Technical Report 838

Project A:
Improving the Selection, Classification, and Utilization of Army Enlisted Personnel

Weighting Criterion Components to Develop Composite Measures of Job Performance

Robert Sadacca and John P. Campbell
Human Resources Research Organization

Leonard A. White
U.S. Army Research Institute

Ani S. DiFazio
Human Resources Research Organization

May 1989

United States Army Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON **Technical Director**	**JON W. BLADES** **COL, IN** **Commanding**

Research accomplished under contract for the Department of the Army

Human Resources Research Organization

Technical review by

Lawrence M. Hanser
Michael G. Rumsey

DTIC COPY INSPECTED

Accesion For	
NTIS CRA&I	✓
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) HumRRO IR-PRD-88-14	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Technical Report 838
6a. NAME OF PERFORMING ORGANIZATION Human Resources Research Organization 6b. OFFICE SYMBOL (If applicable) HumRRO	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) 1100 South Washington St. Alexandria, VA 22314-4499	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Ave. Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION -- 8b. OFFICE SYMBOL (If applicable) --	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-82-C-0531
8c. ADDRESS (City, State, and ZIP Code) --	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 63731A PROJECT NO.: 792 TASK NO.: 232 WORK UNIT ACCESSION NO.: DA303189

11. TITLE (Include Security Classification)
Weighting Criterion Components to Develop Composite Measures of Job Performance

12. PERSONAL AUTHOR(S) Sadacca, Robert, and Campbell, John P. (HumRRO), White, Leonard A. (ARI), and DiFazio, Ani S. (HumRRO)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 85/01 TO 87/09	1989, May	69

16. SUPPLEMENTARY NOTATION Prepared under Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel (Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, U.S. Army Research Institute

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Composite criteria; Criterion measures; Performance components; Personnel classification; Personnel selection; Project A; Weighting

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

Project A is the Army's long-term program to develop a complete personnel system for selecting and classifying all entry-level Army enlisted personnel. During the Concurrent Validation phase, a wide variety of predictor and criterion measures were administered to 9,500 soldiers in various military occupational specialties (MOS). These data were used to refine a model of job performance for entry-level personnel in terms of five basic components: MOS-specific technical skills, general soldiering skills, effort and leadership, personal discipline, and military fitness and bearing. This report describes efforts to develop from these components a composite index of performance to use in determining the validity of the Trial Battery measures for each job. Experiments were conducted to determine the best method(s) of weighting the basic components for an overall composite index; weighting judgments were then gathered from a sample (totaling 712) of officers and noncommissioned officers (NCOs) familiar with each Project A MOS. Analysis of these data showed that both scaling methods tested (direct estimation and conjoint paired-comparison) produced highly

(Continued)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Lawrence M. Hanser	22b. TELEPHONE (Include Area Code): (202) 274-8275 22c. OFFICE SYMBOL: PERI-RS

DD Form 1473, JUN 86 Previous editions are obsolete.

ARI Technical Report 838

19. ABSTRACT (Continued)

reliable construct weights, with the conjoint method slightly favored. There was relatively high agreement, although sometimes different emphasis, between the officer and NCO judging groups.

16. SUPPLEMENTARY NOTATION (continued)

Lawrence M. Hanser, Contracting Offocer's Representative.

Technical Report 838

Project A:

Improving the Selection, Classification, and Utilization of Army Enlisted Personnel

Weighting Criterion Components to Develop Composite Measures of Job Performance

Robert Sadacca and John P. Campbell
Human Resources Research Organization

Leonard A. White
U.S. Army Research Institute

Ani S. DiFazio
Human Resources Research Organization

Selection and Classification Technical Area
Frances C. Grafton, Acting Chief

Manpower and Personnel Research Laboratory
Curtis Gilroy, Acting Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

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

May 1989

Army Project Number
2Q263731A792

Manpower and Personnel

FOREWORD

This document describes research conducted as part of the Army's large-scale manpower and personnel 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 U.S. military selection/classification test battery) and other selection variables as predictors of training and performance.

The overall research effort, referred to as "Project A," is devoted to developing and validating Army Selection and Classification Measures under the Selection and Classification Technical Area (SCTA) of the Manpower and Personnel Research Laboratory (MPRL) at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). This research supports the MPRL and SCTA mission to improve the Army's capability to select and classify its applicants for enlistment or reenlistment by ensuring that fair and valid measures are developed for evaluating applicant potential based on expected job performance and utility to the Army.

Project A was authorized through a letter, DCSOPS, "Army Research Project to Validate the Predictive Value of the Armed Services Vocational Aptitude Battery," effective 19 November 1980; and a Memorandum, Assistant Secretary of Defense (MRA&L), "Enlistment Standards," effective 11 September 1980.

To ensure that Project A research achieves its full scientific potential and will be maximally useful to the Army, a governance advisory group comprised of Army general officers, interservice scientists, and experts in personnel measurement, selection, and classification was established. Members of the last component provide guidance on technical aspects of the research, while general officer and interservice components oversee the entire research effort; provide military judgment and periodic reviews of the research progress, results, and plans; and coordinate within their commands. Members of General Officers' Advisory Group during the period covered by this report included MG W.G. O'Leksy (DMPM) (Chair), MG J.B. Allen, Jr. (DCSOPS), MG T.J.P. Jones (FORSCOM, DCSPER), MG G. Mallory (TRADOC, DCS-T), and BG P.M. Mallory (USAREUR, ADCSOPS). The General Officers' Advisory Group was briefed in May 1987 on the results of the concurrent validation, the preliminary results of the second-tour job analysis, and the plans for the longitudinal validation data collection. Members of Project A's Scientific Advisory Group (SAG) guide the technical quality of the research. During the period covered by this report, they included Drs. Philip Bobko, Thomas Cook, Milton Hakel (Chair), Lloyd Humphreys, Lawrence Johnson, Robert Linn, Mary Tenopyr, and Jay Uhlaner. The SAG was briefed in March 1987 on the status of the second-tour job analysis, the final resolution of utility measurement issues, and the reanalysis of the aptitude area composite and in

September 1987 on the results of the utility and construct weighting research and the plans for second-tour criterion measurement.

