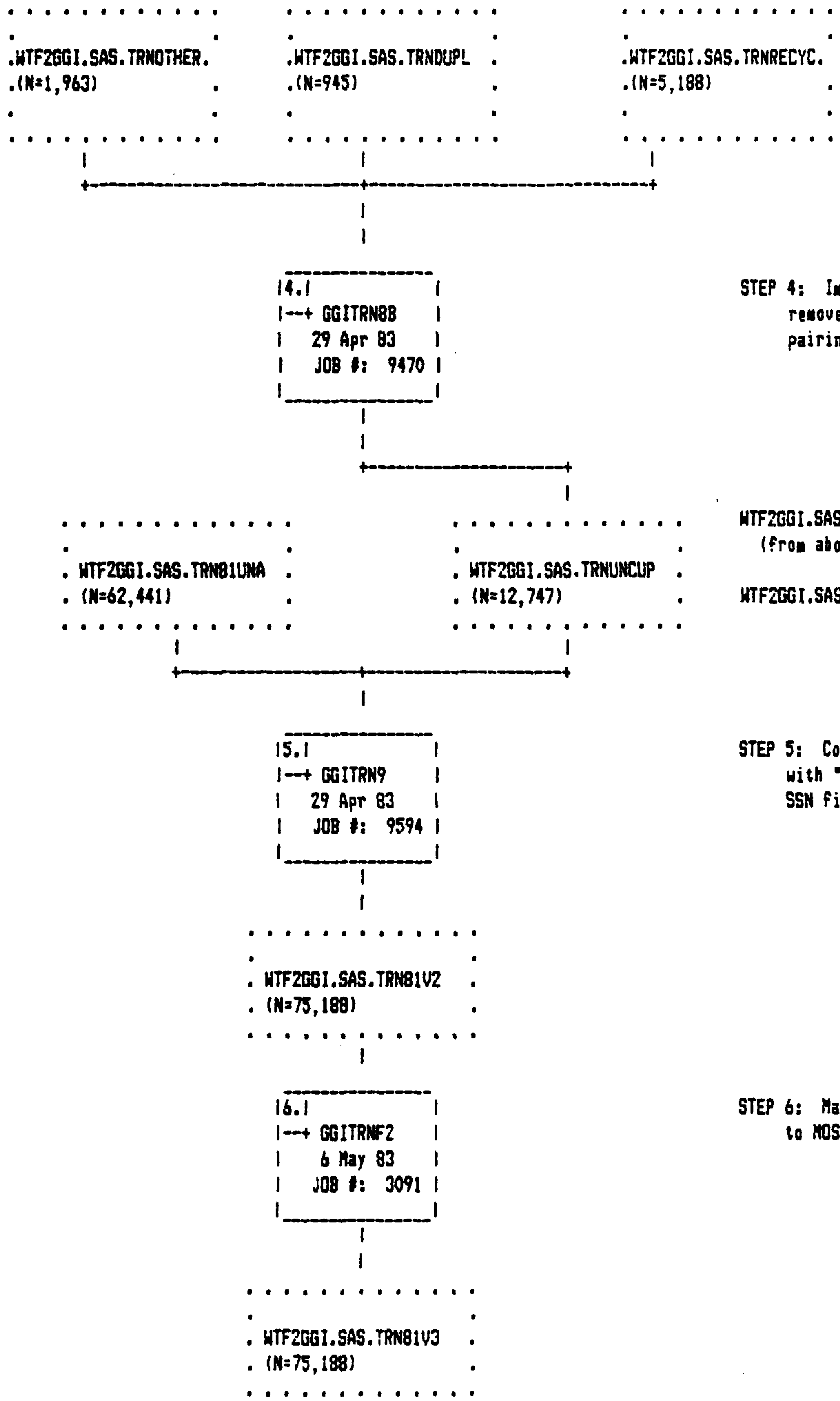
Some few graduates lacked either an enrollment date or a graduation date. For these cases, the missing dates were replaced with the modal date for that class/course. These can be summarized as follows:

Number	School/MOS/course/class	New Date
N=16	804/19D/9D/506	T1ENRDTE=20MAR81
N=1	807/51B/BB/012	T1ENRDTE=02MAR81
N=1	807/51C/BC/023	T1ENRDTE=18MAY81
N=1	929/91B/01/002	T1ENRDTE=25OCT81
N=1	805/63B/3B/021	T1GRDDTE=01MAY81
N=1	805/63B/3B/025	T1GRDDTE=20JUN81
N=1	805/76Y/6Y/042	T1GRDDTE=11SEP81
N=1	810/13E/3E/141	T1GRDDTE=10APR81
N=1	929/91B/01/021	T1GRDDTE=11SEP81

Neither enrollment nor graduation dates exist for classes 18 through 29 of course AF (MOS 12F at school 807); they were absent in the original training data file. It would be a straightforward task to insert plausible dates since class 17 and class 30 of course AF have dates, and the classes appear to have started weekly.

NOTE: What follows is a program flow chart showing each step in the editing process. For each step, the program name, date run, and job number are shown, along with a brief description of the purpose of the run. The intermediate file names and record counts are also shown.

```
. . . . . . . . . . . . .          . . . . . . . . . . . . .     WTF2WYO.TRAIN81.SAS - Combined training
.                       .          .                       .         data
. WTF2WYO.TRAIN81.SAS   .          . WTF2WYO.TRNSSN.SAS    .
. (N=89,889)            .          . (N=1,138)             .     WTF2WYO.TRNSSN.SAS - SSN corrections
. . . . . . . . . . . . .          . . . . . . . . . . . . .
            |                                  |
            +----------------+-----------------+
                             |
                      ________________
                     |1.|             |                          STEP 1:  Correct the SSNs, listing
                     |--+ GGITRN2A    |                               cases where there were duplicate
                     |   30 Mar 83    |                               records with the same correction.
                     |   JOB #:  7202 |
                     |________________|
                             |
                 . . . . . . . . . . . . .
                 .                       .
                 . WTF2GGI.SAS.TRAIN81A  .
                 . (N=89,889)            .
                 . . . . . . . . . . . . .
                             |
                      ________________
                     |2.|             |                          STEP 2:  Divide file into those with
                     |--+ GGITRDUP    |                               unique SSNs and those with more
                     |   1 Apr 83     |                               more than 1 record per SSN.
                     |   JOB #:  5085 |                               Output pairs of "prior" and "previous"
                     |________________|                               entries for those with duplicate SSNs.
                             |
            +----------------+-----------------+
            |                                  |
. . . . . . . . . . . . .          . . . . . . . . . . . . .     WTF2GGI.SAS.TRN81UNA - Unique SSNs
.                       .          .                       .
. WTF2GGI.SAS.TRN81UNA  .          . WTF2GGI.SAS.TRN81DUP  .
. (N=62,441)            .          . (N=8,096)             .     WTF2GGI.SAS.TRN81DUP - duplicate SSNs
. . . . . . . . . . . . .          . . . . . . . . . . . . .
                                               |
                             +-----------------+
                             |
                      ________________
                     |3.|             |                          STEP 3:  Attach "Keep/Del" flags to data
                     |--+ GGITRN7     |                               pairs in all 3 subsets of data with
                     |   28 Apr 83    |                               duplicate SSNs and output 3 subsets.
                     |   JOB #:  3890 |
                     |________________|
                             |
      +----------------------+------------------------+
      |                      |                        |
. . . . . . . . . . . .  . . . . . . . . . . . .  . . . . . . . . . . . .
.                    .   .                    .   .                      .
.WTF2GGI.SAS.TRNOTHER.   .WTF2GGI.SAS.TRNDUPL .   .WTF2GGI.SAS.TRNRECYC.
.(N=1,963)           .   .(N=945)             .   .(N=5,188)             .
.                    .   .                    .   .                      .
. . . . . . . . . . . .  . . . . . . . . . . . .  . . . . . . . . . . . .
```



STEP 4: Implement "Keep/Delete" flags to remove duplicates and undo earlier pairing.

WTF2GGI.SAS.TRN81UNA - Unique SSNs (from above)

WTF2GGI.SAS.TRNUNCUP - dupl deleted

STEP 5: Concatenate unique SSN file with "unpaired" non-unique SSN file.

STEP 6: Make minor editing changes to MOS, school, and course

```
                    . . . . . . . . . . . . .
                    .                       .
                    . WTF2GGI.SAS.TRN81V3   .
                    . (N=75,188)            .
                    . . . . . . . . . . . . .
                              |
                              |
                     _________________
                    |7.1              |
                    |--+ GGITRNFB     |
                    |   18 May 83     |
                    |   JOB #:  2044  |
                    |_________________|
                              |
                              |
                    . . . . . . . . . . . . .
                    .                       .
                    . WTF2GGI.SAS.TRN81V3B  .
                    . (N=75,188)            .
                    . . . . . . . . . . . . .
                              |
                              |
                     _________________
                    |8.1              |
                    |--+ GGITRNFD     |
                    |   18 May 83     |
                    |   JOB #:   613  |
                    |_________________|
                              |
                              |
          +-------------------+-------------------+
          |                                       |
. . . . . . . . . . . . .             . . . . . . . . . . . . .
.                       .             .                       .
. WTF2GGI.SAS.TRN81V3C  .             . WTF2GGI.SAS.NUSCORE   .
. (N=69,365)            .             . (N=5,823)             .
. . . . . . . . . . . . .             . . . . . . . . . . . . .
          |                                       |
          +-------------------+-------------------+
                              |
                              |
                     _________________
                    |9.1              |
                    |--+ GGITRNFE     |
                    |   20 May 83     |
                    |   JOB #:  4785  |
                    |_________________|
                              |
                              |
                    . . . . . . . . . . . . .
                    .                       .
                    . WTF2GGI.SAS.TRN81V3D  .
                    . (N=75,184)            .
                    . . . . . . . . . . . . .
```

STEP 7: Make more minor editing changes to MOS, school, and course & convert rank to numeric var

STEP 8: Concatenate T1SCORE and T1RANK for selected MOS's to produce accurate value of score variable

WTF2GGI.SAS.TRN81V3C - not need new score

WTF2GGI.SAS.NUSCORE - concatenate 2 parts of score

STEP 9: Correct attrition and disposition codes and delete 4 cases with missing SSN

```
. . . . . . . . . . . . .
.                       .
. WTF2GGI.SAS.TRN81V3D  .
. (N=75,184)            .
. . . . . . . . . . . . .
           |
   _______________
  |10.|           |        STEP 10: Further corrections to attrition
  |---+GGITRFGJ   |            and disposition codes and corrections
  |    1 Jun 83   |            to obvious problems with enrollment
  |   JOB #: 9538 |            and graduation dates
  |_______________|
           |
. . . . . . . . . . . . .
.                       .
. WTF2GGI.SAS.TRN81V3E  .
. (N=75,184)            .
. . . . . . . . . . . . .
           |
   _______________
  |11.|           |        STEP 11: Add 3 score type variables
  |---+GGITRFGN   |            and correct some scores, as well as
  |    7 Jun 83   |            set up 2nd score variable for MOS's
  |   JOB #: 2210 |            that had 2 scores
  |_______________|
           |
. . . . . . . . . . . . .
.                       .
. WTF2GGI.SAS.TRN81V3F  .
. (N=75,184)            .
. . . . . . . . . . . . .
           |
   _______________
  |12.|           |        STEP 12: Change some values of 2 score
  |---+GGITRFGV   |            type variables, based on further data,
  |   11 Jul 83   |            also fix many enroll & grad dates and
  |   JOB #: 2255 |            insert modal dates for grads
  |_______________|            with missing dates
           |
           |
. . . . . . . . . . . . .
.                       .
. WTF2GGI.SAS.TRN81V3G  .
. (N=75,184)            .
. . . . . . . . . . . . .
           |
   _______________
  |13.|           |        STEP 13: Make 1 more correction to score
  |---+GGITRFGX   |            types and correct 2 minor errors in
  |   14 Jul 83   |            other variables
  |   JOB #: 4716 |
  |_______________|
```

```
 _______________
|13.|           |        STEP 13:  Make 1 more correction to score
|---+GGITRFQX   |            types and correct 2 minor errors in
|   14 Jul 83   |            other variables
|   JOB #:  4716|
|_______________|
        |
        |
. . . . . . . . . . . . .
.                       .
. WTF2GGI.SAS.TRN81V3H  .
. (N=75,184)            .
. . . . . . . . . . . . .
        |
        |
 _______________
|14.|           |        STEP 14:  Make corrections for training
|---+WYOTRN4A   |                  data SSN's, make scrambled ID,
|   14 Jul 83   |                  & drop other identifying data.
|   JOB #:  5277|
|_______________|
        |
        |
. . . . . . . . . . . . .
.                       .
. WTF2GGI.SAS.TRN81V4A  .
. (N=75,184)            .
. . . . . . . . . . . . .
        |
        |
 _______________
|15.|           |        STEP 15:  Merge accessions file and training
|---+WYOACTRN   |                  data; Keep multiple training
|   14 Jul 83   |                  records if match accessions data
|   JOB #:  5283|
|_______________|
        |
        |
. . . . . . . . . . . . .
.                       .
. WTF2GGI.SAS.ACCTRNV1  .
. (N=52,559)            .
. . . . . . . . . . . . .
```

APPENDIX C

GRADUATION RATES FOR TRAINING COURSES WITH AT LEAST 100 STUDENTS

Training Graduation Rates fro Courses With at Least 100 Students
for FY 81 Nonprior-Service Accessions

End cf Course Disposition
FREQUENCY
ROW PWECENT

MOS COURSE	GRAD, SCORE	GRAD, NO SCORE	COURSE FAILURE	DSCP, AW OL, DSRT	NONACADE MIC ATT	UNKNOWN STATUS	TOTAL
05B2A	514 92.28	0 0.00	3 0.54	0 0.00	40 7.18	0 0.00	557
05C2D	563 79.30	0 0.00	42 5.92	0 0.00	105 14.79	0 0.00	710
11BIN	920 91.36	0 0.00	56 5.56	15 1.49	16 1.59	0 0.00	1007
11CIN	534 92.87	0 0.00	23 4.00	5 0.87	13 2.26	0 0.00	575
11HIN	406 91.86	0 0.00	22 4.98	1 0.23	13 2.94	0 0.00	442
11OIN	309 90.35	0 0.00	17 4.97	4 1.17	12 3.51	0 0.00	342
12BAB	129 94.16	0 0.00	2 1.46	1 0.73	5 3.65	0 0.00	137
12FAF	195 87.44	0 0.00	3 1.35	3 1.35	22 9.87	0 0.00	223
13B3B	611 91.33	0 0.00	28 4.19	4 0.60	26 3.89	0 0.00	669
13E3E	411 88.96	0 0.00	38 8.23	1 0.22	12 2.60	0 0.00	462
13F3F	608 87.11	0 0.00	51 7.31	1 0.14	38 5.44	0 0.00	698
15D5D	271 88.27	0 0.00	10 3.26	0 0.00	26 8.47	0 0.00	307
15E5E	260 88.44	0 0.00	4 1.36	2 0.68	28 9.52	0 0.00	294
16BBA	96 78.05	0 0.00	15 12.20	0 0.00	12 9.76	0 0.00	123
16EEB	109 87.20	0 0.00	8 6.40	1 0.80	7 5.60	0 0.00	125

Training Graduation Rates for Courses With at Least 100 Students
for FY 81 Nonprior-Service Accessions (Continued)

End of Course Disposition
FREQUENCY
ROW PWECENT

MOS COURSE	GRAD, SCORE	GRAD, NO SCORE	COURSE FAILURE	DSCP, AW OL, DSRT	NONACADE MIC ATT	UNKNOWN STATUS	TOTAL
16RRA	226 66.28	0 0.00	70 20.53	6 1.76	39 11.44	0 0.00	341
16SSA	472 84.89	0 0.00	34 6.12	5 0.90	45 8.09	0 0.00	556
17C7C	177 90.77	0 0.00	12 6.15	4 2.05	2 1.03	0 0.00	195
17KGA	121 88.32	0 0.00	7 5.11	2 1.46	6 4.38	1 0.73	137
19D9D	181 90.50	0 0.00	5 2.50	2 1.00	12 6.00	0 0.00	200
19E9E	158 88.76	0 0.00	8 4.49	4 2.25	8 4.49	0 0.00	178
19F9F	115 87.12	0 0.00	8 6.06	3 2.27	6 4.55	0 0.00	132
27E7E	113 76.35	0 0.00	14 9.46	0 0.00	19 12.84	2 1.35	148
31M4D	408 60.18	0 0.00	178 26.25	1 0.15	91 13.42	0 0.00	678
31N4C	159 86.41	0 0.00	0 0.00	2 1.09	23 12.50	0 0.00	184
31V1V	406 86.57	0 0.00	51 10.87	12 2.56	0 0.00	0 0.00	469
32D80	36 23.08	0 0.00	80 51.28	0 0.00	40 25.64	0 0.00	156
36CAA	176 83.02	0 0.00	3 1.42	1 0.47	32 15.09	0 0.00	212
36KAC	641 93.58	0 0.00	18 2.63	2 0.29	24 3.50	0 0.00	685
45KK9	108 82.44	0 0.00	13 9.92	0 0.00	9 6.87	1 0.76	131

Training Graduation Rates for Courses With at Least 100 Students
for FY 81 Nonprior-Service Accessions (Continued)

End of Course Disposition
FREQUENCY
ROW PWECENT

MOS COURSE	GRAD, SCORE	GRAD, NO SCORE	COURSE FAILURE	DSCP, AW OL, DSRT	NONACADE MIC ATT	UNKNOWN STATUS	TOTAL
51KBK	107 67.72	0 0.00	48 30.38	1 0.63	2 1.27	0 0.00	158
54CSS	117 89.31	0 0.00	7 5.34	2 1.53	5 3.82	0 0.00	131
55B5B	186 90.73	1 0.49	4 1.95	0 0.00	12 5.85	2 0.98	205
57HG1	169 94.94	0 0.00	0 0.00	1 0.56	8 4.49	0 0.00	178
61CH1	103 75.74	0 0.00	20 14.71	3 2.21	10 7.35	0 0.00	136
62BCB	214 94.69	1 0.44	6 2.65	2 0.88	3 1.33	0 0.00	226
63B3B	834 89.20	3 0.32	62 6.63	11 1.18	15 1.60	10 1.07	935
63DSA	268 98.89	0 0.00	0 0.00	0 0.00	3 1.11	0 0.00	271
63GM7	100 99.01	0 0.00	0 0.00	0 0.00	1 0.99	0 0.00	101
63HH1	215 97.73	0 0.00	2 0.91	0 0.00	3 1.36	0 0.00	220
63NTS	268 95.37	0 0.00	6 2.14	0 0.00	7 2.49	0 0.00	281
63TFI	467 98.32	0 0.00	5 1.05	0 0.00	3 0.63	0 0.00	475
63WW1	276 100.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	276
63YTV	129 98.47	0 0.00	1 0.76	0 0.00	1 0.76	0 0.00	131
64CEC	198 97.54	0 0.00	1 0.49	1 0.49	3 1.48	0 0.00	203

Training Graduation Rates for Courses With at Least 100 Students for FY 81 Nonprior-Service Accessions (Continued)

End of Course Disposition
FREQUENCY
ROW PWECENT

MOS COURSE	GRAD, SCORE	GRAD, NO SCORE	COURSE FAILURE	DSCP, AW OL, DSRT	NONACADE MIC ATT	UNKNOWN STATUS	TOTAL
64C4C	398 100.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	398
67N65	149 92.55	0 0.00	1 0.62	6 3.73	5 3.11	0 0.00	161
67UP1	175 95.11	0 0.00	0 0.00	0 0.00	9 4.89	0 0.00	184
67V18	151 92.64	0 0.00	2 1.23	6 3.68	4 2.45	0 0.00	163
67YS1	111 82.84	0 0.00	11 8.21	1 0.75	11 8.21	0 0.00	134
68JW6	95 77.24	0 0.00	7 5.69	0 0.00	21 17.07	0 0.00	123
68MW8	94 83.93	0 0.00	9 8.04	1 0.89	8 7.14	0 0.00	112
71NL1	103 84.43	0 0.00	14 11.48	1 0.82	4 3.28	0 0.00	122
73C5R	193 98.47	0 0.00	1 0.51	0 0.00	2 1.02	0 0.00	196
75B5E	456 93.06	0 0.00	27 5.51	0 0.00	7 1.43	0 0.00	490
75D5D	172 74.46	0 0.00	56 24.24	0 0.00	3 1.30	0 0.00	231
75E5E	191 68.71	0 0.00	77 27.70	3 1.08	7 2.52	0 0.00	278
76CEC	1035 87.94	0 0.00	102 8.67	10 0.85	30 2.55	0 0.00	1177
76P5F	548 97.51	0 0.00	9 1.60	0 0.00	5 0.89	0 0.00	562
76VEV	362 100.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	362

Training Graduation Rates for Courses With at Least 100 Students
for FY 81 Nonprior-Service Accessions (Continued)

End of Course Disposition
FREQUENCY
ROW PWECENT

MOS COURSE	GRAD, SCORE	GRAD, NO SCORE	COURSE FAILURE	DSCP, AWOL, DSRT	NONACADEMIC ATT	UNKNOWN STATUS	TOTAL
76WDB	131 98.50	0 0.00	1 0.75	0 0.00	1 0.75	0 0.00	133
76WPW	204 99.51	0 0.00	0 0.00	0 0.00	1 0.49	0 0.00	205
76YEY	355 90.56	0 0.00	22 5.61	0 0.00	15 3.83	0 0.00	392
76Y5G	282 91.56	0 0.00	15 4.87	3 0.97	8 2.60	0 0.00	308
76Y6Y	411 87.26	1 0.21	49 10.40	2 0.42	8 1.70	0 0.00	471
82C2C	319 83.95	0 0.00	50 13.16	3 0.79	7 1.84	1 0.26	380
91B01	685 93.96	1 0.14	38 5.21	1 0.14	4 0.55	0 0.00	729
91C02	202 85.59	0 0.00	18 7.63	4 1.69	12 5.08	0 0.00	236
91E05	146 94.19	0 0.00	8 5.16	0 0.00	1 0.65	0 0.00	155
92B25	103 84.43	0 0.00	18 14.75	0 0.00	1 0.82	0 0.00	122
94BKA	615 94.47	0 0.00	6 0.92	0 0.00	30 4.61	0 0.00	651
94B4B	591 92.49	0 0.00	34 5.32	7 1.10	7 1.10	0 0.00	639
95BSB	701 76.86	0 0.00	15 1.64	18 1.97	178 19.52	0 0.00	912
TOTAL	22062 88.16	7 0.03	1575 6.29	169 0.68	1196 4.78	17 0.07	25026 100.00

* Six categories of End of Course Disposition are defined: Graduation with score; graduation but no score recorded; course failure because of lack of ability, motivation, or skills; attrition because of desertion, AWOL, or disciplinary problem; attrition for other nonacademic reasons (e.g., hospitalization or recall by control organization); unknown status (no information available).

APPENDIX D

TRAINING MEASUREMENT INTERVIEW GUIDE

PROJECT A

Interview Record

Post ______________________________ Date ____________

School ______________________________

Course No. and Title ______________________________

Army Personnel: ______________________________
Name/Section/Position

____

____

____

HumRRO Personnel ____________________ Phone No. ____________
(Civilian)

____________________ ____________

Before starting the interview, check each of the following if present, or enter NA if not used in this course: POI ____, Lesson Plans ____, Course Manager's Guide ____, Facilitator's Guide ____, Instructor's Guide ____, Students's Guide ____, All Tests ____, Answer Sheets or Checklists ____, Blank Student Record Forms ____, and any available documents regarding test development ____.

(After introductions, read the following)

We work for the Human Resources Research Organization or HumRRO. I would like to explain the reason for our visit. The Assistant Secretary of Defense has directed all services to pursue a long-range program to increase the efficiency of the Armed Services Vocational Aptitude Battery (ASVAB) and enlistment standards in predicting performance in training and on the job.

In response to this directive, the Army Research Institute for the Behavioral and Social Sciences has contracted our organization to gather information on Army training programs and to develop new assessment procedures for selected programs.

As part of this effort we are interested in getting information about the different ways student progress is measured in various training courses throughout the Army. For now we want to focus only on the information you have concerning this course. Your answers will help us gather information that is only available from instructors and administrators closely associated with a specific course. You may have been asked some of these same questions in March or April, but we want to be sure that our information is currently correct.

1. Would you give us a general description of how a student goes through a course?

(Record responses and mark those questions in the interview form which the respondent may answer during the description.)

2. Is this course self-paced, group-paced or lock step? (circle one or more)

2.1 (If group-paced) What exactly does group-paced mean?

3. Are these all the tests that are administered to students in this course? Y N

(Have the respondent go through the tests one by one if he/she has not already done so. If any tests are missing find out from the SME whom to contact to get them ASAP. Record the titles or descriptions given by the SME in case the tests do not arrive or cannot be found.

With the SME's assistance organize the tests in the order administered. Number all tests using a red felt marker or pencil on the upper right-hand corner. Use the same number followed by a lower case letter for alternate forms.

If additional tests are brought forth number them following the already numbered tests and indicate where they belong in sequence. Example: 1, 2, 3, 4, 5, 6 - administered directly after test 2.

Write the number we assign to the test on a blank student record form in the area where the score for that test would be recorded.)

4. Which test scores are put in the trainee's school record? (Write test numbers assigned)

5. Which of these tests must a trainee pass to graduate?
(Write numbers assigned to test or "all")

6. For each test, what is a passing score?
(Write the test number/score for each test)

7. Which of these tests has a strict time limit?
(Write number of tests or "all")

8. About what percentage of all trainees normally finish each timed test?
(Write number of test/percentage who finished)

9. Is an end-of-course test given that covers the entire course?

Y N

10. How much does each test contribute to the final grade? (Write test number/percent of final grade or each tests's weight)

11. Do all trainees take the same tests? Y N

11.1 (If no) What tests are used as sets? (Record numbers that make up sets)

12. Can a trainee skip or miss one or more of the tests? Y N

12.1 (If yes) Which ones? (Write test numbers)

12.2 (If yes) For what reasons can a trainee skip a particular test?

13. What is the procedure used to calculate a trainee's grade on each test?

14. How is a trainee's end-of-course grade calculated?

15. (If a numeric or percentile score is given as an end-of-course grade) What is the lowest score possible for an end-of-course grade? (Including trainees who get recycled or dropped)

15.1 What is the highest score possible?

15.2 Can a score be any number between these two extremes?

16. How is passing the course determined?

17. Is any record kept showing how many times a trainee may have taken a test before passing? Y N

17.1 (If yes) How is it recorded?

17.2 (If no) Could this be recorded also?

18. For each test how many times can a student be retested? (Write test number/ number of times retested)

19. How soon after failing a test can a student be retested? (respondents answer)

20. Can trainees take a pre-test if they think they can pass it without going through the training on any of these lessons? Y N

20.1 (If Yes) Is the score recorded as a 1st time GO or NO GO on the trainee's individual record?

20.2 Is there any way for us to identify such a score as a pre-test score?

21. If only GO/NO GO is recorded on a student's record, would it be possible to record actual points or checks earned?

22. Can we tell from the training record whether a trainee was dropped from the course for academic reasons or for some other reasons? Y N

22.1 (If yes) How is this indicated on the training records? (Respondent's reply)

22.2 (If no) How hard would it be to use this code to record the reasons for the drop? (Hand respondent a copy of the TRADOC code)

23. Are there any written guidelines concerning recycling trainees? Y N

23.1 (If yes) Where could we get a copy?

23.2 (If no) What are your usual procedures?

24. Are there any written guidelines telling when a trainee should be dropped from the course? Y N

24.1 (If yes) Where can we get a copy?

25. For each paper and pencil test, is a monitor or test supervisor present? (Test number of each test monitored)

26. For each hands-on test, is prompting permitted when a step in the procedure is not known? Y N

26.1 (If Yes) Is a record made of the prompt?

26.2 (If Yes) Is it used to adjust the score?

26.3 (If Yes) How?

27. Which of the tests require constant checking of the examinee's performance?

28. Are other trainees allowed to watch a trainee take a test prior to their taking it? Y N

29. For each test, how many trainees are tested simultaneously by one person? (Record test number/number of persons tested)

30. Is test security important for any of these tests? Y N

30.1 (If yes) What procedures are used to maintain test security?

30.2 (If yes) Are there any guidelines describing the procedures followed to maintain test security? Y N

30.3 (If 30.2 yes) Where can we get a copy of this document?

31. In some courses a trainee's performance is evaluated in ways other than paper-and-pencil or hands-on tests. Do such things as inspections, attitude, military bearing, instructor and CO comments, and other such information enter into grading a trainee? Do such things affect whether or not a trainee graduates from the course, gets recycled, or gets to take a retest?

31.1 (If other factors included) Is such information recorded? (Record the type of factors, the type of scores or method of assessment.)

31.2 (If other factors recorded) Where can this information be found?

31.2 (If other factors included) How much weight does such information carry in any decision to drop, retest, or recycle a trainee?

32. Are any changes to the course content, measures or recording procedures anticipated during the period 1 July 83 to 30 June 84? Y N

32.1 (If yes) What are they?

33. Is there a file or pool of knowledge test questions? Y N

33.1 (If yes) How is the pool used?

34. Do you have information about the statistical properties of these tests? Y N

34.1 (If no) Would anyone at this post have this type of information? Y N

34.2 (If yes) Who?

35. Do you have any information on procedures or steps that went into the development of these tests? (If not, try to find out who does.) Y N

36. Do you know of any written document that describes the test development procedures used for any of the tests? Y N

36.1 (If yes) What is the title, and where could I get a copy of this document?

37. What office or group wrote the tests that are currently being used? (Trace process)

38. How was the content of the tests selected from the content of the entire course?

39. How was the content of the course determined? Who determined the course content?

40. How was the type of course (self-paced vs lock step, performance oriented vs knowledge oriented, etc.) selected?

41. How was it decided what instruction activities the trainees and instructors would do?

42. What media and/or methods of instruction are used in the course? (List) (Unless this has been determined from POI)

43. Are different methods/media used to teach different kinds of content?

44. How were these particular methods/media chosen?

45. How does the training system respond to individual variations in performance?

46. Is there information available about how well parts of the course are meeting their goals? Y N

46.1 (If yes) Where could I obtain this information?

47. Is there information available about the performance of students after they leave the course? Y N

47.1 (If yes) Where could I obtain this information?

(For 63B, 64C, 76Y, and 94B only:)

48. This course is taught at:
   63B - Ft. Dix, Ft. Jackson and Ft. Leonard Wood
   64C - Ft. Dix and Ft. Leonard Wood
   76Y - Ft. Jackson and Ft. Lee
   94B - Ft. Dix, Ft. Jackson and Ft. Leonard Wood

   Do you know whether the course content, training measures and individual training records are approximately the same at each of these posts? Y N

   48.1 (If there are differences) Can you tell us what these differences are?

(Find out how far back existing records go. When this is determined:)

49. Is there a control sheet or record which shows when changes to the course were made, e.g., when a new POI, or lesson, or test was implemented? Y N

   (If yes, use this record to obtain the superseded documents that were in effect for the past training records that can be obtained.)

   49.1 (If no) Who can tell us what training documents have been changed during the period over which training records can be obtained?

   (Obtain the superseded documents.)

APPENDIX E

SAMPLE OF GROUP CODAP STATEMENTS FOR MOS 13B

PREVIOUS PAGE
IS BLANK

061-266-1504 Store Ammunition in Preparation for Firing

Q-13 Store Cannon Ammunition at Cannon Site

Q-12 Remove Ammunition from Containers

Q-11 Unload Ammunition from Vehicles

071-326-0601 Use Visual Signals To Control Movement

A-6 Guide Wheel/Track Vehicles During Daylight

A-7 Guide Wheel/Track Vehicles During Darkness

113-600-1015 Install and Operate a Field Telephone

V-17 Install Field Telephones

V-18 Prepare Field Telephones for Operation

V-19 Perform Operator Checks and Services on Field Telephone

V-26 Communicate Information Over Tactical Wire Lines

INSTRUCTIONS

Here is a stack of 165 cards stating job tasks that might be performed by a 13B10 soldier. Today you are going to be working with these task statements. You will be asked as a job expert to do two things:

1. Judge the importance of the tasks.

2. Sort the tasks into groups that are similar in content.

The results of your work will be used to help describe the important things that 13B Skill Level One soldiers do on the job so that the Army can do a better job of selecting recruits for that MOS.

Instructions for what we want you to do are on the next page. Please take your time and work carefully. If you have any questions about what you are to do, please ask. This work is very important to the Army and to the 13B MOS.

TASK IMPORTANCE

The first thing we want to do is find out how important you think each of these tasks is. But you will want to know, Important to what? So begin by reading the following mission scenario:

> Your unit is deployed to Europe as part of a U.S. Corps. The Corp's mission is to defend and maintain the host country's border during a period of escalating hostilities. The Corps maneuver terrain is inhibiting, weather is expected to be inclement. The enemy approximates a combined arms army and has nuclear and chemical capability. Air parity does exist. Enemy adheres to same environmental and tactical constraints as does U.S.Corps.

Now . . . put yourself in the shoes of a first-line supervisor and think about what tasks are important for Skill Level One 13B soldiers to perform so the unit can accomplish its support mission. The soldiers may be assigned duty positions normally held in the Battalion. A good way to look at it is to ask yourself, "If a 13B10 soldier can't do all of these tasks, which ones are the most important for the situation described?"

With this in mind, read through the 165 tasks. Then go back through them and pick out the one task you think is the most important in the situation and put it face down on your right. When you've done that, go through the stack again and pick out the one you think is least important in the situation and put that card face up on your left.

Now do that again for the remaining tasks—pick the one that is most important and put it face down on top of the one on your right, then put the least important face up on the one to your left.