A comprehensive set of new selection/classification tests and job performance/training criteria have been developed and field tested, and the revised tests have been administered in a large-scale concurrent validation and in longitudinal validation phases. The present report describes the work done to estimate performance component weights for the Batch A and Batch Z military occupational specialties (MOS) tested during the concurrent validation. Results from this and other Project A research activities will be used to link enlistment standards to required job performance standards and to more accurately assign soldiers to Army jobs.

EDGAR M. JOHNSON
Technical Director

WEIGHTING CRITERION COMPONENTS TO DEVELOP COMPOSITE MEASURES OF JOB PERFORMANCE

EXECUTIVE SUMMARY

Requirement:

The Army's Selection and Classification Project (Project A) is a long-term research program to improve the system used to select and classify personnel for enlisted Military Occupational Specialties (MOS). A major goal of this research is to validate current and new predictors of performance against job success in enlisted occupations.

Results of the Project A research revealed that performance in entry-level MOS can be summarized in terms of five distinct performance factors: MOS-specific technical skills, general soldiering skills, effort and leadership, personal discipline, and military bearing and fitness. However, to relate scores on selection and classification measures to overall job success, the Army needs to combine the information about performance in these different areas to form a composite score. This report describes the effort to scale the contribution of these five performance factors to overall effectiveness.

Procedure:

A series of workshops was conducted to compare various methods for judging the relative importance of the performance factors. The two most successful methods were subsequently used in a series of workshops in which MOS job incumbents provided importance weights for the five performance factors. A total of 471 commissioned and 231 noncommissioned officers participated in these workshops. Importance weights were obtained for 20 MOS. These 20 MOS were selected for Project A as a representative sample of enlisted job requirements.

Findings:

In 13 of the 20 MOS, core technical skills were judged to be most important. Job effort and leadership received the highest relative weight in 6 of the 20 MOS. For all 20 MOS, military bearing and fitness was the performance factor judged to be least important.

The mean importance ratings for the five performance factors were generally highly reliable (above .80). Use of a conjoint scaling procedure, as compared with a direct estimation technique, yielded more reliable importance weights for the performance factors. Also, officers agreed more than NCOs regarding the relative contribution of the performance factors to overall effectiveness.

Utilization of Findings:

The information collected in this research provides a defensible basis for combining scores on the performance factors into a single composite or composites (e.g., "will-do" and "can-do") to measure an enlisted soldier's performance. Thus, for the representative sample of Army jobs in the Project A research, we can now evaluate the validity of the Armed Services Vocational Aptitude Battery (ASVAB) and other new selection and classification measures as predictors of overall performance on the job.

WEIGHTING CRITERION COMPONENTS TO DEVELOP COMPOSITE MEASURES OF JOB PERFORMANCE

CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

WEIGHTING CRITERION COMPONENTS TO DEVELOP COMPOSITE MEASURES OF JOB PERFORMANCE

OVERVIEW OF PROJECT A

Project A is a comprehensive long-range research and development program the U.S. Army has undertaken to develop an improved system for selecting and classifying enlisted personnel. The Army's goal is to increase its effectiveness in matching first-tour enlisted manpower requirements with available personnel resources, through use of new and improved selection/classification tests that will validly predict carefully developed measures of job performance. The project addresses the Army's 675,000-person enlisted personnel system encompassing several hundred military occupations.

The program began in 1980, when the U.S. Army Research Institute (ARI) started planning the extensive research needed to develop the desired system. In 1982 ARI selected a consortium, led by Human Resources Research Organization (HumRRO) and including American Institutes for Research (AIR) and Personnel Decisions Research Institute (PDRI), to undertake the 9-year project. It is utilizing the services of 40 to 50 ARI and consortium researchers working collegially in a variety of professional specialties. The Project A objectives are to:

- Validate existing selection measures against both existing and project-developed criteria (including both Army-wide job performance measures based on rating scales, and direct hands-on measures of MOS-specific task performance).
- Develop and validate new selection and classification measures.
- Validate intermediate criteria such as training performance, as predictors of later criteria, such as job performance, so that better informed decisions on reassignment and promotion can be made throughout a soldier's career.
- Determine the relative utility to the Army of different performance levels across MOS.
- Estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility for making decisions.

The research design incorporates three main stages of data collection and analysis in an iterative progression of development, testing, evaluation, and further development of selection/classification instruments (predictors) and measures of job performance (criteria). In the first iteratio. file data from fiscal years (FY) 1981/1982 were evaluated to explcre relationships between scores of applicants on the Armed Services Vocational Aptitude Battery (ASVAB), and their later performance in training and their scores on first-tour Skill Qualification Tests (SQT).

For the ensuing research, 19 Military Occupational Specialties (MOS) were selected as a representative sample of the Army's 250+ entry-level MOS. The selection was based on an initial clustering of MOS derived from rated similarities of job content. These MOS account for about 45 percent of Army accessions and provide sample sizes large enough so that race and sex fairness can be empirically evaluated in most MOS.

In the second iteration, a Concurrent Validation design was executed with FY83/84 accessions. A "Preliminary Battery" of perceptual, spatial, temperament, interest, and biodata predictor measures was developed and tested with several thousand soldiers as they entered four MOS. The data from this sample were then used to refine the measures, with further exploration of content and format. The revised set of measures was field tested to assess reliabilities, "fakability," practice effects, and other factors. The resulting predictor battery, the "Trial Battery," was administered together with a comprehensive set of job performance indexes based on job knowledge tests, hands-on job samples, and performance rating measures, in the Concurrent Validation during the summer and fall of 1985. The results of the Concurrent Validation were used to form five performance constructs and to report to the Army incremental validities of the Trial Battery components over ASVAB predictors.

On the basis of testing experience, the "Trial Battery" was revised as the "Experimental Predictor Battery," which in turn is being administered in the third iteration, the Longitudinal Validation stage, which began in the late summer of 1986. All measures are being administered in a true predictive validity design. About 50,000 soldiers across 21 MOS are included in the FY86-87 administration and subsequent first-tour measurement. About 3,500 of these soldiers are expected to be available for second-tour performance measurement in FY91. Three MOS were added to the original 19 (19K, 29E, and 96B), and one of the original MOS was dropped (76W).