Continue this procedure thinking each time just about the tasks still in the stack you started with and choosing from those the one most important and the one least important. You will be through when all task cards have been divided into two piles. Please be careful to keep the two piles in order as you do this. Let the group leader know when you are finished.

TASK CONTENT

Here is another set of 165 cards covering the tasks you worked with earlier. We now want you to do something completely different. We want you to sort them into groups on the basis of what the soldier does in performing the tasks.

Sort the tasks into groups according to your own knowledge of the activities required. Tasks which require similar procedures or principles should be placed in the same group. The objective is to set up groups of tasks so that the task activities tend to be alike within a group but different from one group to another.

You might want to go through the tasks and sort them fairly quickly first. Then go over your groups more carefully and reassign tasks as you think it is necessary. You may want to break large groups into smaller ones, or combine small groups into larger ones.

Take your time and do a careful job. There are no right or wrong answers here, only your own best judgment as a job expert as to how tasks are alike or different.

Because of the variety of tasks, we expect that you will end up with 15 to 25 groups of tasks. But it's OK if you believe fewer or more groups are necessary.

When you are through, clip together the task cards in each group and let the group leader know you are finished.

We are working on a project to reassess the personnel aptitude requirements for successful performance in Army jobs. One of the MOSs to be studied intensively is 13B10.

One of the crucial activities of the project is to develop tests of job proficiency. Since we want to build the tests around important aspects of job performance, we need to identify the most critical and representative tasks of all the tasks that a 13B at Skill Level One could be expected to perform.

The first step is to identify the tasks that are performed by 13B10 soldiers. We want to use your subject matter expertise to identify those tasks. We would like you to do four things:

- Review grouping of task statements
- Suggest explanations of differences in task lists
- Indicate doctrinal conflicts within list
- Suggest additional tasks

Review Groupings of Task Statements

We are drawing together two task lists: the MOS and Common Task Soldier's Manuals (SM) and the CODAP Surveys from the Army Occupational Survey Program. The first thing we want you to do is to review the way we have grouped the CODAP statements under the SM tasks. You have three choices for each CODAP statement:

- Leave it where it is
- Put it under a different SM task
- Move it to a non-SM list

We want the CODAP statement to relate to the task that best describes what an SL1 soldier does when he performs the CODAP behavior. To some extent that placement involves judging how the soldiers probably interpreted the CODAP statement. The current groupings represent our judgment; do not hesitate to disagree.

We also want the final task list to include all the meaningful functions for a 13B10. If a CODAP statement represents a meaningful function that is not part of an SM task, we need to consider it separately. The second part of your review will be to check our groupings of CODAP statements that we think represent behavior that is not covered by any SM task. During this review you have four choices for each CODAP statement:

- Leave it where it is
- Put it under an SM task
- Put it under a different non-SM task
- Make it a task by itself

The last part of your review of the groupings will be to decide whether the CODAP statements that you extract from the SM list should be grouped with other statements or be considered a task in its own right.

Suggest Explanations of Differences Between Task Lists

As we go through the list you will find that there are some Soldier's Manual task statements that do not have comparable CODAP statements. We are confident that there is a good reason in each case (such as equipment introduced since the CODAP survey), but we need to be able to account for the differences. When we come to examples of such tasks, I will ask for possible explanations.

Indicate Doctrinal Conflicts Within List

The results of your review will be used to prepare a list of tasks for a detailed survey of the importance of each task to the job of the 13B10. Before we conduct that survey it would be helpful if you would indicate tasks that a 13B10 should not perform. Here we are interested in cases where it would be wrong for a Skill Level One soldier to do the task. For example we understand that it is a violation of doctrine for a Skill Level One soldier to authenticate a nuclear message, yet a significant number reported doing the task. If we are right about the doctrine, we want to delete that task before we conduct our importance survey. We will make a separate pass through the higher skill level and non-SM tasks for this factor.

Suggest Additional Tasks

If there are important parts of the 13B10's job that are not covered by these tasks, we need to add them. If you think of an additional task while we are reviewing the CODAP and Soldier's Manual tasks, jot it down. We may have the task later. If not, we will discuss adding the task when we finish the review.

APPENDIX F

INITIAL SETS OF PERFORMANCE FACTORS

MOTOR TRANSPORT OPERATOR PERFORMANCE CATEGORIES

1. Driving Vehicles: Operating Army vehicles (e.g., trucks, jeeps, tractors and semi-trailers) in a safe, effective, and lawful manner; using appropriate driving principles and procedures in hazardous weather conditions (rain, snow, ice, fog) and in other special situations (field, blackout, etc.); taking appropriate action as necessary to prevent accidents or lessen their severity; driving effectively in convoys; using ground guides when appropriate.

2. Vehicle Coupling: Coupling/uncoupling trucks, tractors, and trailers according to standard operating procedures; making proper connections (e.g., air hoses) between vehicles.

3. Checking and Maintaining Vehicles: Performing PMCS; checking vehicles for problems before, during, and after commitments; recognizing vehicle problems and taking appropriate action; fueling vehicles properly and safely; performing maintenance duties (e.g., changing and inflating tires, changing oil, servicing batteries, etc.) effectively, safely, and according to procedures; ensuring that vehicles have proper equipment (e.g., fire extinguishers, tools, etc.).

4. Using Maps/Following Proper Routes: Securing proper maps as needed; using maps effectively; following prescribed routes; becoming familiar with routes ahead of time when appropriate.

5. Loading Cargo and Transporting Personnel: Supervising the loading of cargo; checking that cargo is properly distributed, secured, and blocked; securing safety straps and tailgate when hauling personnel; following special instructions when hauling dangerous or hazardous cargo; observing vehicle capacity limits.

6. Parking and Securing Vehicles: Setting the brakes and transmission properly when parking vehicles; using chock blocks and similar equipment as necessary; securing vehicles to avoid theft or damage to cargo while vehicle is not in operation.

7. Performing Administrative Duties: Preparing various forms such as DD Form 1970, DA Form 2404, DA Form 2408-1, SF 91, DD Form 518 completely, neatly, and accurately; reading forms carefully; obtaining needed forms before departing on commitment; turning in forms to proper persons; keeping account of cargo picked up and delivered; obtaining necessary licenses; reporting accidents and other incidents according to procedures.

8. Self-Recovering Vehicles: Taking appropriate action when vehicles are disabled; using winch or other equipment to perform vehicle self-recovery; following proper procedures when recovering or towing vehicles; repairing, replacing, or patching parts or equipment to enable crippled vehicles to complete mission.

9. Safety-Mindedness: Knowing and following safety procedures; being alert to possible dangerous situations and taking steps as appropriate to avoid them; exercising concern for the safety of other soldiers and the general public; using proper safety equipment; recognizing, reporting, and where appropriate, taking corrective action in unsafe conditions.

10. Assisting Others/Teamwork: Assisting others with vehicle problems; instructing, helping and encouraging others; pitching in to get the job done; getting along with others.

11. Hard Work/Effort: Working hard to complete assignments; willing to work long hours or on own time to complete jobs or prepare for assignments; take on additional duties; taking initiative.

12. Following Orders: Following oral and written instructions from higher ranking officers; carrying out commitments according to orders.

13. Performing Dispatch Duties: Dispatching other MTO's effectively; operating phones and radio.

ADMINISTRATIVE SPECIALIST PERFORMANCE DIMENSIONS

1. <u>Preparing Documents for Typing</u>

   . Identifying the correct form and format to use for each type of action or knowing where to go for help to identify the correct form and format (e.g., Army regulation manuals, sample completed forms, other co-workers, etc.);

   . Screening all drafts of documents for accuracy and for complete information, correcting grammatical and spelling errors before typing;

   . Following additional instructions to ensure the document is prepared in the correct manner.

2. <u>Typing and Proofreading Documents to Meet Scheduled Deadlines</u>

   . Completing all document actions in a timely fashion;

   . Submitting typed documents that contain no typographical errors or misspelled words and few if any erasures, strike overs, etc.;

   . Checking to make sure all required information appears accurately on the typed document;

   . Utilizing office equipment (e.g., memory typewriters, word processors, etc.) to ensure all work is completed on time.

3. <u>Answering Phones and Taking Messages</u>

   . Making a complete record of all calls received;

   . Providing proper identification information when answering phones (e.g., name and section or office);

   . Taking appropriately detailed messages (e.g., name, phone number, date, time and message);

   . Relaying phone messages to office personnel quickly and accurately.

4. Preparing Completed Documents (Actions) for Dispatch

- Using the appropriate distribution scheme to determine the number of copies needed;
- Obtaining the required number of document copies for proper distribution;
- Verifying copies for legibility;
- Reviewing documents to ensure the document contains the proper authorization, codes, etc. before dispatching;
- Assembling documents properly including all enclosures in the proper order.

5. Properly Routing/Dispatching Incoming and Outgoing Documents/Mail

71L

- Routing all paperwork, correspondence, messages, etc. to the correct units following the appropriate distribution scheme;
- Using the appropriate mailing procedure for special documents or packets (e.g., registered or certified mail, air courier services, hand delivery on post, etc.);
- Verifying that addresses are legible and accurate before distributing;
- Using designated information codes to sort documents properly;
- Becoming familiar with post sections and personnel to quickly and accurately sort distribution;
- Processing incoming mail as required (e.g., open and sort official mail, sort personal mail, etc.).

71L Fox 5

- Examining all mail pieces and packages to ensure each is properly addressed and wrapped;
- Using appropriate information to sort incoming mail properly;
- Becoming familiar with post sections and personnel to sort mail with partial addresses accurately;
- Using locator cards to re-route mail properly;

- Picking up and distributing mail at designated time and location;
- Sorting or routing all mail to ensure timely deliveries;
- Properly re-directing and re-routing mail according to Army/ Postal regulations.

6. Ordering/Obtaining and Distributing Office Supplies

- Accurately assessing office needs by inventorying office supplies or asking co-workers to determine quantities or types of supplies needed;
- Ordering supplies based on office needs (e.g., blank forms, paper, etc.) and ordering supplies as requested (e.g., publications);
- Using proper codes to process orders quickly and accurately and to ensure prompt delivery;
- Picking up supplies as needed to avoid shortages;
- Distributing supplies according to needs, requests, etc.

7. Posting Regulations

- Properly posting all regulations to ensure all actions are IAW current regulations (e.g., noting changes on the cover page, inserting sheets appropriately, etc.);
- Recognizing when publications, pamphlets, etc. require posting and obtaining the necessary information to post changes;
- Posting changes promptly to avoid unnecessary backlogs;
- Notifying other office personnel of regulation changes to ensure all actions are valid.

8. Establishing and/or Maintaining Files IAW TAFFS [This includes files for messages, completed forms, blank forms, correspondence, and classified, secret and top-secret documents.]

- Including files for messages, completed forms, blank forms, correspondence, and classified, secret and top-secret documents;
- Creating file folders, labels and categories in accordance with Army regulations;

- Verifying the accuracy and completeness of documents before filing, coding and marking or stamping all documents properly (e.g., preparation dates, disposition dates, etc.);

- Placing documents in the proper folder and in the proper order to ensure easy retrieval;

- When appropriate, developing new category areas for new section activities that are IAW Army regulations;

- Filing classified documents in accordance with security status codes;

- Reviewing filing system periodically to ensure all items are properly ordered and in good condition, and to ensure filing system is complete (e.g., no missing documents, incorrect labels or missing categories, etc.).

9. Processing and Maintaining Suspense File System

- Monitoring and processing the suspense file on a daily basis to ensure proper actions are taken to meet scheduled deadlines;

- Prioritizing suspense files to ensure all actions are completed on schedule;

- Logging in documents as they arrive and filing suspenses for each properly.

10. Maintaining Accountability of Classified Documents/Mail

71L

- Matching document codes with form codes before signing for documents;

- Attaching the required cover sheets to classified documents before dispatching;

- Logging in all classified documents upon receipt, ensuring all appropriate forms are attached;

- Obtaining a signed receipt for all dispatched accountable documents.

71L Fox 5

- Maintaining accurate records of all accountable mail items such as money orders, stamps, and registered certified and ensured mail;

- Logging in mail correctly upon receipt;

- Matching mail codes with form codes before signing for mail items;

- Ensuring only the addressees sign for accountable mail items upon receipt.

11. Safeguarding and Monitoring Security of Classified Documents/Mail

71L

- Following proper procedures (e.g., proper authorization) and using accurate disposition dates to destroy classified documents;

- Ensuring all documents are secure at all times (e.g., locking the safe when unattended, leaving no documents on desks unattended);

- Following security procedures when photocopying, transporting or shredding documents.

71L Fox 5

- Ensuring the mailroom is always secure while unattended (e.g., locking mailroom during lunch hours, or after hours, etc.);

- Properly securing all classified mail in safe or other security container;

- Ensuring mail is properly stored and protected at all times;

- Properly packaging and securing mail for transportation;

- Checking all incoming items to ensure security of mail;

- Safeguarding the mail in unusual circumstances (e.g., vehicle breakdown, traffic accidents, etc.).

12. Maintaining Office and Other Army Equipment

71L

- Understanding or becoming familiar with procedures for operating equipment to avoid unnecessary damage or downtime (e.g., typewriters, word processors, photocopiers, etc.);

- Allowing only authorized or experienced personnel to have access to office equipment;

- Cleaning and performing simple maintenance on equipment to ensure it operates efficiently;
- Notifying maintenance personnel promptly when equipment needs repair.

71L Fox 5

- Maintaining assigned vehicle in good operating condition;
- Cleaning vehicle as requested or as needed;
- Notifying maintenance promptly when vehicle needs repair.

13. <u>Maintaining Accurate Records</u>

71L

- Maintaining up-to-date personnel records including unit rosters, records of assignments;
- Maintaining and updating promotion records, duty status changes, orders and awards records, etc.;
- Maintaining and updating leave control log book;
- Maintaining accurate records for special events (e.g., promotion board scores, CO's calendar, etc.).

71L Fox 5

- Maintaining and updating locator card file;
- Correctly recording values on mail manifests.

14. <u>Providing Customer Service</u>

- Cheerfully helping customers and other SMs requiring assistance;
- Establishing or following procedures to help process customers quickly and accurately;
- Providing accurate information or referring customer to appropriate sources;
- Volunteering to help SMs needing prompt paperwork processing.

15. <u>Displaying Tact and Courtesy to Co-workers, Officers and Other SMs</u>

- Cheerfully greeting visitors, VIPs, etc. and offering refreshments and comfortable seating;

- Tactfully explaining problems or reasons for delays to others and handling complaints or outbursts;

- Using proper titles or ranks to address other Army personnel; showing the proper respect in all situations.

16. Compiling Lists from Office Records and Files

- Searching and screening available records to generate an accurate listing of personnel (e.g., unit strength records, recent promotions or awards information, etc.);

- Ensuring all requested information is included and is in the proper order;

- Verifying accuracy of list by checking with other office or outside sources;

- Compile list as requested to meet scheduled deadline.

17. Preparing Special Reports, Document Drafts or Other Materials

- Drafting letters, correpondence, new forms, etc. that are clear, concise, and require few changes;

- Developing presentations and presentation materials (e.g., briefing charts) that are clear, informative and accurate;

- When tasked, editing or re-writing handwritten drafts that more clearly convey the intended message;

- Thoroughly investigating assigned topic areas and writing reports that accurately and clearly summarize the information.

18. Contributing to Office Work Effort

- Completing all assigned tasks with little or no supervision;

- Assisting co-workers to avoid or reduce a backlog of work;

- Voluntarily working later or extra hours to complete work duties;

- Offering to perform extra job duties.

19. Providing Course Training or Informal On-the-Job Training

. Being fully prepared and knowledgeable about the topic area;

. Providing clear and accurate instruction on office/Army procedures and keeping up-to-date with all regulation changes;

. Voluntarily offering on-the-job training to newly assigned personnel.

20. Demonstrating Supervisory (Leadership) Skills

. Prioritizing work and assigning actions to others to meet scheduled deadlines;

. Distributing work assignments appropriately (e.g., assigning difficult tasks to a more experienced worker, etc.);

. Reviewing completed actions for accuracy before submitting them for authorization;

. Providing firm and clear guidance and direction to subordinate personnel.

21. Improving or Enhancing Job Knowledge and Job Skills

. Enrolling in courses to improve job skills;

. Conducting independent research to learn more about job duties (e.g., reviewing Army regulations during off hours);

. Developing a broad knowledge of all job duties and office actions;

. Studying materials to prepare for new job duties;

. Studying for special exams (e.g., SQT, soldier-of-the-month, etc.).

22. Displaying Conscientiousness Toward Work

. Being prompt and on-time for work or other scheduled activities (e.g., SQTs);

. Beginning work immediately upon arrival;

- Informing supervisor of appointments, scheduling conflicts or other events during work hours;

- Attending to work activities regardless of personal problems or conflicts (e.g., financial problems, duty transfer or re-assignment, etc.).

MILITARY POLICE PERFORMANCE DIMENSIONS

1. Providing Traffic Control Services

Directing traffic using regulation hand and arm signals to ensure a smooth accident-free flow of vehicles and pedestrians; monitoring compliance with traffic control laws and issuing traffic citations.