For administrative purposes, Project A is divided into five research tasks: Task 1, Validity Analyses, and Data Base Management; Task 2, Developing Predictors of Job Performance; Task 3, Developing Measures of School/Training Success; Task 4, Developing Measures of Army-Wide Performance; Task 5, Developing MOS-Specific Performance Measures.

Activities during the course of Project A research have been reported as follows: FY83, ARI Research Report 1347 and its Technical Appendix, ARI Research Note 83-37; FY84, ARI Research Report 1393 and two related reports, ARI Technical Report 660 and ARI Research Note 85-14; FY85, ARI Technical Report 746 and ARI Research Note 87-54; FY86, ARI Technical Report 792 and ARI Research Note 88-36; FY87, ARI Technical Report (in preparation), and ARI Research Note (in preparation); FY88, ARI Technical Report (in preparation). These reports list other publications on specific Project A activities.

INTRODUCTION

The data from the Concurrent Validation sample have been used to revise and develop more completely a model of job performance for entry-level performance in terms of five basic components. This process was described in the Project A Annual Report for FY 1986 (Campbell, 1986a; also Campbell, 1986b, and Wise, Campbell, McHenry, & Hanser, 1986).

Results have indicated that each of the components can be predicted with considerable validity and that the validity of the different predictor domains varies systematically across criterion components. Yet to be determined is how a composite index of performance can be formed and what the validity of the Trial Battery is for each job, when just one composite of performance is used.

This report describes research conducted to determine the best method to weight the importance of the components of an overall composite index of performance. Weighting judgments were then gathered from noncommissioned officers (NCOs) and officers familiar with each Project A MOS. Analyses of these data are presented in the final sections of the report.

BACKGROUND

Several methods are available for assigning weights to performance dimensions in such a way that they reflect the factors' relative importance to overall performance. Four procedures that have been emphasized in the literature are: 1) the Two-Factor-at-a-Time conjoint procedure; 2) the Full-Profile conjoint procedure; 3) the Kelly Bids system; and 4) the Kane method.

In a conjoint procedure the respondents are asked to rank order, rate, or otherwise choose among two or more sets of profile descriptions that vary along the dimensions of interest. The relative weights for the dimensions can be inferred from the relationships between the dimension values built into the descriptions and the rank orders or ratings (the dependent variable) of the profiles. The Two-Factor-at-a-Time and the Full-Profile approaches have been generally used in conjoint procedures.

The Two-Factor-at-a-Time is also referred to as the Trade-off procedure (Johnson, 1974). In this procedure the performance factors are evaluated on a two-at-a-time basis. The evaluators are usually asked to rank the various combinations of each pair from most preferred to least preferred (Green & Srinivasan, 1978). The advantages of using this procedure are that it is simple, reduces information overload, and lends itself to mail questionnaire administration. It does, however, have some limitations. It has been criticized as being unrealistic because there are other factors that must also be taken into consideration in the overall evaluation. Some researchers (Green, 1974; Johnson & VanDyk, 1975) have pointed out that the total number of required evaluations is quite large when there are multiple levels within the dimensions. In these circumstances the respondents may attend to one dimension first before considering the other (Johnson, 1974).

The Full-Profile approach attempts to address some of the limitations of the Two-Factor-at-a-Time procedure, following the same procedure but utilizing the complete set of factors in the descriptions. It gives a more realistic description of the stimuli being judged by defining the levels on all of the factors, and possibly taking into account the potential environmental correlations between the factors in real stimuli (Green & Srinivasan, 1978). It is, however, not devoid of limitations. Information overload is highly likely as the number of factors in the profile increases. Furthermore, the respondents may simplify the task by ignoring variations in the less important factors or by simplifying the factor levels themselves (Green & Srinivasan, 1978). For these reasons, use of this procedure is generally limited to five or six factors.

The measurement scale used for these conjoint procedures is either non-metric (paired comparisons, rank order) or metric (rating scales assuming interval scales, ratio scales obtained by constant-sum paired comparisons). For the Two-Factor, the non-metric scale is more appropriate because the rank order of the cells in a trade-off table need not depend on the levels of the missing factors, except if the attributes are correlated (Green & Srinivasan, 1978).

The effectiveness of these two procedures has been evaluated by several researchers. Montgomery, Wittink, and Glaze (1977) reported that the Two-Factor procedure yielded higher predictive validity. Their research focused on job choices made by MBAs and used a total of eight attributes. In a study of commuters' choice of transportation modes that varied along nine attributes, Alpert, Betak, and Golden (1978) reported better goodness-of-fit for the Two-Factor procedure. Jain, Acito, Malhotra, and Mahajan (1978), on the other hand, reported that the two methods yielded approximately the same level of cross-validity in the context of choosing checking accounts offered by various banks when the accounts were described via five attributes. Oppedijk van Veen and Beazley (1977) found that the utilities determined by the two methods were roughly similar in the context of a durable good product class when using three attributes.

In the Kelly Bids system for weighting purposes, the respondents are asked to allocate 100 points across the criterion dimensions on the basis of their relative importance. Schmidt (1977) found this procedure better than others because the focus is on the hypothetical "true" criterion.

Kane (1980) maintained that observability and uncertainty should also be considered critical in all appraisal situations. He therefore proposed the Kane method for assigning weights to performance factors. An important aspect to this procedure is the designation of a level of specificity for assigning importance weights (e.g., task level) prior to any activity. The respondents are then asked to identify the component having the least importance for measuring overall effectiveness; this component is assigned a weight of 1.0. The respondents are then asked to compare the remaining factors to the least important component, assigning weights to reflect how many times more important each factor is compared to the least important factor.

All four procedures for assigning weights to performance factors have been shown to work well in a variety of settings. The appropriateness of the methodology depends to a great extent on the purposes and the type of factors and variables of the research endeavor. Consequently, the Project A staff conducted a series of exploratory studies with the various procedures before proceeding to the actual determination of the weights, using the two best methods.