2. Operating a Circulation Control Point

Camouflaging one's CCP position using materials available near the CCP; monitoring movement of enemy troops and personnel in the vicinity of the CCP; using one's camouflaged position to surprise and capture enemy troops straying near one's CCP; preventing friendly troops and vehicles from straying into disputed or enemy territory; offering assistance to stragglers and refugees.

3. Providing Escort Services

Conducting route reconnaissances to ensure that roads and bridges are safe for the vehicles which are to be escorted and to identify potential ambush sites and alternate routes; briefing drivers regarding the escort or convoy route; staying alert for ambush or attack during the escort; dismounting, providing cover for the convoy, and directing the convoy to an alternate route in case of attack; preventing bunching among escort or convoy vehicles to minimize losses in case of an attack; ensuring that particularly valuable cargo and important passengers are provided with extra protection or camouflage; maintaining reasonably close physical proximity with the vehicles being escorted.

4. Making Arrests and Processing Prisoners, Internees, and/or EPWs

Using only the amount of force necessary to effect the arrest or apprehension; refraining from physical abuse of the prisoners; informing prisoners of the reasons for their arrest; reading prisoners their rights; searching prisoners and confiscating all weapons, drugs, evidence, etc.; calling for assistance and/or following SOP for processing opposite-sex prisoners; segregating prisoners and maintaining silence; using hand irons to maintain control of prisoners during processing and transport; keeping prisoners under observation at all times; providing for the safety of prisoners during processing and transport.

5. Providing Entry/Exit Physical Security

Restricting site access at strategic locations or in emergency situations to authorized and emergency vehicles and individuals; checking names and IDs of individuals seeking site access against access rosters; using sign/countersign challenge; evacuating civilian and military personnel from unsafe areas; staying at one's assigned guard post until relieved of duty; blocking traffic when necessary to keep unauthorized vehicles and pedestrians off the road or to promote the flow of emergency or military priority vehicles; setting up road blocks to apprehend suspected criminals or enemy personnel.

6. Conducting Security Checks

Checking thoroughly for buildings which have been left unsecured or for signs of forced entry (e.g., broken windows, open doors, etc.) into a secured building or area.

7. Demonstrating a Proper Respect for the Law and for One's MP Position

Obeying all laws, including traffic laws, both when on-duty and off-duty; pursuing all violations of military and civil law; not showing favoritism toward friends and colleagues or prejudice against any groups when enforcing the law; not abusing or attempting to profit from one's position as an MP; maintaining confidentiality regarding ongoing investigations and classified material.

8. Showing Respect for Authority

Obeying all orders issued by senior personnel; listening respectfully to criticism without talking back or becoming defensive; saluting officers; showing up on-time for formations and/or assignments; listening carefully to instructions and orders.

9. Maintaining a Proper Personal Appearance

Keeping one's uniform cleaned and pressed; keeping boots and brass shined; shaving, showering, and grooming oneself daily before duty; maintaining a neat, trim haircut; staying in full uniform, including head gear and arm bands, while on duty.

10. Keeping in Proper Physical Condition

Participating regularly in exercise and PT; maintaining a proper body weight; abstaining from drinking prior to duty; obtaining adequate sleep prior to duty in order to stay awake and alert during duty.

11. Demonstrating a Commitment to the Job

Working late or extra-long hours to complete one's assigned duties; working extra shifts when needed due to manpower shortages or emergencies; intervening in emergency situations when off-duty; working under difficult physical conditions or in temporary, undesirable assignments; working without complaining about working conditions; taking extra courses and tests to improve one's MP skills; not allowing personal problems to interfere with one's job performance.

12. Providing Training and Instruction for Other MPs

Giving on-the-job and formal classroom or field instruction to new MPs; serving as an example to new MPs while on the job; observing new MPs closely and providing help when they make mistakes; preparing to give formal lectures and demonstrations by reading books and manuals and developing thorough lecture notes and visual aids.

13. Maintaining Availability During Patrol Duty

Staying in one's assigned patrol area; informing the dispatcher when leaving the radio net during non-emergency situations (e.g., during lunch break); acknowledging all communications with the dispatcher; keeping one's radio tuned to the proper frequency; refraining from personal and leisure activities which would detract from one's ability to respond quickly to an emergency.

14. Providing Community Welfare and Crime Prevention Services

Providing assistance to individuals and families in non-emergency situations (e.g., escorting lost children to their home, changing tires or providing directions for motorists, searching for and retrieving lost pets, obtaining keys to let families into their quarters when they have locked themselves out); warning residents of unsafe conditions (e.g., approaching severe weather); requesting and/or arranging counseling for troubled soldiers and their families; participating in crime prevention programs and demonstrations; warning individuals to secure unsecured personal property; talking to residents on one's patrol beat to promote better relations between MPs and residents; volunteering for community service activities (e.g., scouts, operating a MARS station, etc.).

15. Cooperating with Civilian Authorities

Assisting civilian authorities in searching for suspected criminals who have fled onto post; ensuring that civilian authorities are informed of suspected criminals and AWOLs who have fled the post; assisting local authorities off-post at their request in the event of emergency, life-threatening situations; refraining from involvement in law enforcement activities off-post.

16. Caring for and Handling Weapons

Keeping one's weapon clean, lubricated, properly assembled, and properly sighted; exercising care with one's weapon in public (e.g., refraining from using the gun as a toy or playing "quick draw"); turning in one's weapon and all unused rounds each day at the end of duty or field exercises; ensuring that one's weapon is cleared and that the magazine has been removed before turning the weapon in; complying with all weapons safety rules and regulations.

17. Caring for and Maintaining Vehicles and Equipment

Conducting a thorough preventative maintenance check (PMC) on one's vehicle before driving it; adding oil and antifreeze to one's vehicle as needed; immediately reporting all vehicle maintenance problems to mechanics; ensuring that one's vehicle has an adequate supply of gas; ensuring that all equipment which may be needed during patrol duty (e.g., radar, radio, gas mask, etc.) is functioning properly; avoiding road conditions which are likely to result in damage to one's vehicle; ensuring that one's vehicle is secured when not in use.

18. Being Alert for and Responding to Suspicious or Unlawful Activities and Unsafe Conditions

Recording information from police broadcasts and seeking clarification when broadcast instructions are not clear; looking for persons matching the description of suspects; watching for stolen cars and property; monitoring parking lots, housing areas, buildings, strategic sites, etc. for possible criminal or sabotage activity; enforcing military rules and martial law, including lights and noise discipline; conducting physical searches to investigate unusual noises or activity; questioning suspicious persons; informing appropriate authorities of unusual noises or activity; using knowledge of the area to one's advantage.

19. Maintaining Personal Safety

Using emergency equipment (i.e., lights and siren) appropriately; driving carefully, especially near traffic signs and signals, when responding to emergencies; requesting and waiting for backup when assistance might be needed; maintaining radio communication so that the dispatcher and/or desk sergeant always know one's location and activities; recognizing potential threats to personal safety; using cover appropriately in potentially life-threatening situations.

20. Using Interpersonal Communication (IPC) Skills

Answering general information and law enforcement questions accurately and politely; refraining from verbal altercations with members of the public, suspects, or prisoners; refraining from profane or obscene language in public; maintaining an even temper, even when provoked; speaking with subjects in a serious, business-like manner; listening to all parties involved in an incident or dispute before taking action.

21. Responding to Medical Emergencies

Determining the subject's injury or condition; administering prompt emergency first-aid; transporting subjects to the hospital or calling for emergency assistance, as appropriate; rescuing subjects from life-threatening situations (e.g., fires, boating and swimming accidents, etc.).

22. Testifying in Court

Reviewing notes and records prior to testifying; demonstrating proper courtroom etiquette; explaining probable cause for arrest; explaining all details of the incident.

23. Gathering Information and Evidence

Securing crime and accident scenes; searching crime scenes; interviewing and obtaining names of suspects, victims, and witnesses; obtaining and verifying damage reports; obtaining descriptions of stolen property; obtaining descriptions of suspects; obtaining information needed to complete all paperwork; collecting evidence from the crime scene; maintaining the integrity of the evidence (e.g., not destroying the evidence as one is collecting it, maintaining the chain of custody, etc.); following up on leads; completing all paperwork accurately.

APPENDIX G

SOLDIER EFFECTIVENESS WORKSHOP
GENERAL PROTOCOL

Sponsoring Organization:

U.S. Army Research Institute
Alexandria, VA

Soldier Effectiveness Workshop Agenda

Time	Topic
0900 - 0915	Description of the project
0915 - 0945	Briefing on the day's activities
0945 - 1200	Generating performance examples
1200 - 1245	LUNCH
1245 - 1445	Generating more performance examples
1445 - 1515	Discussion of performance categories emerging in the workshop
1515 - 1600	Generating more performance examples
1600 - 1630	Review of the day's activities and discussion of next steps

Overview

The U.S. Army annually selects and assigns to specific MOS tens of thousands of new recruits. This initial selection and assignment is obviously important, and it is only the first in a series of personnel decisions regarding soldiers arising at key choice points at different times and stages in a soldier's career. This initial assignment triggers a series of decisions in which soldiers are trained, managed, led, utilized, retained, promoted, retrained, reassigned, etc. These decisions impact on the performance of individuals within the Army, and, as a consequence, on the effectiveness of Army units. The challenge is to make these decisions in a systematic manner, maximizing total system performance and minimizing the costs resulting from inappropriate selection, assignment, utilization and promotion decisions. Gaining the most from these decisions depends on the Army's success in predicting at time of entry the future performance of potential recruits in the full spectrum of MOS. Developing these prediction systems and following them up to ensure they are functioning effectively is what this Army Selection and Classification Project is designed to accomplish.

The project represents a major five to seven year commitment by the Army to fully utilize recent advances in the measurement and prediction of human performance. The end goal is an Army selection and classification system that assigns recruits to MOS according to their abilities and potential for performing effectively in each MOS. This should result in a system that puts the right people in the right jobs, minimizing the number of person-job mismatches that typically occur. Major parts of the project include:

1. Develop new and improved tests and other predictor measures of soldier effectiveness. Currently, the Army uses essentially an intelligence test (ASVAB) to help select and assign recruits. However, several other kinds of tests and predictors might be used in this selection and assignment process--biographical inventories, psychomotor and physical abilities tests, and personality and vocational interest measures are all possible candidates for a more comprehensive selection and classification test/predictor battery. During the project, these types of predictor tests will be developed.

2. Refine existing performance measures and develop new, more comprehensive and accurate measures of soldier effectiveness. The tests/predictors discussed in (1) above are intended to select recruits who will perform effectively in the Army. In other words, predictor tests will be designed to indicate when a recruit has the proper skills, abilities, personal characteristics, etc. to perform well as a soldier. However, we feel strongly that the accuracy of these performance predictors should be checked on. Are soldiers with high scores on the tests/predictors also the more successful performers, and do those soldiers with lower test/predictor scores perform less effectively? The only way to test this is to

*Editor's note: Later versions will more correctly refer to the ASVAB as a cognitive aptitude test.

administer the predictor tests to recruits and then evaluate their later Army performance to see if the pattern is as stated above--high scorers on the predictor tests perform effectively as soldiers and low scorers perform less effectively in their Army careers. During the project, we will develop improved soldier effectiveness measures and conduct this check on accurate selection and assignment.

Problem This Workshop Addresses

To conduct the check described above, it is critical to obtain a faithful, accurate picture of these soldiers' effectiveness levels. How well is each actually performing? If we are unable to get this accurate look at performance for individual soldiers, we cannot accomplish the test on prediction discussed previously.

Unfortunately, the EER ratings are unlikely to help us here. In the past few years, scores on the EER have all bunched up near the top of the scale. This, of course, makes it impossible to tell who's actually performing effectively and who is performing less effectively. A special evaluation of individual soldier performance is needed, an evaluation that doesn't go into the 201 File or any other records, but is used for research only in the present project. This actual performance rating in turn requires a special rating form developed to help raters provide an accurate picture of a soldier's performance and effectiveness. This is where you and these workshops come in.

We are going to work together to develop a state-of-the-art rating form. This form should help raters (for example, First Sergeant or fellow squad member) make accurate judgments of the actual performance effectiveness of soldiers they supervise or work with. Together, we will design what are referred to as behavior-based rating scales, a rating form that offers an opportunity for relatively objective assessments of performance.

Before introducing the behavior-based rating scale concept, here are some other kinds of rating forms.

Alternative Performance Rating Scale Formats

1. Examples of trait ratings:

Excerpts from a U.S. Army document regarding performance appraisal

Lower Senaca Town
August 15, 1813

Sir:

I forward a list of the officers of the 27th Regt. of Infty. arranged agreeably to rank. Annexed thereto you will find all the observations I deem necessary to make them.

Respectfully,
I am, Sir,
Yo. Cot. Servt.

Lewis Cass
Brig. Gen.

27th Infantry Regiment

Alex Denniston - Liet. Col., Comdg.	-a good natured man.
Clarkson Crolins - First Major	-a good man, but no officer.
Captain Shotwell	-a man of whom all unite in speaking ill, a knave despised by all.
" Allen Reynolds	-an officer of capacity, but imprudent and a man of most violent passions.
First Lieut. Wm Perrin " " Danl. Scott " " Jas. I. Ryan " " Robt. McElwrath	-low vulgar men, with exception of Perrin, Irish and from the meanest walks of life-possessing nothing of the character of officers or gentlemen.
" " Robt. P. Ross	-willing enough-has much to learn-with small capacity.
2nd Lieut. Nicholas G. Carner	-a good officer but drinks and disgraces himself and the services.

SOURCE: The First Recorded Efficiency Report in the Files of the War Department, August 15, 1813.

Leadership

Poor — Good

Aggressiveness

Low — High

Self-Confidence

Low — High

2. Examples of numerically anchored ratings:

a.

Quality of Work

1 2 3 (4) 5

b. Quality of Work: Judge the amount of scrap, consider general care and accuracy of work.

poor, 1-6; average, 7-18; good, 19-25

20

c.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Quality																	X								
	POOR					BELOW AVERAGE					AVERAGE					ABOVE AVERAGE					EXCELLENT				

3. Example of a behavior-based rating scale:

ESTABLISHING AND MAINTAINING GOOD RELATIONSHIPS IN THE COMMUNITY

Contacting and working effectively with high school counselors, newspaper editors, radio and TV personnel, and others capable of helping recruiters to enlist prospects; building a good reputation for the Navy by developing positive relationships with persons in the community; establishing and maintaining good relationships with parents and family of prospects; presenting a good Navy image in the community.

9 or 10

EXTREMELY EFFECTIVE PERFORMANCE

Is exceptionally adept at cultivating and maintaining excellent relationships with school counselors, teachers, principals, police, news media persons, local business persons, and other persons who are important for getting referrals and free advertising.

Is innovative in informing the public about the Navy; actively promotes the Navy and makes friends for the Navy while doing it; always distributes the most current Navy information.

Volunteers off-duty time to work on community projects, celebrations, parades, etc.

6, 7, or 8

EFFECTIVE PERFORMANCE

Spends productive time with individuals such as police, city government, or school officials; may lunch with them, distribute calendars, appointment books, buttons, etc., to them, and/or invite them for cocktails.

Arranges for interested persons such Navy activities as trips to the Naval Academy; keeps relevant persons informed of Navy activities.

Encourages principals, counselors, and other persons important to a prospect to call if they have any questions about the Navy.

3, 4, or 5

MARGINAL PERFORMANCE

Contacts school officials only sporadically; keeps them waiting for information they want; relationships with counselors, teachers, etc., and persons important to an applicant or recruit are distant and under-developed.

Is not alert to opportunities to promote the Navy; rarely volunteers off-duty time to promote the Navy and is unenthusiastic when approached to do something for the community; rarely accepts speaking invitations.

Is, at times, discourteous to persons in the community; for example, sends form letters to persons who have assisted him or other Navy recruiters; is not always alert to the family's desire for more information about the Navy and the program in which their son or daughter enlisted.

---

1 or 2

INEFFECTIVE PERFORMANCE

Does not contact high school counselors; does not accept speaking engagements; drives around in car instead of getting out and meeting people.

Alienates persons in community or persons important to an applicant or recruit by ignoring them, not answering their questions, responding rudely, demanding information, encouraging high school students to drop out of school; sometimes does not appear at recruiting presentations for which he/she is scheduled.

Presents negative image of the Navy by doing things like driving while intoxicated or speeding and honking impatiently at other drivers; may express dislike for the Navy or recruiting.

Advantages of Behavior-Based Scales

1. Scales constructed to reflect performance requirements regarded as important by those knowledgeable about the job.

2. Scales define in concrete terms the relevant and important performance requirements.

3. Job experts agree on the effectiveness levels of scaled job behaviors used as performance effectiveness "anchors."

4. Rating task with these scales emphasizes objective observation rather than subjective evaluation.

5. In sum, raters can compare the observed performance of a soldier to behavioral benchmarks or standards of effectiveness, resulting in more objective performance judgments.

HOW TO WRITE SOLDIER EFFECTIVENESS INCIDENTS

To write a performance example or incident, try to remember what the soldier actually did or failed to do that made him or her effective or ineffective in a situation. These can be examples of extremely effective, ineffective, or even average performance. The important thing is that the incident is described specifically as it happened.

When writing an incident, describe only what you saw or what the person did, not what you inferred from the action. For example, in writing an incident, rather than writing that the soldier "displayed loyalty," you should describe what this soldier did to make you believe he or she was loyal. As examples, the soldier "worked all night to accomplish a job," or "speaks very highly of his/her C.O." Both of these behaviors or actions might be described as displays of loyalty; they are things a soldier did to make the writer believe he or she was loyal. Thus, we are asking you to describe specific behaviors or actions, not traits or personal characteristics.

The features of a good incident are:

1. It concerns the actions of an individual soldier.

2. It tells what the soldier did (or did not do) that made you feel he or she was effective or ineffective.

3. It describes clearly the background of the incident, along with the consequences of what happened.

4. It is concise in that it is short, to the point, and does not go to great lengths specifying unimportant details of the background, the activity itself, or the consequences of what the soldier did.

On the following pages are some hypothetical examples we will use to get you "up to speed" to write behavioral examples or incidents.

PERFORMANCE INCIDENT FORM

Problems

1. What were the circumstances leading up to the incident?

Excessive and irrelevant detail.

While on a minefield exercise, a soldier cut his left forearm to the bone on his buddy's bayonet. The buddy, Joe Rads, put a 3" x 3" sterile field dressing on the wound and then began escorting him back to camp (about a two-mile hike). However, about one mile from camp the soldier passed out and went into shock. Joe ran back to camp for help.

2. What did the individual do that made you feel he or she was a good, average, or poor performer?

Labels the behavior rather than indicating what the actual behavior was and/or should have been.

Joe's first-aid efforts were stupid; he really blew it.

3. In what soldier effectiveness category would you say this incident falls?

Knowledge of First-Aid

4. Circle the number below that best reflects the correct effectiveness level for this example?

1	(2)	3	4	5	6	7	8	9
extremely ineffective		ineffective		about average		effective		extremely effective

Example 1a

PERFORMANCE INCIDENT FORM

Problems

1. What were the circumstances leading up to the incident?

   While on a field training exercise, this soldier's buddy deeply cut his arm and it was bleeding profusely. This soldier applied a standard field dressing, but the injured soldier continued to bleed.

2. What did the individual do that made you feel he or she was a good, average, or poor performer?

   Though realizing his buddy continued to lose considerable amounts of blood, this soldier failed to apply a pressure dressing or to use a tourniquet. As a result, the soldier continued to bleed and eventually went into shock.

3. In what soldier effectiveness category would you say this incident falls?

   Knowledge of First-Aid

4. Circle the number below that best reflects the correct effectiveness level for this example:

1	(2)	3	4	5	6	7	8	9
extremely ineffective		ineffective		about average		effective		extremely effective

Example 1b

PERFORMANCE INCIDENT FORM

Problems

1. What were the circumstances leading up to the incident?

M16 target practice

Insufficient information to evaluate the soldier's behavior.

2. What did the individual do that made you feel he or she was a good, average, or poor performer?

Told another soldier he was clearing his rifle improperly.

Doesn't indicate consequences or result of this action.

3. In what soldier effectiveness category would you say this incident falls?

Safety Consciousness

4. Circle the number below that best reflects the correct effectiveness level for this example:

1	2	3	4	5	6	7	(8)	9
extremely ineffective		ineffective		about average		effective		extremely effective

Example 2a

PERFORMANCE INCIDENT FORM

Problems

1. What were the circumstances leading up to the incident?

   This soldier was on the firing range with several other soldiers getting M16 target practice using live ammunition. He observed another soldier attempting to clear a jammed live round by poking a cleaning rod down the bore.

2. What did the individual do that made you feel he or she was a good, average, or poor performer?

   This soldier immediately approached the soldier and advised him of the extreme danger of clearing his weapon in this way. As a result, the soldier ceased his clearing effort, thereby avoiding a potentially serious accident.

3. In what soldier effectiveness category would you say this incident falls?

   Safety Consciousness

4. Circle the number below that best reflects the correct effectiveness level for this example:

1	2	3	4	5	6	7	(8)	9
extremely ineffective		ineffective		about average		effective		extremely effective

Example 2b

PERFORMANCE INCIDENT FORM

Problems

1. What were the circumstances leading up to the incident?

   During a field training exercise involving a 10-mile hike, this soldier observed a barracks-mate having great difficulty keeping pace due to the weight of the equipment he was carrying.

2. What did the individual do that made you feel he or she was a good, average, or poor performer?

   This soldier began teasing his barracks-mate about "dogging it." But after his mate got angry, this soldier helped carry some of the mate's equipment.

Double-barreled statement. Two behaviors are described, so it's unclear which one is being rated for effectiveness. No consequences indicated.

3. In what soldier effectiveness category would you say this incident falls?

   Teamwork

4. Circle the number below that best reflects the correct effectiveness level for this example:

1	2	(3)	4	5	6	7	8	9
extremely ineffective		ineffective		about average		effective		extremely effective

Example 3a

PERFORMANCE INCIDENT FORM

Problems

1. What were the circumstances leading up to the incident?

   During a field training exercise involving a 10-mile hike, this soldier observed a barracks-mate having great difficulty keeping pace due to the weight of the equipment he was carrying.

2. What did the individual do that made you feel he or she was a good, average, or poor performer?

   This soldier carried some of the other soldier's equipment for a couple of miles enabling that soldier to recover his strength and thereafter keep pace with the others.

3. In what soldier effectiveness category would you say this incident falls?

   Teamwork

4. Circle the number below that best reflects the correct effectiveness level for this example:

1	2	3	4	5	6	(7)	8	9
extremely ineffective		ineffective		about average		effective		extremely effective

Example 3b

PERFORMANCE INCIDENT FORM

Problems

1. What were the circumstances leading up to the incident?

   During a field training exercise involving a 10-mile hike, this soldier observed a barracks-mate having great difficulty keeping pace due to the weight of the equipment he was carrying.

2. What did the individual do that made you feel he or she was a good, average, or poor performer?

   This soldier began to tease his mate about "dogging it" so that everyone in their group could hear it. The episode caused considerable embarrassment to his mate.

3. In what job performance category does this incident fall?

   Getting Along With Others

4. Circle the number below that best reflects the correct effectiveness level for this example:

1	2	(3)	4	5	6	7	8	9
extremely ineffective		ineffective		about average		effective		extremely effective

Example 3c

APPENDIX H

RECORDS FORM AND GUIDELINE FOR RECORDING ARCHIVAL PERFORMANCE DATA

ID ______

RECORDS COLLECTION FORM

IDENTIFICATION

PMOS ______ DMOS ______ SMOS ______ SQT ______ DOT ______

REENLIST ELIG ______ EDUC. MIL ______ CIV ______

APPOINTMENTS & REDUCTIONS (2-1 # 18)

GRADE	DATE OF ELIG./RANK

_____ Promotion packet to E5 in Action Pending

AWARDS, DECORATIONS & CAMPAIGNS (2-1 #9)
(DO NOT RECORD: ASR, OSR, NPDR)

#1 TYPE ______ DATE AWARDED (YYMMDD) ______

#2 TYPE ______ DATE AWARDED (YYMMDD) ______

#3 TYPE ______ DATE AWARDED (YYMMDD) ______

ARMS QUALIFICATION (M16): EXP SPS MKM ______ DATE (YYMMDD) ______

GRENADE RESULTS: EXP 1C 2C ______ DATE (YYMMDD) ______

ARMS QUALIFICATION ( ) ______ DATE (YYMMDD) ______

LOCALLY DESIGNED CERTIFICATES

#1 FOR ______ DATE (YYMMDD) ______

#2 FOR ______ :ATE (YYMMDD) ______

(Adapted from DA Form 1059)

SERVICE SCHOOL ACADEMIC EVALUATION REPORTS

DATE

3. GRADE | 4. BR | 5. SPECIALTY/MOSC

6. COURSE TITLE | 7. NAME OF SCHOOL | 8. COMP

9. TYPE OF REPORT
☐ RESIDENT
☐ NONRESIDENT

10. PERIOD OF REPORT (Year, month, day) From: Thru:

11. DURATION OF COURSE (Year, month, day) From: Thru:

13. PERFORMANCE SUMMARY

*a. ☐ EXCEEDED COURSE STANDARDS *(Limited to 20% of class enrollment)*

b. ☐ ACHIEVED COURSE STANDARDS

*c. ☐ MARGINALLY ACHIEVED COURSE STANDARDS

*d. ☐ FAILED TO ACHIEVE COURSE STANDARDS

**Rating must be supported by comments in ITEM 18.*

14. DEMONSTRATED ABILITIES

a. WRITTEN COMMUNICATION
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

b. ORAL COMMUNICATION
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

c. LEADERSHIP SKILLS
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

d. CONTRIBUTION TO GROUP WORK
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

e. EVALUATION OF STUDENT'S RESEARCH ABILITY
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

(SUPERIOR/UNSAT rating must be supported by comments in ITEM 18)

15. HAS THE STUDENT DEMONSTRATED THE ACADEMIC POTENTIAL FOR SELECTION TO HIGHER LEVEL SCHOOLING/TRAINING?
☐ YES ☐ NO ☐ N/A *(A "NO" response must be supported by comments in ITEM 18)*

Comments ____________________

DATE

3. GRADE | 4. BR | 5. SPECIALTY/MOSC

6. COURSE TITLE | 7. NAME OF SCHOOL | 8. COMP

9. TYPE OF REPORT
☐ RESIDENT
☐ NONRESIDENT

10. PERIOD OF REPORT (Year, month, day) From: Thru:

11. DURATION OF COURSE (Year, month, day) From: Thru:

13. PERFORMANCE SUMMARY

*a. ☐ EXCEEDED COURSE STANDARDS *(Limited to 20% of class enrollment)*

b. ☐ ACHIEVED COURSE STANDARDS

*c. ☐ MARGINALLY ACHIEVED COURSE STANDARDS

*d. ☐ FAILED TO ACHIEVE COURSE STANDARDS

**Rating must be supported by comments in ITEM 18.*

14. DEMONSTRATED ABILITIES

a. WRITTEN COMMUNICATION
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

b. ORAL COMMUNICATION
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

c. LEADERSHIP SKILLS
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

d. CONTRIBUTION TO GROUP WORK
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

e. EVALUATION OF STUDENT'S RESEARCH ABILITY
☐ NOT EVALUATED ☐ UNSAT ☐ SAT ☐ SUPERIOR

(SUPERIOR/UNSAT rating must be supported by comments in ITEM 18)

15. HAS THE STUDENT DEMONSTRATED THE ACADEMIC POTENTIAL FOR SELECTION TO HIGHER LEVEL SCHOOLING/TRAINING?
☐ YES ☐ NO ☐ N/A *(A "NO" response must be supported by comments in ITEM 18)*

Comments ____________________

EDUCATION & SCHOOLS

AIT COMPLETION (YYMMDD) ________________ (Record only service courses after this date).

CIVILIAN EDUCATION & MILITARY SCHOOLS (2-1 #17)				
SCHOOL	MAJOR/COURSE/MOSC	DURAT	COMP	YEAR

Military Education: Any training/courses taken after entry and not shown in #17:

____

____

Correspondence Courses: Type of notice code, title, credit hours, evaluation, date

___ ____

___ ____

___ ____

___ ____

___ ____

___ ____

___ ____

___ ____

___ ____

___ ____

___ ____

___ ____

Civilian Education

Any semester hours taken after entry and not shown in #17: No. Hours/Courses ____________

_____ Transcript _____ DA 2496 _____ Cert. of Continuing Education _____ Cert. of Completior

(Adapted from DA Form 2166-6)

ENLISTED EVALUATION REPORT

(AR 623-205)

I. PERIOD OF REPORT					
FROM	YEAR	MONTH	THRU	YEAR	MONTH

J. RATED MONTHS	K. NONRATED MONTHS	L. NONRATED CODES

RATER	INDORSER	A. PROFESSIONAL COMPETENCE
		1. Demonstrates initiative.
		2. Adapts to changes.
		3. Seeks self-improvement.
		4. Performs under pressure.
		5. Attains results.
		6. Displays sound judgment.
		7. Communicates effectively.
		8. Develops subordinates.
		9. Demonstrates technical skills.
		10 Physical fitness.
		SUBTOTALS

RATER	INDORSER	B. PROFESSIONAL STANDARDS
		1. Integrity
		2. Loyalty.
		3. Moral courage.
		4. Self-discipline
		5. Military appearance.
		6. Earns respect
		7. Supports EO/EEO.
		SUBTOTALS

PART VI. SCORE SUMMARY

PART	RATER SCORE	INDORSER SCORE
III		
IV		
Sum		
REPORT SCORE (R + I ÷ 2) =		

I. PERIOD OF REPORT					
FROM	YEAR	MONTH	THRU	YEAR	MONTH

J. RATED MONTHS	K. NONRATED MONTHS	L. NONRATED CODES

RATER	INDORSER	A. PROFESSIONAL COMPETENCE
		1. Demonstrates initiative
		2. Adapts to changes.
		3. Seeks self-improvement.
		4. Performs under pressure.
		5. Attains results.
		6. Displays sound judgment.
		7. Communicates effectively.
		8. Develops subordinates.
		9. Demonstrates technical skills.
		10. Physical fitness.
		SUBTOTALS

RATER	INDORSER	B. PROFESSIONAL STANDARDS
		1. Integrity
		2. Loyalty.
		3. Moral courage.
		4 Self-discipline
		5. Military appearance
		6. Earns respect
		7. Supports EO/EEO.
		SUBTOTALS

PART VI. SCORE SUMMARY

PART	RATER SCORE	INDORSER SCORE
III		
IV		
Sum		
REPORT SCORE (R + I ÷ 2) =		

MAXIMUM	
85	85
40	40
125	125
	125

LETTERS

#1 Appreciation ________ Reprimand ________ Date of Letter (YYMMDD) ________

Commendation ________ Other ________

Content ________

____

____

Directed to: ____ MPRJ ____ OMPF ____ No reference made

#2 Appreciation ________ Reprimand ________ Date of Letter (YYMMDD) ________

Commendation ________ Other ________

Content ________

____

____

Directed to: ____ MPRJ ____ OMPF ____ No reference made

#3 Appreciation ________ Reprimand ________ Date of Letter (YYMMDD) ________

Commendation ________ Other ________

Content ________

____

____

Directed to: ____ MPRJ ____ OMPF ____ No reference made

CERTIFICATES OF ACHIEVEMENT, COMMENDATION, APPRECIATION

#1 DATE (YYMMDD) ________ CONTENT ________

____

____

#2 DATE (YYMMDD) ________ CONTENT ________

____

____

(Adapted from DA Form 2-2 1 Nov 1974)

INSERT SHEET TO DA FORM 2
RECORD OF COURT-MARTIAL CONVICTION

For use of this form, see AR 640-2-1; the proponent agency is the office of TJAG.

1. TYPE OF COURT MARTIAL	a. NUMBER	b. HEADQUARTERS	c. ARTICLE

2. SYNOPSIS OF SPECIFICATION AND DATE OF OFFENSE

3. SENTENCE AS APPROVED, INCLUDING DATE ADJUDGED AND DATE APPROVED *(after insertion, complete certification)*

4. ACTION ON SUPERVISORY OR APPELLATE REVIEW, INCLUDING HEADQUARTERS AND DATE *(after insertion complete certification)*

5. MODIFICATION, SUSPENSION, OR SETTING ASIDE OF TRIAL RESULTS *(insert action taken, headquarters and date) (after insertion, complete certification)*

6. SUSPENDED SENTENCE VACATED *(insert headquarters and date) (after insertion, complete certification)*

____ Court-Martial proceedings in Action Pending

(Adapted from DA Form 4126-R, 1 Apr 1975)

BAR TO REENLISTMENT CERTIFICATE (Face)

DATE

3. GRADE	4. ETS	5. DEROS

6. TOTAL ACTIVE SERVICE	7. CONDUCT	8. EFFICIENCY
YRS MOS DAYS		

9. RECORD OF COURT-MARTIAL CONVICTIONS *(Indicate type, offense, sentence, date adj and app)*

10. RECORD OF NON-JUDICIAL PUNISHMENT *(Art 15) (Indicate offense, sentence and date)*

11. RECORD OF NON-PAYMENT OF JUST DEBTS *(Indicate dates of Letters of Indebtedness, Counseling, and Results)*

12. OTHER FACTUAL AND RELEVANT INDICATORS OF UNTRAINABILITY OR UNSUITABILITY *(See para 1-34, AR 601-280*

ARTICLES 15/FLAG ACTION

#1 Date issued (YYMMDD) ______________ Location ______________

Violation of article(s) ________ ________ ________ ________

if vio. art. 86 record duration ______________

Crime/Reason ______________

Punishment		SUSPEND	VACATE
extra duty:	______		
forfeiture:	______		
restriction:	______		
reduction:	______		
confinement:	______		
other action:	______		

#2 Date Issued (YYMMDD) ______________ Location ______________

Violation of article(s) ________ ________ ________ ________

if vio. art. 86 record duration ______________

Crime/Reason ______________

Punishment		SUSPEND	VACATE
extra duty:	______		
forfeiture:	______		
restriction:	______		
reduction:	______		
confinement:	______		
other action:	______		

3 Date Issued (YYMMDD) ______________ Location ______________

Violation of article(s) ________ ________ ________ ________

if vio. art. 86 record duration ______________

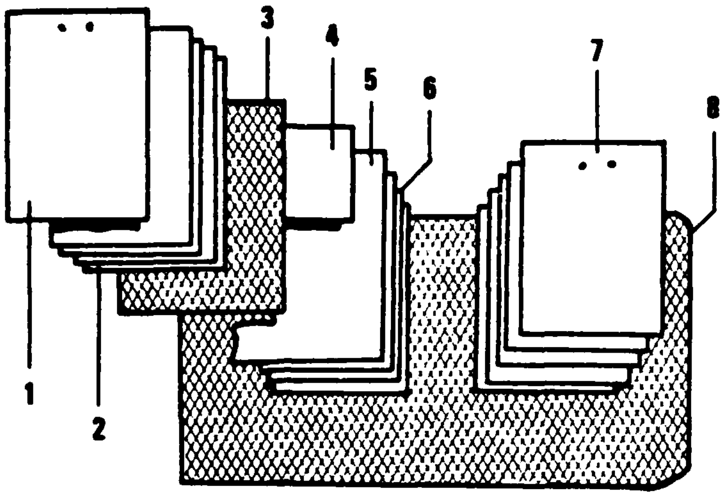
Crime/reason ______________

Punishment		SUSPEND	VACATE
extra duty:	______		
forfeiture:	______		
restriction:	______		
reduction:	______		
confinement:	______		
other action:	______		

GUIDELINES FOR RECORDS COLLECTION FORM

TASK 4

NOTE: These guidelines are intended for use by data collectors who have been trained in MPRJ data extraction.