PILOT TESTS OF METHODS FOR WEIGHTING CRITERION COMPONENTS

Three pilot experiments were conducted to select the procedures to be used in weighting constructs for Project A. The primary focus in these studies was on the weighting procedures themselves, not on the weights of the constructs for a given MOS. Our interest in conducting the experiments was in selecting one or more construct weighting procedures that would be acceptable to the Army and would yield a reliable, valid set of weights for each of the sampled MOS when the procedures were applied by the appropriate subject matter experts. The three pilot experiments were related in the sense that the weighting procedure selected as a result of the first experiment was also used in the second and third experiments to further evaluate that and other procedures. The experiments and their results will be described briefly prior to describing the actual factor weighting procedure.

Experiment 1: Procedure and Results

Sixteen Army officers stationed at Fort Meade, Maryland, and Fort Monroe, Virginia, participated in the first experiment. Their task was to assign relative weights to six performance constructs for three Military Occupational Specialties (MOS) -- Infantryman (11B), Wheel Vehicle Repairer (63B), and Administrative Specialist (71L).

At the time the experiment was conducted in the summer of 1985, the Project A performance constructs had not yet been selected. Therefore, a plausible set of six constructs whose weights might be expected to vary considerably was used instead. The six performance constructs were 1) dependability, 2) MOS-specific task performance, 3) MOS knowledge, 4) military bearing, 5) performance under adverse conditions, and 6) performance on common, general soldiering tasks (e.g., putting on a gas mask). The construct weights for the three MOS were assigned by the officers under a replicated 3 x 3 Graeco-Latin square design in which three weighting procedures were used under three different military scenarios (see Figure 1).

The three procedures all involved direct judgments of the relative weight that each performance construct should receive in forming an overall composite performance score. In procedure A, the officers were first asked to rank order the six constructs, and then to assign 100 points to the first-ranked construct and to scale the other constructs accordingly (this is a variant of the Kane method). In procedure B, the officers were instructed to divide 100 points among the six constructs in a manner that reflected the relative weight that should be given the constructs. In

Scaling Methods: Maximum 100 points (A), divide 100 points (B), paired comparison (C).

Military Scenario: Wartime (a), period of heightened tensions (b), peacetime (c).

Number of Subjects	MOS 11B	MOS 63W	MOS 71L
2	Aa	Bb	Cc
1	Bc	Ca	Ab
1	Cb	Ac	Ba
	MOS 63W	MOS 71L	MOS 11B
2	Aa	Bb	Cc
2	Bc	Ca	Ab
2	Cb	Ac	Ba
	MOS 71L	MOS 11B	MOS 63W
2	Aa	Bb	Cc
2	Bc	Ca	Ab
2	Cb	Ac	Ba

Figure 1. Replicated Graeco-Latin Square design.

procedure C, 15 pairs of the six factors were presented in a paired comparison protocol; the order of presentation followed the optimization procedure worked out by Ross (1934). The officers' task was to divide 100 points between the two constructs being judged in any given pair.

The judgments were made in the context of three different scenarios (see Figure 2). The scenarios described respectively a peacetime condition, a period of heightened tensions, and a wartime setting in which hostilities had just broken out. The site (i.e., Europe) of the three scenarios was the same.

After completing the construct weighting judgments, each officer used four 7-point scales to evaluate the weighting methods on four dimensions.

PEACETIME SCENARIO

Europe is in the peacetime condition currently prevailing there. Your Corps' mission is to defend and maintain the host country's border should war break out. The potential enemy approximates a combined arms Army and has nuclear and chemical capability. Air parity does exist. The Corps has personnel and equipment sufficient to make it mission capable for training and evaluation. The training cycle includes periodic field exercises, command and maintenance inspections, ARTEP evaluations, and individual soldier training/SQT testing.

HEIGHTENED TENSIONS SCENARIO

Europe is in a period of heightened tensions. There is an increasing probability that hostilities will break out in the next several months. Your Corps' mission is to defend and maintain the host country's border should war break out. The potential enemy approximates a combined arms Army and has nuclear and chemical capability. Air parity does exist. The Corps' training and other preparatory activities have been substantially increased. Most combat and associated support units are participating in frequent field exercises. Most units are being actively resupplied.

WARTIME SCENARIO

Hostilities have broken out in Europe and your Corps' combat units are engaged. Your Corps' mission is to defend, then reestablish, the host country's border. Pockets of enemy airborne/heliborne and guerilla elements are operating throughout the Corps sector area. Limited initial and reactive chemical strikes have been employed but nuclear strikes have not been initiated. Air parity does exist.

Figure 2. Three different military scenarios.

1) Acceptability to the Army,

2) Ease of making the judgments called for by the method.

3) Their confidence in the validity of the judgments made.

4) The amount of agreement with other workshop participants that could be expected.

The relevant mean ratings across the four dimensions are shown in Table 1. After the officers completed rating the methods, an informal discussion was held to solicit their opinions about the methods.

The design permitted testing for the significance of differences in mean ratings on the four dimensions for procedures and for scenarios, and for any Procedure X Scenario interactions. None of the main effects due to the scaling procedure or scenario were significant. However, significant ($p<.05$) Procedure X Scenario interactions were obtained for the acceptability to the Army and the raters' confidence in their judgment scales, and for the average of the four scales. Procedure A (in which 100 points were assigned to the first-ranked construct) had particularly low ratings when combined with the peacetime scenario, but had relatively high ratings when combined with the wartime and heightened tension scenarios.

The officers generally expressed preference for procedures A and C over procedure B, and thought that the time they spent in Procedure B in making sure that the sum of their weights equaled 100 detracted from their ability to judge the relative importance of the performance factors. It was evident that if a larger number of constructs were ultimately identified, procedures B and C could become fairly onerous.

The officers also expressed a general preference for the heightened tensions and wartime scenarios over the peacetime scenario as the setting for the judgments. Based on the discussion, it also seemed that a heightened tension scenario would evoke a more uniform frame of reference across the many different kinds of subject matter experts providing the MOS construct weights than a wartime scenario would, unless the wartime scenario was made quite specific. However, specificity in the scenario could produce unwanted dependency of the construct weights on particular elements in the scenario, which could detract from the validity of the weighted composite as an overall, general measure of MOS performance.

Experiment 2: Procedure and Results

The second pilot experiment was conducted in the winter of 1985 at Fort Bragg, North Carolina, using two 4-hour workshops. One workshop was attended by 15 officers, the other by 15 NCOs. The workshop participants were asked to weight five performance constructs for the Infantry MOS: 1) demonstrating commitment to the Army, 2) technical proficiency and knowledge, 3) physical fitness and military bearing, 4) performance under adverse conditions, and 5) maintaining and servicing weapons and equipment.

Table 1

Experiment 1: Mean[a] Ratings[b] of Nine Weighting Procedure/Scenario Combinations

	Scenario		
Procedure	Peacetime	Heightened Tensions	Wartime
A. Maximum = 100 points	2.85	4.75	4.79
B. Divide 100 points	4.95	5.12	4.20
C. Paired Comparison	4.62	4.60	4.35

[a] Separate means based on ratings of five or six officers.
[b] Seven-point rating scales in which 1 = Low and 7 = High.

Each participant used the three different weighting methods described in the following instructions:

1) Rank order the five constructs, assign 100 points to the first ranked construct, and then scale the other constructs accordingly (same as procedure A in Experiment 1).

2) Based upon their scores on the separate constructs, rank order 25 infantrymen in order of their overall performance. For each infantryman, a different set of performance scores on the five constructs was given on 7-point scales that range from the lowest level of performance to the highest. A sample profile is shown in Figure 3.

3) Based upon their scores on two constructs, rank order 10 sets of 13 infantrymen in order of their overall performance. In each set, the performance scores on two constructs are given on the same 7-point scales used in the second method above. A set of 13 infantrymen is given for each of the 10 possible pairs of the five constructs. (See Figure 4.)

The second and third methods are variants of the conjoint approach to scaling in which, instead of obtaining the relative importance of the performance constructs directly, the judges' weights for the performance constructs are inferred from the rank order they give sets of hypothetical soldiers whose performance on the constructs has been systematically varied. Multiple regression weights are calculated from the interrelationships between the rank orders provided by the judges and the performance construct levels given in the performance descriptions. In the paired comparison method, these regression weights are then used to derive the construct weights, using a ratio scaling procedure described by Torgerson (1958, pp.

Soldier ____ Rank Order ________

Overall Score ________

MOS: Infantryman (11B)

A. DEMONSTRATING COMMITMENT TO THE ARMY
Maintaining Army traditions, spirit and fellowship.

Shows lack of dedication to Army traditions and values.		Generally supports Army traditions and values.			Shows constant devotion to Army tradition and values.	
1	2	3	(4)	5	6	7

B. PHYSICAL FITNESS AND MILITARY APPEARANCE
Maintaining military standards of physcial fitness; maintaining proper military appearance and standards of cleanliness and grooming.

Maintains self in poor physical condition. Fails to meet military standards for dress and personal hygiene.		Meets Army standards of physical fitness. Dresses neatly and meets Army standards of personal hygiene.			Exceeds Army standards and expectations set for physical fitness. Maintains excellent personal hygiene and proper appearance.	
1	2	(3)	4	5	6	7

C. MAINTAINING AND SERVICING WEAPONS AND EQUIPMENT
Keeping weapons and equipment clean and serviced and prepared for the field.

Fails to perform or improperly performs checks and preventive maintenance on weapons and equipment.		Performs routine checks and preventive maintenance on weapons and equipment.			Always keeps assigned weapons and equipment in ready-for-inspection condition.	
1	2	3	(4)	5	6	7

D. TECHNICAL PROFICIENCY AND KNOWLEDGE
Effectiveness in applying technical knowledge and proficiency in carrying out MOS tasks.

Does not display the knowledge/skill required to perform many job assignments and tasks.		Displays the knowledge/skill required to perform most job assignments and tasks properly, but may need help for harder tasks			Displays the knowledge/skill to perform all job assignments and tasks properly.	
1	2	3	4	(5)	6	7

E. PERFORMANCE UNDER ADVERSE CONDITIONS
Continuing to execute appropriate soldier skills under combat conditions or under hardship, stressful or otherwise difficult circumstances.

Makes frequent mistakes in combat situations or otherwise stressful situations.		Makes mistakes infrequently in combat or otherwise stressful situations.			Almost never makes mistakes in combat or otherwise stressful situations.	
1	2	3	4	(5)	6	7

Judge No.______

Figure 3. Sample MOS 11B Profile Form.

Name______________________ Sheet No. 01

OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Demonstrating Commitment to the Army	Technical Proficiency and Knowledge	Rank Order	Overall Score
1	5	5	____	____
2	1	4	____	____
3	2	6	____	____
4	4	7	____	____
5	4	4	____	____
6	6	5	____	____
7	6	2	____	____
8	3	2	____	____
9	4	1	____	____
10	5	6	____	____
11	2	3	____	____
12	3	3	____	____
13	7	4	____	____

Performance Scales:

DEMONSTRATING COMMITMENT TO THE ARMY
Maintaining Army traditions, spirit and fellowship.

Shows lack of dedication to Army traditions and values.		Generally supports Army traditions and values.			Shows constant devotion to Army tradition and values.	
1	2	3	4	5	6	7

TECHNICAL PROFICIENCY AND KNOWLEDGE
Effectiveness in applying technical knowledge and proficiency in carrying out MOS tasks.

Does not display the knowledge/skill required to perform many job assignments and tasks.		Displays the knowledge/skill required to perform most job assignments and tasks properly, but may need help for harder tasks			Displays the knowledge/skill to perform all job assignments and tasks properly.	
1	2	3	4	5	6	7

Figure 4. Example of Overall Performance Score Sheets.

105-112). This procedure results in a set of scale values or weights for the constructs whose geometric mean is equal to 1.0.