THE MPRJ OF A REGULAR ARMY MEMBER ON ACTIVE DUTY

KEY

1 - DA FORM 268 (IF APPLIES)

2 - ACTION PENDING DOCUMENTS

3 - DA FORM 201A BETWEEN ACTION PENDING AND PERMANENT SECTIONS

4 - DD FORM 93

5 - VA FORM 29 - 8286

6 - ALL OTHER PERMANENT DOCUMENTS

7 - TEMPORARY DOCUMENTS

8 - DA FORM 201

GUIDELINES FOR RECORDS COLLECTION FORM

GENERAL: The odd numbered pages are where most of the recording will be done. If no information appears for an item, leave the space blank.

COVER SHEET: given - NAME, SSN, SEX, RACE, MOS, BASD, ID, MILPO CODE
To be removed upon completion of data collection for each MPRJ.

IDENTIFICATION (p.1): All items should be available from 2A (computer sheet).

PMOS DMOS SMOS	5 character alpha-numeric with possibility of 4 additional characters
EDUC MIL/CIV	1 character alpha or numeric code each
REENLIST ELIG	2 character numeric or alpha-numeric code
SQT	100 maximum score
DOT(SQT)	date of test YYMM

NOTE: Sometimes SQT information has not been entered on 2A, but you may find an Individual Soldier Report (ISR) in the Action Pending section. If person took new format there will be an "interim score" near the top right of ISR; date tested is above that. If person took old format there will be a percentage score near the center of the report with date tested to the right of that. If person is E5 or above there should be a 10A in the Permanent section. This gives the final SQT score, which is what you want to record. SQT dates for our 5 MOS: 64C JAN 83-JUN 83 (old); 71L MAR 83-MAY 83 (new); 05C APR 83--JUN 83 (new); 11B JUN 83-AUG 83 (new); 91B not scheduled.

APPOINTMENTS & REDUCTION (p.1): Is located on 2-1 #18 (green card).

Record exactly the information you find, e.g., PV1 810715.

Check (✓) if there is a promotion packet in Action Pending section.

AWARDS, DECORATIONS & CAMPAIGNS (p.1): NOTE: Do not record ASR, OSR, NPDR.

Applies only to those awards listed below, excluding Certificate of Achievement. (The abbreviations we know are given.)

Record of an award will be in abbreviated form on 2-1 #9 (green card).

To find the date, you must look in the Permanent section. Can be on a separate letter, form, or certificate, or can be included in orders. If latter, person's name is usually highlighted, checked, or underlined. (The award of the Army Commendation Medal is usually shown in orders.)

DSM	Distinguished Service Medal	Parachutist Badge
LM	Legion of Merit	Divers Badge
SM	Soldier's Medal	Explosive Ordnance Disposal Badge (Permanent awards only)
BSM	Bronze Star Medal (Valor or Merit)	Pathfinder Badge
	Defense Meritorious Service Medal	Aircraft Crewman Badge (Permanent awards only)
MSM	Meritorious Service Medal	Nuclear Reactor Operator Badge
AM	Air Medal (Valor or Merit)	Ranger Tab
	Joint Service Commendation Medal	Driver and Mechanic Badge
ARCOM	Army Commendation Medal (Valor or Merit)	Air Assault Badge
	Foreign Decoration (Individual Award or Decoration)	Drill Sergeant Identification Badge
AAM	Army Achievement Medal	US Army Recruiter Badge
PH	Purple Heart	Expert Marksmanship Qualification Badge (Most recent score on individual weapon only.)
CIB	Combat Infantry Badge	Campaign Star (Battle Star)
CMB	Combat Medical Badge	Certificate of Achievement (DA Form 2442)
	Good Conduct Medal	
EIB	Expert Infantry Badge	
EFMB	Expert Field Medical Badge	

Arms Qualification will be in same box near the bottom. The marksmanship level and date will be shown. The levels given on our form are regulation codes. Circle the recorded level. (Assume the following: EX=EXP; SS, SH, & SHS=SPS; MM & MK=MKM.) If a score or "NQ" is given, record it in the blank.

Grenade Results will be in same location as above. The levels given on our form are regulation codes. Circle the recorded level. (Assume EX=EXP; make no other assumptions.) Record different entries on the blank. Copy date.

In same location a different weapon's qualifications may be given; copy as shown with name of weapon; include date.

LOCALLY DESIGNED CERTIFICATES (p.1):

Are recognition for acts not covered by the Certificate of Achievement, Commendation, or Appreciation, e.g., honor graduate status, soldier of the month, selection as commander's orderly, high or perfect SQT, high or perfect APRT, training exercise. These will be found in the Permanent section.

SERVICE SCHOOL EVALUATION REPORTS #1059 (p.2): Instructions are included in next section.

EDUCATION & SCHOOLS (p.3):

AIT Completion Date is determined by looking inside 2-1 (green card) Section VII #35; finding column "Duty MOSC" and moving down it until you find MOS with skill level "1" in 4th position; moving directly left under column "Effective Date" and copying that date.

EFFECTIVE DATE		DUTY MOSC
810615		71L00
810815		71L00
811215	⬅	71L10

On 2-1 #17 (green card) look for any schools or courses taken after the above date. This is not always easy because only year is recorded in #17. If something is entered that looks like a possibility, you must peruse the Permanent section for a complete date from which to judge. It may be in the form of a certificate of training or a diploma. REMEMBER - you are trying to determine whether to use an entry in #17, not copying the certificate or diploma you find verifying the entry.

For those certificates and/or diplomas which refer to training courses taken after AIT but not listed in #17, record on the blanks under Military Education.

Do not record SIDPERS User Manual Training Course.

Record GED earned after entry.

A person who has completed a service school should have a #1059 in the Permanent section. Record the identifying information and evaluations on page 2 of our form.

Correspondence Courses will be found in Action Pending section (subcourses) until a Program of Instruction (POI, course) has been completed. This will be found in the Permanent section.

Use the following for "Type of Notice" code (should correspond to the box that has X in it):

A. enrollment notice	G. reexamination deficiency
B. phase completion notice	H. SSN correction notice
C. term of enrollment	I. subcourse completion notice
D. retirement point credit notice	J. reissue of failed subcourse notice
E. exemption notice	K. waiver notice
F. unavailability of subcourse	L. course completion notice
	M. 90 day warning notice

Record course title, number of credit hours, evaluation, and date.

Civilian Education which has not resulted in a degree, etc., is entered in pencil near the bottom of 2-1 #17. Record entry unless work was obviously done prior to entry into service; use AIT completion date as a guide. If year is the same, look for supporting documentation in Action Pending or Permanent sections.

AIT Completion Date 820811					
Tabor Coll	Health	1 yr	32SH	81	(obviously prior)
NVCC	Health		12SH	82	(must check further)
NECI	Health		6SH	83	(obviously after)

Record CLEP and DANTES entries.

Do not record a string like this:

MATH/SCI/BIO/CHEM/GEO, which has no other information entered.

ENLISTED EVALUATION REPORT #2166-6 (p.4):

Should exist in Permanent section for a person at E5. Information called for on our form comes from front and back of #2166-6.

LETTERS (p.5): Usually in the Permanent section, but can be in Action Pending.

There should be a SUBJECT line at the top that will indicate the type of letter. If not, read the first line of the body to determine, e.g., "I want to commend you ...," or "I congratulate you on...." "Others," that we know of, can be congratulations, admonition, and censure.

Beside the type of letter write the signer's rank, e.g., Appreciation CPT.

Copy the date.

Read the body to summarize the content.

Check (✓) if letter is directed to MPRJ and/or OMPF, or if no direction for filing is given.

If there is another letter(s) with the same content, the only additional recording needed is signer's rank. If SUBJECT is the same, make a slash (/) after first signer's rank and add rank of second letter, and so on, e.g., Appreciation CPT/MAJ/GEN. If SUBJECT is different, then record rank in the appropriate space.

CERTIFICATES OF ACHIEVEMENT, COMMENDATION, APPRECIATION (p.5):

Will always be so labeled and located in the Permanent section.

Record date and a summary of the reason for award of the certificate.

RECORD OF COURT-MARTIAL CONVICTION #2-2 (p.6):

Will be inserted in 2-1 (green card).

Record any information entered in items #1-6.

Check (✓) if you find court-martial proceedings in Action Pending section.

BAR TO REENLISTMENT CERTIFICATE #4126-R (p.6): Will be in Permanent section.

Record any information entered in date and items #3-12.

ARTICLES 15/FLAG ACTION (p.7):

Let's take the easier one first! Flag Action.

There will be an 8-1/2" x 11" form attached to the outside front of the MPRJ indicating a flag action on a person; it's hard to miss.

The information needed will be in Action Pending on a #2496 Disposition Form or a #268 Report for Suspension of Favorable Personnel Actions.

"Violation of article(s)" lines do not apply to a flag action.

Under "Punishment" use "other action" blank.

An Article 15 #2627 will be in the Permanent section.

"Date issued" is uppermost date you read in left side boxes.

Item #1 is the narrative of the charge (crime) and will contain location, violation number(s), duration if 86 and applicable.

Item #4 contains the punishment(s). Record all punishments; in "Suspend" column write number of days suspended. If suspension is vacated there will be a #2627-2 filed with the Article 15 #2627; check (✓) under "Vacate."

APPENDIX I

DATA BASES SEARCHED AS PART OF LITERATURE SEARCH FOR PREDICTOR SELECTION AND DEVELOPMENT

PSCYINFO. (Commonly known as Psyc Abstracts) This file is produced by the American Psychological Association and covers the world's literature in psychology and related behavioral and social sciences such as psychiatry, sociology, anthropology, education, pharmacology, and linguistics. The following general fields are covered: applied psychology, educational psychology, experimental human and animal psychology, experimental social psychology, general psychology, personality, physical and psychological disorders, physiological intervention, physiological pathology, professional personnel and issues, psychometrics, social processes and issues, treatment and prevention.

GPOM. (Government Printing Office Monthly Catalog) This file is produced by the Superintendent of Documents, United States Government Printing Office and indexes the public documents generated by the legislative branch, executive branch, and all agencies of the United States Federal Government. Some publications from the judicial branch are also included. The subjects covered are agriculture, commerce, defense, health and human services, education energy, housing, interior, justice, labor, state, transportation, and treasury.

NTIS. (National Technical Information Service) This file is produced by the National Technical Information Service of the U.S. Department of Commerce. The data base consists of government-sponsored research, development, and engineering reports as well as other analyses prepared by government agencies, their contractors, or grantees. The following are representative of the subject areas: administration and management; aeronautics and aerodynamics; agriculture and food; astronomy and astrophysics; atmospheric sciences; behavior and society; biomedical technology and engineering; building industry technology; business and economics; chemistry; civil engineering; communication; computers, control, and information theory; electrotechnology; energy; environmental pollution and control; health planning; industrial and mechanical engineering; library and information sciences; materials sciences; mathematical sciences; medicine and biology; military sciences; missile technology; natural resources and earth sciences; navigation, guidance, and control; nuclear science and technology; ocean technology and engineering; photography and recording devices; physics; propulsion and fuels; space technology; transportation; urban and regional technology.

ERIC. (Educational Resources Information Center) This data file is produced by The National Institute of Education and covers the following subject areas: adult, career, and vocational education; counseling and personnel services; early childhood education; educational management; handicapped and gifted children; higher education; information resources; junior colleges; languages and linguistics; reading and communication skills; rural education and small schools; science, mathematics, and environmental education; social studies/social science education; teacher education; tests, measurement, and evaluation; and urban education.

SSCI & SSCB. (Social Scisearch) These files are produced by the Institute for Scientific Information (ISI) and constitute an international, multidisciplinary index to the literature of the social, behavioral, and related sciences. Subjects included in the data base are anthropology, archaeology, area studies, business and finance, communication, community health, criminology and penology, demography, economics, education research, ethnic group studies, geography, history, information/library science, international relations, law, linquistics, management, marketing, philosophy, political science, psychology, psychiatry, sociology, statistics, and urban planning and development.

SSIE. (Smithsonian Science Information Exchange) This file is produced by the Smithsonian Science Information Exchange and contains abstracts of research either in progress or completed in the past two years. The data bases encompass all fields of basic and applied research in the physical, social, engineering, and life sciences including: agricultural sciences, behavioral sciences, biological sciences, chemistry and chemical engineering, earth sciences, electronics, engineering materials, mathematics, medical sciences, physics, social sciences and economics.

DTIC. (Defense Technical Information Center) This file is produced by the Defense Logistics Agency. It makes available from one central repository the thousands of research and development reports produced each year by U.S. military organizations and their contractors and grantees. Defense facilities and their contractors are required to submit to DTIC copies of each report (up to and including SECRET) that formally records scientific and technical results of Defense-sponsored research, development, test, and evaluation. Although created originally to serve the military, DTIC services habe been extended to all federal government agencies and their contractors, subcontractors, and grantees.

APPENDIX J

THE TOTAL ARRAY OF PREDICTOR CONSTRUCTS AND PERFORMANCE FACTORS THAT WILL BE USED IN THE TASK 2 TECHNICAL REVIEW

PREDICTOR CONSTRUCTS

Construct Name	Definition
Verbal Comprehension	Measures knowledge of the meaning of words and their relationships to each other.
Numerical Computation	Measures speed and accuracy in performing simple arithmetic operations, i.e., addition, subtraction, multiplication and division.
Use of Formulations and Number Problems	Measures the ability to correctly use algebraic formulae to solve number problems.
Word Problems	Measures the ability to select and organize relevant information to correctly solve mathematical word problems.
Reading Comprehension	Measures the ability to read and understand written material.
Two-Dimensional Mental Rotation	Measures the ability to identify a two-dimensional figure when seen at different angular orientations within the picture plane.
Three-Dimensional Mental Rotation	Measures the ability to identify a three-dimensional object, projected on a two-dimensional plane, when seen at different angular orientations either within the picture plane or about the axis in depth.
Inductive Reasoning: Concept Formation	Measures the ability to discover a rule or principle and apply it in solving a problem.
Spatial Visualization	Measures the ability to mentally manipulate the components of a two- or three-dimensional figure into other arrangements.
Deductive Logic	Ability to use logic and judgment in drawing conclusions from available information. Given a test of facts and a set of conclusions, deductive logic refers to the ability to determine whether the conclusions flow logically from the facts.
Field Dependence	Ability to find a simple form when it is hidden in a complex pattern. Given a visual percept or configuration, field dependence (or independence, more accurately) refers to the ability to hold it in mind so as to disembed it from other well-defined perceptual material.
Perceptual Speed and Accuracy	Ability to perceive visual information quickly and accurately and to perform simple processing tasks with it (e.g., comparisons). This requires the ability to make rapid scanning movements without being distracted by irrelevant visual stimuli, and also measures memory, working speed, and sometimes eye-hand coordination.