The judgments were made in the context of a worldwide increase in tensions (see Figure 5). The weighting methods were applied in counterbalanced order by the 15 participants in each workshop. After completing each method, the participants rated the method on the four 7-point scales used in the first experiment.

The world is in a period of heightened tensions. There is an increasing probability that hostilities will break out in Europe, Asia, the Caribbean, Latin America, and Africa. The Army's mission is to support U.S. treaty obligations and to help defend the borders of allied and friendly nations. Some of the potential enemies have nuclear and chemical capability. Air parity does exist between allied forces and potential hostile nations. U.S. Army training and other preparatory activities have been substantially increased. Most combat and associated support units are participating in frequent field exercises. Most units are being actively resupplied.

Figure 5. Worldwide Increase in Tensions Scenario.

Table 2 presents the mean ratings given the three weighting methods by the 30 workshop participants, along with the results of analysis of variance tests of the significance of the method mean differences. The ratings clearly favored the direct estimation method, while the full-profile conjoint method, which involved rank ordering the descriptions of 25 hypothetical infantrymen, generally received the lowest ratings. A breakout of these ratings by type of judge indicated that both the officers and the NCOs generally preferred the direct estimation method most and the full-profile conjoint method least.

The methods were also compared on three other dimensions: judge reliability (intraclass correlation), correlation between mean weights assigned by the officers and NCOs, and the intercorrelations among the sets of mean weights obtained by the three methods for all participants. These statistics are shown in Table 3.

In general, the conjoint paired-comparison method yielded the highest intraclass correlations for both officers and NCOs while the conjoint full-profile method had the lowest values. The correlation between the mean officer and NCO weights obtained from the conjoint paired-comparisons method also was the highest (r = .91), while the conjoint full-profile officer/NCO correlation was the lowest (r = .60). The mean weights obtained from the

Table 2

Experiment 2: Mean Ratings[a] of Weighting Methods (n = 15 officers, 15 NCOs)

Weighting Method	Acceptability	Ease	Validity	Agreement	Average Rating
Direct estimation	4.30	5.13	5.80	4.77	5.00
Conjoint paired-comparison	4.23	4.13	5.17	4.50	4.51
Conjoint full-profile	4.27	3.87	5.10	4.23	4.37
Significance	.020	.002	.048	NS	.04

[a] Seven-point scales in which 1 = Low and 7 = High.

Table 3

Experiment 2: Agreement Indexes for Weighting Methods

	One-Rater Reliability			Correlation	Intercorrelation	
Weighting Method	Officer	NCO	All	Off/NCO Means	Full Profile	Paired Comp
Direct estimation	.27	.24	.25	.81	.17	.93
Conjoint full-profile	.23	.01	.11	.60		.15
Conjoint paired-comparison	.54	.32	.42	.91		

direct estimation and the conjoint paired-comparisons were highly correlated (r = .93) while the correlations of these weights with those obtained from the conjoint full-profile method were quite low. On the basis of these results and the participant method evaluations described earlier, it was decided to drop the conjoint full-profile method from further consideration.

Experiment 3: Procedure and Results

The third pilot study was also conducted in the winter of 1985, at Fort Bragg, using two 4-hour workshops. One workshop was attended by seven officers, the other by eight NCOs. The workshop participants were asked to weight seven performance constructs for the infantry MOS. The seven constructs included the five used in the second weighting method experiment plus two additional ones--avoiding serious disciplinary problems and providing peer leadership and support.

Each participant used the three different weighting methods described below and in the following order:

1) Based on scores on two constructs, participants were asked to rank order 21 sets of 13 infantrymen in order of their overall performance. This is the same basic conjoint paired-comparison procedure used in the second experiment. In this case, however, in addition to rank ordering the 13 infantrymen, the judges assigned performance scores that reflected the soldiers' relative overall performance.

2) The participants were then asked to rank order the seven constructs, assign 100 points to the first-ranked construct, and then scale the other constructs accordingly (the direct estimation procedure used in Experiments 1 and 2).

3) The third method was a variant of the second and incorporated a Delphi procedure. Participants first indicated why the performance factors were ranked and weighted as they were in method 2 above. These reasons were passed around to the other workshop participants; also passed around were the average and range of the weights given each performance factor by the workshop participants in method 2. After considering this feedback information, the participants reassigned weights to the performance factors using method 2 above. The Delphi procedure was then repeated once more.

The above judgments were made in the same context of a worldwide increase in tensions that was used in Experiment 2. After completing each method, the participants rated the method on the same four 7-point scales used in the first and second experiments.

Table 4 presents the mean ratings given the three weighting methods by the 15 workshop participants, along with the results of analysis of variance tests of the significance of the method mean differences. The ratings for the direct estimation and modified Delphi methods were generally higher than those given the conjoint paired-comparison method.

Table 4

Experiment 3: Mean Ratings[a] of Weighting Methods
(n = 7 Officers, 8 NCOs)

Weighting Method	Acceptability	Ease	Validity	Agreement	Average Rating
Conjoint paired-comparison	3.43	4.20	4.60	3.86	4.02
Direct estimation	4.21	5.27	5.80	4.57	4.95
Modified Delphi	4.46	5.43	5.93	4.62	5.09
Significance	NS	.049	.010	NS	.002

[a] Seven-point scales in which 1 = Low and 7 = High.

It is interesting to note that while the mean ratings given the direct estimation method in Experiments 2 and 3 (see Tables 2 and 3) were generally quite similar, the conjoint paired-comparison method generally received lower ratings in Experiment 3 than in Experiment 2, although only the mean acceptability ratings for this conjoint method were significantly different across the two experiments (4.23 vs. 3.43).

The weighting methods used in Experiment 3 were also compared on interjudge reliability (intraclass correlation), correlation between mean weights assigned by officers and NCOs, and intercorrelations among the sets of mean weights obtained by the three methods. For the conjoint paired-comparison method, weights could be derived by using only the rank orders provided by the judges, or by using the overall performance scores assigned the sets of 13 infantrymen. Similarly, for the modified Delphi method, weights could be obtained from the participants' judgments after the first round of feedback or after the second and final round of feedback.

One-rater reliabilities were therefore calculated for five different procedures of obtaining weights from the judgments provided by the workshop participants. These reliabilities, along with the correlations of the mean weights of the officer and NCO participants, are shown in Table 5. The correlations obtained between the five sets of mean weights are shown in Table 6. Also shown in Table 6 are the intercorrelations across weights of the five common constructs used in Experiments 2 and 3 for all the methods used in the two experiments.

Table 5

Experiment 3: Agreement Indexes for Weighting Methods

Weighting Method	One-Rater Reliability Officer	One-Rater Reliability NCO	One-Rater Reliability All	Correlation Off/NCO Means
Conjoint PC Ranking	.43	.27	.35	.84
Conjoint PC Scores	.32	.20	.27	.87
Direct Estimation	.28	.20	.25	.84
Delphi Round 1	.26	.18	.22	.75
Delphi Round 2	.32	.18	.24	.77

Table 6

Experiments 2 and 3: Intercorrelations of Mean Weights Obtained From the Weighting Methods Used in Both Experiments

Weighting Method	No. of Constructs	Conjoint PC Ranking	PC Scores	Direct Est.	Delphi Round 1	Delphi Round 2	Direct Est.	Conjoint Full Profile
Conjoint PC Ranking	7	-						
Conjoint PC Scores	7	.96	-					
Direct Estimation	7	.73	.86	-				
Delphi Round 1	7	.65	.80	.96	-			
Delphi Round 2	7	.64	.80	.99	.97	-		
Direct Est. (Exp 2)	5	.82	.91	.96	.93	.93	-	
Conj Full-Prof (Exp 2)	5	.12	.19	.36	.44	.44	.17	-
Conj Paired-Comp (Exp 2)	5	.97	.98	.87	.37	.31	.93	.15

Several inferences can be drawn from the data presented in Tables 5 and 6. First, there is no evidence that the one-rater reliabilities or the correlations obtained from the officers and the NCOs are improved substantially by adding the requirement to provide overall performance scores as well as rankings in the conjoint paired-comparison method. Nor are these agreement indexes improved by adding one or two rounds of Delphi feedback to the direct estimation method. Moreover, the correlations between weights obtained through the two basic methods (conjoint paired-comparisons ranking and direct estimation) and the weights obtained through their respective extensions (conjoint paired-comparison scores and Delphi-rounds 1 and 2) ranged from .96 to .99.

Two other considerations led us to decide not to require that the judges assign overall performance scores in addition to rank ordering the sets of soldiers in any future application of the conjoint paired-comparison method. First, from a practical point of view, the requirement to assign performance scores added about two minutes, on the average, to the amount of time a judge takes to complete the judgment for one set of 13 hypothetical soldiers.

The second consideration has to do with the assumption one makes about the soldiers' scores on the constructs that are <u>not</u> being immediately compared in the paired-comparison protocol. The overall performance scores assigned the set of soldiers for the pair of constructs being judged might be different if one assumes that these other construct scores are all high, than if one assumes that these scores are low, average, or mixed. The rank orders, on the other hand, should not be so influenced.

Similar considerations led us to decide not to use the modified Delphi method in addition to the direct estimation method.

The choice between the direct estimation method and the conjoint paired-comparison ranking method was not an easy one. The direct estimation method generally received higher evaluation ratings in both Experiment 2 and 3 and would obviously take less time to administer than the conjoint method. On the other hand, the officer and NCO one-rater reliabilities obtained for the conjoint method were higher than for direct estimation in both experiments. However, for both the direct estimation and paired-comparison methods the correlations between the officer and NCO mean weights were above .80 in both experiments. The correlations between the mean weights obtained in Experiment 2 and those obtained in Experiment 3 were very high for both methods (.96 for the direct estimation and .97 for the conjoint method).

In short, although each method might have some advantages over the other, both appeared to be sound methods of obtaining performance construct weights. We therefore decided to use both methods to weight the performance constructs for the Project A MOS sample.

OBTAINING PERFORMANCE CONSTRUCT WEIGHTS FOR PROJECT A MOS

Procedure

The component weighting judgments for Project A MOS were collected in a series of 2-hour workshops. Separate workshops were held for NCOs and officers at each of two posts for each of 20 MOS. One of these posts housed the proponent school for the MOS and the other housed field units having officers and NCOs with expert knowledge of the MOS.

At each workshop, after a briefing on Project A, the participants were first given general instructions which covered the background and purpose of the workshop, and descriptions of the performance components (constructs) and the two methods (direct estimation and conjoint paired-comparison ranking) that would be used to obtain weights for the components.

The components to be weighted were the five job performance criterion factors that had been developed as part of Project A's performance modeling effort (Campbell, 1986, Chapter 7). The components were:

1) Task proficiency: MOS-specific technical skills.

2) Task proficiency: General soldiering skills.

3) Exercise of leadership, effort, and self-development.

4) Maintaining personal discipline.

5) Military bearing/appearance and physical fitness.

The two scaling methods were then administered, always in the same order. The participants were given a short break between methods.

Sample of Judges

The sample plan called for a total of 36 judges for each MOS, half coming from field units (FORSCOM and USAREUR) and half from proponent posts (TRADOC). The judges were to be evenly divided among NCOs, company grade officers, and field grade officers. However, the target sample composition was not attained for every MOS. In some cases where sufficient numbers of officers and/or NCOs were not available, warrant officers who knew the jobs well were used in lieu of company or field grade officers. Table 7 shows the total sample of 712 judges identified by MOS, type of post, and grade level. Although some individual MOS proportions did not meet the target, overall the proportions of officers to NCOs and of judges from field units to proponent MOS posts were quite close to the desired composition.

The Scaling Methods

On the basis of the results of the earlier exploratory experiments, two methods were used to obtain importance weights for the five performance constructs.