Construct Name	Definition
Mechanical Comprehension	Ability to learn, comprehend, and reason with mechanical terms. More specifically, this is the ability to perceive and understand the relationship of physical forces and mechanical elements in practical situations.
Rote Memory	Measures the ability to recall previously learned but unrelated item pairs.
Place Memory (Visual Memory)	Ability to remember the configuration, location, and orientation of figural material.
Ideational Fluency	Ability to rapidly generate ideas about a given topic or exemplars of a class of objects.
Follow Directions	Measures ability to follow simple and complex directions.
Analogical Reasoning	Measures the ability to identify the underlying principles governing relationships between pairs of objects.
Figural Reasoning	Measures ability to generate and apply hypotheses about principles governing the relationship among several figures.
Spatial Scanning	Measures the ability to visually survey a complex field to find a particular configuration representing a pathway through the field.
Omnibus Measures of Intelligence/Aptitude	Measures general mental ability or general aptitude.
Word Fluency	Ability to rapidly think of words.
Verbal and Figural Closure	Measures ability to identify objects or words given sketchy or partial information.
Processing Efficiency	Speed of reactions to simple stimuli.
Selective Attention	This is the ability to attend to a target stimulus when presented with two or more stimuli simultaneously.
Time-Sharing	Time-sharing is the ability to perform two or more tasks simultaneously.
Multilimb Coordination	Multilimb coordination is the ability to coordinate the simultaneous movement of two or more limbs. This ability is general to tasks requiring coordination of any two limbs (e.g., two hands, two feet, one foot and one hand). It is most common to tasks where the body is at rest (e.g., seated or standing) while two or more limbs are in motion.

Construct Name	Definition
Control Precision	Control precision is the ability to make fine, highly controlled (but not over-controlled) muscular movements necessary to adjust or position a machine or equipment control mechanism. This ability is general to tasks requiring motor adjustments in response to a stimulus whose speed and/or direction of movement are perfectly predictable. This ability is critical in situations where the motor adjustments must be both rapid and precise. The ability extends to arm-hand movements as well as to leg movements.
Rate Control	Rate control is the ability to make continuous anticipatory muscular movements necessary to adjust or position a machine or equipment control mechanism. This ability is general to tasks requiring motor adjustments or movements in response to a moving stimulus which is changing speed and/or direction in a random or unpredictable manner. The ability applies to compensatory tracking of the stimulus as well as following pursuit of the stimulus.
Manual Dexterity	Manual dexterity is the ability to make skillful, coordinated movements of the hand or the arm and hand. This ability most typically applies to tasks involving manipulation of moderately large objects (e.g., blocks, pencils, etc.) under speeded conditions.
Finger Dexterity	Finger dexterity is the ability to make skillful, coordinated, highly controlled movements of the fingers. This ability applies primarily to tasks involving manipulation of objects with the fingers.
Track Tracing Test	Designed to measure arm-hand steadiness.
Wrist-Finger Speed	The ability to carry out very rapid, discrete movements of the fingers, hands, and wrists. This ability applies primarily to tasks in which the accuracy of the movement is not a major concern. This ability is determined entirely by the speed with which the movement is carried out.
Aiming	The ability to make very precise, accurate hand movements under highly-speeded conditions. This ability is dependent upon very precise eye-hand coordination.
Speed of Arm Movement	This ability involves the speed with which discrete arm movements can be made. The ability deals with the speed with which the movement can be carried out after it has been initiated.

Construct Name	Definition
Involvement in Athletics and Physical Conditioning	Frequency and degree of participation in sports, exercise and physical activity. Individuals high on this dimension actively participate in individual and team sports and/or exercise vigorously several times per week.
Energy Level	Characteristic amount of energy and enthusiasm. The person high in energy level is enthusiastic, active, vital, optimistic, cheerful, zesty, and has the energy to get things done.
Cooperativeness	Characteristic degree of pleasantness versus unpleasantness exhibited in interpersonal relations. The highly cooperative person is pleasant, tolerant, tactful, helpful, not defensive, and generally easy to get along with. His/her participation in a group adds cohesiveness.
Sociability	Outgoingness. The person high in sociability is talkative, relates easily to others, is responsive and expressive in social environments, readily becomes involved in group activities, and has many relationships.
Traditional Values	Personal views in areas such as authority, discipline, social change, and religious commitment. The person with traditional values accepts authority and the value of discipline, is likely to be religious, values propriety, and is conventional, conservative, and resistant to social change.
Dominance	Tendency to seek and enjoy positions of leadership and influence over others. The highly dominant person is forceful and persuasive at those times when adopting such characteristics is appropriate.
Self-esteem	Degree of confidence in one's abilities. A person with high self-esteem feels largely successful in past undertakings and expects to succeed in future undertakings.
Conscientiousness	Characteristic amount of behavioral self-control. The highly conscientious person is dependable, planful, well organized, and disciplined. This person prefers order and thinks before acting.
Locus of Control	Characteristic belief in the amount of control people have over rewards and punishments. The person with an internal locus of control expects that there are consequences associated with behavior and that people control what happens to them by what they do. The person with an external locus of control believes that what happens to people is beyond their personal control.

Construct Name	Definition
Emotional Stability	Characteristic degree of stability vs. reactivity of emotions. The emotionally stable person is generally calm, displays an even mood, and is not overly distraught by stressful situations. He/she thinks clearly and maintains composure and rationality in situations of actual or perceived stress.
Nondelinquency	Amount of respect for laws and regulations as manifested in attitudes and behavior. The nondelinquent person is honest, trustworthy, wholesome, and law-abiding. Such persons will have histories devoid of trouble with schools and legal agencies.
Work Orientation	Tendency to strive for competence in one's work. The work-oriented person works hard, sets high standards, tries to do a good job, endorses the work ethic, and concentrates on and persists in completion of the task at hand.
Realistic Interests	Preference for concrete and tangible activities, characteristics and tasks. Persons with realistic interests enjoy, and are skilled in, the manipulation of tools, machines and animals but find social and educational activities and situations aversive.
Investigative Interests	Preference for scholarly, intellectual, and scientifice activities and tasks. Persons with investigative interests enjoy analytical, ambiguous, and independent tasks but dislike leadership and persuasive activities.
Enterprising Interests	Preference for persuasive, assertive and leadership activities and tasks. Persons with enterprising interests may be characterized as ambitious, dominant, sociable and self-confident.
Artistic Interests	Preferences for unstructured, expressive and ambiguous activities and tasks. Persons with artistic interests may be characterized as intuitive, impulsive, creative and non-conforming.
Social Interests	Preferences for social, helping and teaching activities and tasks. Persons with social interests may be characterized as responsible, idealistic, and humanistic.
Conventional Interests	Preferences for well-ordered, systematic and practical activities and tasks. Persons with conventional interests may be characterized as conforming, unimaginative, efficient, and calm.

CRITERION FACTORS

On the following pages are listed 72 criterion factors. The first 53 factors are duty groupings. Each has a brief defining phrase, such as "repair metal," followed by a more elaborate description of the factor. These 53 factors are intended to cover the occupational diversity of Army MOS's. (Note: For purposes of making direction of relationship judgments, high scores on these factors represent better performance).

Following these 53 factors are 19 factors that cover the Army-wide or common parts of soldier performance, including training. These factors have a basically similar format, but they are somewhat disparate in nature. Please read them carefully.

The last page of this booklet contains a list of the names of the 72 criterion factors. You may detach this for use as a ready reference source.

1. Inspect mechanical systems--test , measure , and/or use diagnostic equipment as well as visual, aural and tactile senses, in conjunction with technical information, to compare the operating status of mechanical equipment (e.g., engines, transmissions, machineguns) and mechanical components (e.g., bearings in an electrical generator) to standards of operating efficiency, and to identify malfunctions.

2. Troubleshoot mechanical systems--use test, measuring, and diagnostic equipment, in conjunction with technical information, to determine the cause of malfunctions in mechanical equipment (e.g., engines, transmissions, machineguns) and mechanical components (e.g., bearings in an electrical generator).

3. Repair mechanical systems--perform corrective actions on previously diagnosed malfunctions of mechanical equipment or mechanical components using appropriate tools (e.g., wrenches, screwdrivers, gauges, hammers) in conjunction with technical information.

4. Inspect fluid systems--use test, measuring, and diagnostic equipment, as well as visual, aural and tactile senses, in conjunction with technical information, to determine the operating status of fluid systems (e.g., hydraulic, refrigeration, engine cooling, compressed air) in comparison to standards of operating efficiency, and to identify malfunction.

5. Troubleshoot fluids systems--use test, measuring and diagnostic equipment, in conjunction with technical information, to determine the cause of malfunctions in fluid system (e.g., hydraulic, refrigeration, engine cooling, compressed air).

6. Repair fluids systems--perform corrective actions on previously diagnosed malfunctions of fluids systems using appropriate tools (e.g., wrenches, pressure gauges, soldering equipment, etc) in conjunction with technical information.

7. Inspect electrical systems--use test, measuring, and diagnostic equipment, as well as visual, aural and tactile senses, in conjunction with technical information, to determine the operating status of electrical systems (e.g., generators, wiring harnesses, switches, relays, circuit breakers, motors, lights) in comparison to standards of operating efficiency and to identify malfunction.

8. Troubleshoot electrical systems--use test, measuring and diagnostic equipment, in conjunction with technical information, to determine the cause of malfunctions in electrical systems (e.g., generators, wiring harnesses, switches, relays, circuit breakers, motors, lights).

9. Repair electrical systems--performs corrective actions on previously diagnosed malfunctions of electrical systems and electrical components using appropriate tools (e.g., pliers, wire strippers, soldering irons) in conjunction with technical information.

10. Inspect electronic systems--use test, measuring and diagnostic equipment, and to a limited extent, visual, aural, and tactile senses, in conjunction with technical information, to compare the operating status of electronic systems (e.g., communications equipment, radar, missile and tank ballistics controls) to standards of operating efficiency and to identify malfunctions.

11. Troubleshoot electronic systems--use test, measuring, and diagnostic equipment, in conjunction with technical information, to determine the cause or location of malfunctions in electronics systems (e.g., communication equipment, radar, missile and tank ballistics controls).

12. Repair electronic systems--performs corrective actions on previously diagnosed malfunction of electronic systems and electronic components using appropriate tools (e.g., test sets, screwdrivers, pliers, soldering guns) in conjunction with technical information.

13. Repair metal--performs corrective actions (e.g., bend, cut, drill, saw, weld, rivet, hammer, grind, solder, paint) to refabricate metal structures.

14. Repair plastic and fiberglass structures--perform corrective actions (e.g., measure, cut, saw, drill, sand, fill, paint, glue) to refabricate plastic and fiberglass structures to perform their original function, or to refabricate for modified function.

15. Construct wooden buildings and other structures--perform carpentry activities (e.g., measure, saw, nail, plane) to frame, sheath and roof buildings, or to erect trestles, bridges, piers, etc.

16. Construct masonry buildings and structures--perform masonry activities (e.g., measure, lay brick, pour concrete) to construct walls, columns, field fortifications, etc.

17. Prepare parachutes--inspect cargo and personnel parachutes, repair or replace faulty parachute components, and prepare (i.e., pack) parachute for future air drop.

18. Prepare equipment and supplies for air drop--fabricate and assemble platforms, cushions, and rigging to parachute supplies, equipment and vehicles; load, position and secure supplies and equipment in aircraft.

19. Install electronic components--place and interconnect electronic and communication components and equipment (e.g., radios, antennas, telephones, teletypewriters, radar, power supplies) and check system for operation.

20. Operate electronic equipment--set and adjust the controls of electronic components to operate electronic systems (e.g., radio, radar, computer hardware, missile ballistics controls).

21. Send and receive radio messages--use standardized radio codes and procedures to transmit and receive information.

22. Operate keyboard device--type information using a typewriter, teletype or keypunch, or computer terminal.

23. Use maps in the field--read and interpret map symbols and identify geography features in order to locate geography features and field positions on the map, and to locate map features in the field.

24. Plan placement or use of tactical position and features--using maps and on-site inspection, identify geographic positions or areas to be used for cover and concealment or to place fortifications, mines, detectors, chemicals, etc.

25. Place tactical equipment and materials in the field--without using heavy equipment (e.g., lifts, dozers) place mines, detectors, chemicals, camouflage of other tactical items into position on the battlefield.

26. Detect and identify targets--using primarily sight, with or without optimal systems, locate potential targets, and identify type (e.g., tanks, troops, artillery) and threat (friend or foe); report information.

27. Prepare heavy weapons for tactical use--transport, position and assemble heavy tactical weapons such as missiles, field artillery, anti-aircraft systems.

28. Load field artillery or tank guns--manipulate breech controls and handle ammunition (stow and load) to prepare guns for firing.

29. Fire heavy direct fire weapons (e.g., tank main guns, TOW missile, infantry fighting vehicle cannon)--using optical sighting systems, manipulate weapon system controls, aim, track and fire on designated targets.

30. Operate fire controls of indirect fire weapons (e.g., field artillery)--using map coordinates and ballistics information determine elevation and azimuth needed for firing at designated targets; adjust weapon using fire controls.

31. Fire individual weapons--aim, track and fire hand operated weapons such as rifles, pistols, and machineguns at designated targets.

32. Engage in bayonet and hand-to-hand combat--use offensive and defensive body maneuvers to subdue hostile individuals.

33. Operate wheeled vehicles--use various vehicle controls to drive wheeled vehicles from point to point, generally over paved and unpaved roads, observe traffic regulations; secure cargo.

34. Operate track vehicles--use various vehicle controls to drive track vehicles (e.g., tanks, APCs, scout vehicles, bulldozers); steer in response to terrain features.

35. Operate lifting, loading and grading equipment--operate heavy equipment (e.g., fork lifts, cranes, loader, back-hoes, graders) to load, unload, or move heavy equipment, supplies, construction materials (e.g., culvert pipes, building or bridge trusses), or terrain features (e.g., earth, road, trees).

36. Operate power excavating equipment--use pneumatic hammers and drills, paving breakers, grinders, and backfill tamper in the fabrication and modification of concrete, stone and earthen structures.

37. Reproduce printed materials--operate duplicating machines and offset presses to reproduce printed materials; collate and bind materials using various types of bindery equipment.

38. Make movies and videotapes--use motion picture cameras or videotape equipment to record visual and auditory aspects of assigned subject matter to be used for intelligence analyses, training or documentation.

39. Draw maps and overlays--uses drafting, graphics, and related techniques to prepare and revise maps, with symbols and legends, from aerial photographs.

40. Write and deliver presentations--prepares scripts for formal presentation including radio and television broadcast; make oral presentation.

41. Record and file information--collect, transcribe, annotate, sort, index, file, and retrieve information (e.g., training rosters, personnel statistics, supply inventories).

42. Receive, store and issue supplies, equipment and other material--inspect material and review paperwork upon receipt; sort, transport, and store material; issue or ship material to authorized personnel or units.

43. Prepare technical forms and documents--follow standardized procedures to prepare or complete forms and documents (e.g., personnel records and dispositions, efficiency reports, legal briefs).

44. Translate or decode data--use standardized coding systems and decoding rules to convert coded information to some more usable form (e.g., interpret radar information, decode Morse code, translate foreign languages).

45. Analyze intelligence data--determine importance and reliability of information; integrate information to provide identification, disposition and movement of enemy forces and estimate enemy capabilities.

46. Prepare food--prepare food and beverages according to recipes and meal plans (measure, mix, bake, etc.); inspect fresh food and staples for freshness; maintain sanitary work area.

47. Receive clients, patients, guests--schedule, greet and give routine information to persons seeking medical, dental, legal or counseling services.

48. Interview--verbally gather information from clients, patients, witnesses, prisoners, or other persons.

49. Provide medical and dental treatment--give medical attention to soldiers in the field, or medical or dental clinic, or to animals (e.g., CPR, splinting fractures, administering injections, dressing wounds).

50. Select, lay-out and clean medical or dental equipment and supplies--prepare treatment areas for use by following prescribed procedures for laying-out instruments and equipment; clean equipment and area for subsequent use.

51. Perform medical laboratory procedures--conducts various types of blood tests, urinalysis, cultures, etc.

52. Control individuals and crowds--apprehend suspected criminals, capture enemy soldiers, guard prisoners, participate in riot control operations, etc.

53. Control air traffic--coordinate departing, en route, arriving and holding aircraft by monitoring radar equipment and communicating with aircraft and other air traffic control facilities.

54. Following regulations--consistently complying with Army rules and regulations; conforming appropriately to standard procedures; following the spirit as well as the letter of military and civilian laws, regulations, written orders, etc.

55. Commitment to Army norms--adjusting successfully to Army life; displaying appropriate military appearance and bearing; showing pride in being a soldier.

56. Cooperation with supervisors--responding willingly to orders, suggestions, and other guidance from NCOs and officers; deferring appropriately to superiors' expertise and judgment and being supportive of superior officers/NCOs.

57. Cooperation with other unit members--pitching in when necessary to help other unit members with their job and mission assignments or during training; encouraging and supporting other unit members, as appropriate; showing concern for unit objectives over and above personal interests.

58. Hard work and perseverance--working hard on the job and during training; sustaining maximum effort over long periods of hard duty and on daily assignments; coping well with hardship or otherwise unpleasant conditions to continue to work toward mission completion.

59. Attention to detail--carrying out assignments carefully and thoroughly; consistently completing job and duty assignments on time or ahead of schedule; being conscientious in maintaining own and unit's equipment, and taking care to ensure that own quarters are clean and neat.

60. Initiative--willingly volunteering for assignments; performing extra necessary tasks without explicit orders; anticipating problems and taking action to prevent them.

61. Discipline--consistently concentrating on the job or duty assignment rather than being distracted by opportunities to socialize or otherwise stop working; controlling own emotions and not allowing them to interfere with performance of duty; keeping under control alcohol and other drug intake so that performance is not affected.

62. Emergent leadership--displaying good judgment in making suggestions to others in the unit regarding the job, duty assignments, etc.; appropriately taking charge when placed in a leadership position; where appropriate, persuading others in the unit to accept his/her ideas, opinions, and directions.

63. Survive in the field--react to direct or indirect fire; construct individual fighting position; camouflage self and equipment; use challenge and password; protect against NBC attack.

64. Maintain physical fitness--keep self at physical fitness level appropriate for state of battle readiness.

65. Disciplinary problems--having a record of disciplinary problems as reflected by AWOLS, Article 15s, civil arrests, etc. (Note: for judgment purposes, assume that high scores indicate a large number of disciplinary problems.)

66. Attrition--separating from the Army for "negative" reasons such as discipline or drug-related problems. (Note: For judgment purposes, assume that 0=did not attrite for negative reasons and 1=did attrite for negative reasons.)

67. Reenlistment--signing on for a second tour of duty.

68. Job satisfaction/morale--being satisfied with own MOS and Army life.

69. Training progress/success--successfully completing formal training course in normal amount of time versus washing out, being reassigned, being "set back" or "recycled."

70. Effort/motivation in training--the degree of effort, motivation, and interest that a soldier puts into his/her training, as evidenced by such things as curiosity about course content, not being afraid to be "wrong" or to ask questions, taking notes, being attentive in class, studying on own time, seeking out the instructor to clarify course content.

71. Performance of theoretical, or "classroom" parts of training--learning the theoretical part of a course; performing well on quizzes, tests, and examinations given in a classroom setting that tests the acquisition of concepts, principles, facts, or other information, e.g., learning the basic food groups, understanding the principles of internal combustion, learning the nomenclature of a weapon.

72. Performance of practical, "hands-on" part of training--applying the theory or principles of a course to practical problems and situations, either during simulations, field exercises, or other "hands-on" parts of training, e.g., cooking a meal, repairing an engine, firing a weapon, etc.

LIST OF CRITERION FACTOR NAMES

1. Inspect Mechanical Systems
2. Troubleshoort Mechanical Systems
3. Repair Mechanical Systems
4. Inspect Fluid Systems
5. Troubleshoot Fluids Systems
6. Repair Fluids Systems
7. Inspect Electrical Systems
8. Troubleshoot Electrical Systems
9. Repair Electrical Systems
10. Inspect Electronic Systems
11. Troubleshoot Electronic Systems
12. Repair Electronic Systems
13. Repair Metal
14. Repair Plastic and Fiberglass Structures
15. Construct Wooden Buildings and Other Structures
16. Construct Masonry Buildings and Structures
17. Prepare Parachutes
18. Prepare Equipment and Supplies for Air Drop
19. Install Electronic Components
20. Operate Electronic Equipment
21. Send and Receive Radio Messages
22. Operate Keyboard Device
23. Use Maps in the Field
24. Plan Placement or Use of Tactical Position and Features
25. Place Tactical Equipment and Materials in the Field
26. Detect and Identify Targets
27. Prepare Heavy Weapons for Tactical Use
28. Load Field Artillery or Tank Guns
29. Fire Heavy Direct Fire Weapons
30. Operate Fire Controls of Indirect Fire Weapons
31. Fire Individual Weapons
32. Engage in Bayonet and Hand-to Hand Combat
33. Operate Wheeled Vehicles
34. Operate Track Vehicles
35. Operate Lifting, Loading and Grading Equipment
36. Operate Power Excavating Equipment
37. Reproduce Printed Materials
38. Make Movies and Videotapes
39. Draw Maps and Overlays
40. Write and Deliver Presentations
41. Record and File Information
42. Receive, Store and Issue Supplies, Equipment and Other Material
43. Prepare Technical Forms and Documents
44. Translate or Decode Data
45. Analyze Intelligence Data
46. Prepare Food
47. Receive Clients, Patients, Guests
48. Interview
49. Provide Medical and Dental Treatment
50. Select, Lay-out and Clean Medical or Dental Equipment and Supplies
51. Perform Medical Laboratory Procedures
52. Control Individuals and Crowds
53. Control Air Traffic
54. Following Regulations
55. Commitment to Army Norms
56. Cooperation with Supervisors
57. Cooperation with Other Unit Members
58. Hard Work and Perseverance
59. Attention to Detail
60. Initiative
61. Discipline
62. Emergent Leadership
63. Survive in the Field
64. Maintain Physical Fitness
65. Disciplinary Problems
66. Attrition
67. Reenlistment
68. Job Satisfaction/Morale
69. Training Progress/Success
70. Effort/Motivation in Training
71. Performance of Theoretical, or "Classroom" Parts of Training
72. Performance of Practical, "Hands-On" Parts of Training

SUPPLEMENTARY

INFORMATION

AD-A137117

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION
1100 South Washington Street • Alexandria, Virginia 22314 • (703) 549-3611

DATE: April 10, 1984

MEMORANDUM TO: Addressee

FROM: Jim Harris, JH Project A Research Coordinator

SUBJECT: Corrected Pages for "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Technical Appendix to the Annual Report"
Research Note 83-37

Please insert the attached pages in Research Note 83-37, as follows:

Remove pages	Insert pages
~~ix~~/x	~~ix~~/x
11/11a	11/11a
335/336	335/336
337/338	337/338

JHH/bpm
Enclosure

TABLE OF CONTENTS (Continued)

TABLES

armed services. The project also needed a mechanism for assuring that the research program met the highest standards for scientific quality and state-of-the-art technology in personnel selection and classification research. Finally, because it takes some time in a longitudinal research program to arrive at definitive answers to some questions, a method was needed to receive feedback from senior officers on priorities and objectives, as well as to identify current problems where an appropriate research focus would bring operationally useful early results. An effective mechanism was essential because the research program involved a large number of troops. Their commanders would require justification for use of those assets.

With the active assistance of Dr. Joyce L. Shields, Dr. John P. Campbell, and MG H. N. Schwarzkopf, advisory group participants were identified, commitments to participate were obtained, and the groups were established. Figure 2 shows the structure and membership of the Governance Advisory Group (GAG).

The Scientific Advisory Group (SAG) comprises nationally recognized authorities in psychometrics, experimental design, sampling theory, utility analysis, applied research in selection and classification, and in the conduct of psychological research in the Army environment.

The InterService Advisory Group (ISAG) comprises the Laboratory Directors for applied psychological research in the Army, Air Force, and the Navy, and the Director of Accession Policy from the DoD Office of Assistant Secretary of Defense for Manpower and Research Affairs.

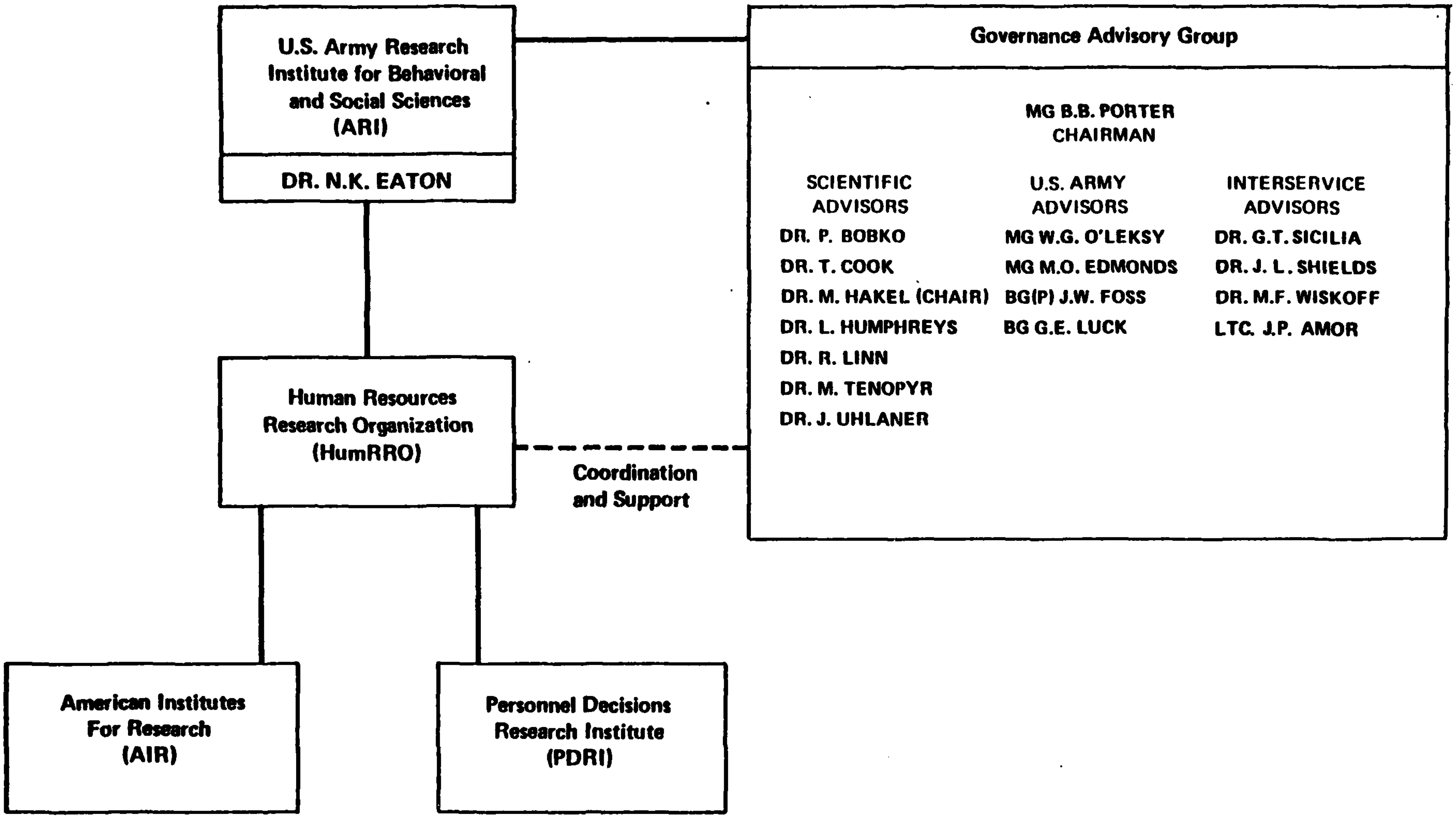


Figure 2. Governance Advisory Group

Results

To date (9/15), instructors and supervisors in the following 11 courses have been interviewed:

05C - Radio Teletype Operator	Ft. Gordon, GA
16S - Manpads Crewman	Ft. Bliss, TX
19E - Tank Crewman	Ft. Knox, KY
19K - M-1 Crewman	Ft. Knox, KY
63B - Light-Wheeled Vehicle Mechanic	Ft. Dix, NJ
63B - Light-Wheeled Vehicle Mechanic	Ft. Jackson, SC
64C - Motor Transport Operator	Ft. Dix, NJ
71L - Administrative Specialist	Ft. Jackson, SC
76Y - Unit Supply Specialist	Ft. Jackson, SC
94B - Food Service Specialist	Ft. Dix, NJ
94B - Food Service Specialist	Ft. Jackson, SC

The interview was concerned primarily with trainee progress and achievement measures in each course. Much of the information gathered is summarized in Table 15. There was surprising unanimity among the courses in these matters. All of the courses are group-paced (GP), except for the self-paced (SP) 05C course, the mostly self-paced 16S course, and the lock-step (LS) 19E and 19K courses. Since group-paced and lock-step are virtually indistinguishable modes of procedure, the 05C and 16S courses are the only real exceptions and both are scheduled to become group paced in the near future.

Table 15

Testing Practices in Selected MOS Training Courses

Items of Information	05C	16S	19E	19K	63B Dix	63B Jackson	64C	71L	76Y	94B Dix	94B Jackson
Type of testing done	HO[a]	HO/P&P[b]	HO/P&P	HO/P&P	HO/P&P	HO/P&P	HO/P&P	HO	HO/P&P	HO/P&P	HO/P&P
Tests pass to graduate	all	all	all	all	all	all	all	all	all	all	all
Can tests be missed	no	no	yes[c]	yes[d]	no	sometimes	no	no	no	no	no
Some test for all	yes	yes	yes	yes	no	no	no	no	no	no	no
Final "grade"	pass/fail	PI[e]	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail
Passing score (HO)	NA	PI ≤ 1.25	NA	NA	NA	NA	NA	NA	NA	NA	NA
Passing score (P&P)	NA	some 100%	80%	80%	70%	75%	70%	NA	100%	70%	62.5-70%
Strictly timed tests	some	one	some	some	all	none	some	none	few	none	none
Trainees finish within time limit	almost all	80%	all	almost all	95%	95%	80%	NA	100%	NA	all
Final grade calculation	NA	observed time / expected time	NA	NA	NA	NA	NA	NA	NA	NA	NA
No. of tests allowed	3+	3	3	3	3	3	3	3	3	3+	3
Interval before retesting	none	variable	none	none	none	none	retrain	none	24 hr. or less	none	variable
Retests recorded	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Challenges allowed	yes[f]	no	no	no	yes	yes	no	no	no	yes	no
Are tests monitored	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Is prompting permitted	no[g]	no	no	no	no	no	no[h]	no[i]	no	no	no[f]

[a]HO=Hands on
[b]P&P=Paper and Pencil
[c]Company commander can change rule
[d]For medical disability or for outstanding performance
[e]PI (Progress Index)=Observed time/Expected time
[f]Extremely rare
[g]Trainee has manual
[h]Minor prompting allowed
[i]Only for one test

Training performance data of considerable detail, e.g., task level or test item information, were often found to be recorded at Army schools. These data, generally not forwarded to centralized files, are not routinely available for research purposes. However, at most of the schools it was possible to make arrangements for these raw data to be forwarded to the Project A data base (LRDB) manager. The format in which detailed training performance data is available varies by installation. It may be maintained in the TRADOC Educational Data System (TREDS) system or on a local computer; it may be recorded in individual or class roster hard copy records; or it may be available only on the original individual test forms and score cards.

In line with the Army's current emphasis on competency training and testing, most tests now employed in these courses are hands-on or performance tests: TRADOC designation E1 or E2, rather than E3, paper-and-pencil, verbally oriented achievement tests. Although 9 of the 11 courses do have at least one written (E3) test, such tests constitute a minor portion of the testing programs. Because the line between E2 and E3 tests is a fine one, some of the tests reported as E3 might be better classified as E2.

Courses are designed to make achievement or competence cumulative throughout the course. In all 11 courses, all trainees must pass, successively, all of the tests given in order to graduate. Only in 19E, 19K, and 63B is a trainee allowed to skip a test, and then only in cases of temporary disability or outstanding performance. Almost all of the hands-on, performance tests (E1 and E2) given are scored on a pass/fail ("GO/NO GO") basis, although the criteria for pass/fail vary from course to course. The only exception to grading on a pass/fail basis is the 16S course, which utilizes a Progress Index (PI), calculated by dividing the observed time by the expected time, and allows a passing score of 1.25 or less. The written

(E3) tests in the 63B, 64C, and 94B courses require 70 percent to pass; those in the 16S and 76Y courses, 100 percent. In the other courses, the written (E3) tests are of minimal significance.

Time stress has been eliminated from most of the testing. The time limits on tests are strictly enforced only when time is the essence of the performance: in the 05C course for IMC (International Morse Code) for typing, and for preparing various pieces of equipment for operation, which must be done quickly under combat conditions; for the "10-Second Drill" in the 16S course; for the nuclear/biological/chemical drills in the 19E and 19K courses. The instructors interviewed generally believe that, for most tests, 90 percent or more of the trainees finish the required tasks well within the time limits of the tests, while the remainder would not be able to complete the test satisfactorily even if time were not a factor.

The coverage of the course by the tests varies. Virtually everything taught in every course is tested during the course, usually at the end of the module or annex. The nature of the course material and time, however, determine whether an end-of-course test (EOCT) will cover everything in the entire course. A relatively limited number of distinct tasks are taught in the 71L course for administrative specialists; this makes it quite possible to devise a test of reasonable length covering all tasks. The 94B course for food service specialists, on the other hand, includes a large number of tasks and a huge number of subtasks-- if each of the recipes in the cook's collection may be considered a subtask-- any one of which could generate a performance test. In such a case, a test of reasonable length can be constructed only by sampling arbitrarily from the vast array of possible performances. Whether such a test can be said to cover the entire course is debatable. Only in the 05C, 19E, 19K, and 71L courses does the EOCT appear to cover the entire course. In the other courses, the EOCT represents a sample of the course material.