Table 7

Composition of Judging Sample[a] for Project A MOS

		Type of Unit					
		Field		Proponent		Total	
MOS		Officer	NCO	Officer	NCO	Officer	NCO
11B	Infantryman	17	6	19	6	36	12
12B	Combat Engineer	17	4	12	6	29	10
13B	Cannon Crewman	6	6	21	6	27	12
16S	MANPADS Crewman	11	6	11	5	22	11
19E	Armor Crewman	11	5	14	6	25	11
27E	TOW/Dragon Repairer		6	16	5	16	11
31C	Single Channel Radio Operator	13	6	12	6	25	12
51B	Carpentry/Masonry Specialist	4	6	27	6	31	12
54E	Chemical Operations Specialist	20	14			20	14
55B	Ammunition Specialist	4	3	24	9	28	12
63B	Light Wheel Vehicle Mechanic	7	2	20	11	27	13
64C	Motor Transport Operator	10	5	12	6	22	11
67N	Utility Helicopter Repairer	12	1	17	12	29	13
71L	Administrative Specialist	13	6	9	7	22	13
76W	Petroleum Supply Specialist	10	11			10	11
76Y	Unit Supply Specialist	15	5	8	5	23	10
91A	Medical Specialist	25	13			25	13
94B	Food Service Specialist	12	7	8	4	20	11
95B	Military Police	23	13			23	13
96B	Military Intelligence Analyst			11	6	11	6
		230	125	241	106	471	231

[a] In addition to the 702 officers and NCOs listed in this table, there were 10 judges whose grades were unknown, making the total sample 712.

Direct Estimation. The judges first rank ordered the five constructs in terms of their relative importance for deriving an overall performance measure in the given MOS. After assigning 100 points to the most important performance construct, the judges scaled the other four constructs by assigning values that reflected the importance they felt each construct should have in the total effectiveness score. The judges were allowed to give any relative weight from 0 to 100 to the other constructs. After they initially assigned points to the constructs, the judges were told to review the weights they had assigned and make sure that they were in correct proportion to one another.

Conjoint Paired-Comparison. The judges were given performance profiles on 10 sets of 15 hypothetical soldiers in the MOS. The 15 soldiers in any one set had different scores on two of the constructs. The judgmental task was to rank the 15 soldiers in order of their overall performance. When the judges were satisfied with their ranking on one set, they proceeded to the next set of 15 soldiers, who had scores on two other constructs. The order of presentation of the ten pairs of five constructs was governed by the optimization procedure worked out by Ross (1934). The order of presentation of the 15 soldiers on the score sheets was originally randomized, but for ease in making the judgments and processing the data the order remained the same for all 10 sets of soldiers for all MOS. However, the order of presentation of the 10 pairs of constructs was randomized across MOS.

In the conjoint method, the weights assigned by the judges must be inferred from their rank ordering of the 15 hypothetical soldiers. Presumably, if a judge consistently gave a higher rank to soldiers with high performance scores on one construct than to soldiers high on the other construct, then the judge considers the first construct more important in overall MOS performance than the second construct.

The judges accomplished the two methods in the order listed above. The full set of instructions and materials used to collect the weighting judgments for the Infantryman MOS (11B) is given in Appendix A. The judges were given definitions of the performance constructs to study before they made their judgments. They were asked to assume that performance scores for the given MOS were available only on the constructs given.

The judges were further asked to assume that the military context or scenario in which the soldiers' performance was being evaluated was the following:

> The world is in a period of heightened tensions. There is an increasing probability that hostilities will break out in Europe, Asia, the Caribbean, Latin America, and Africa. The Army's mission is to support U.S. treaty obligations and to help defend the borders of allied and friendly nations. Some of the potential enemies have nuclear and chemical capability. Air parity does exist between allied forces and potential hostile nations. U.S. Army training and other preparatory activities have been substantially increased. Most combat and associated support units are participating in frequent field exercises. Most units are being actively resupplied.

ANALYSIS

Data Transformation

The direct estimation scaling method yielded weights on a scale ranging from 0 to 100. The range and distribution of direct estimation weights varied considerably among the judges. To better reflect the combined judgments of the construct weights across the judges for each MOS, the data from each judge were standardized prior to averaging--a procedure that would tend to equalize the judges' contributions to the MOS mean even though they may have assigned rather disparate sets of weights to the constructs.

To preserve the relative size of the weights that each judge had assigned the constructs, each judge's weights were transformed by multiplying them by a constant (the ratio, 100/sum of the judge's weights). This caused the five construct weights of each judge to sum to 100, but did not change the relative values of the judge's weights. Consequently, the average of the five construct weights of all judges was set at 20.0, and the average of the five weights for any group of judges within and across MOS was also set at 20.0. The mean weight of a given construct obtained by averaging the judges' individual weights could, of course, be different from 20.

For the conjoint method, the data from each judge were scaled using a method developed by Comrey (1950) which is described in Torgerson (1958). Essentially, the multiple regression equation predicting the judge's rank orders of the two performance construct scores of the 15 hypothetical soldiers was first obtained for each of the 10 sets of soldiers. The ratio of the two regression weights for each pair of constructs then became the basic data entering into the scaling procedure. Since the correlation between the two construct scores of the 15 hypothetical soldiers on each performance rating sheet was specified to be zero, the ratio of the regression weights is directly proportional to the correlation of each set of construct scores with the judge's rank order of the soldiers. (The mean and standard deviations of the construct scores were equal for all constructs.)

The scaling procedure employs a least squares solution to obtain a set of weights that best fit the observed ratios. The resultant weights are so scaled that their geometric mean is 1.0. To facilitate comparing the conjoint weights with those obtained by the direct estimation method, the conjoint weights for each judge were also linearly transformed so that their sum was equal to 100 and their average equal to 20.0.

One reason for effecting the transformation concerned the practical application of the weights to the construct scores. The final intent is to apply a set of weights to the construct variance/covariance matrix such that the covariance of each construct with the composite total score is equivalent to the construct weight obtained from the judges. In other words, the contribution to the total MOS performance variance of each construct would be directly proportional to its weight. A separate algorithm will be used to calculate the weights that, when applied to the variance/covariance matrix, yield the desired (the obtained scaled) weights or contributions to the total composite variance.

<u>Examination of Missing Data</u>

For the conjoint scaling method, 73 of the judges either had failed to complete the entire judgmental sequence or had recorded judgments that were inconsistent with the assumptions of the scaling method involved. For example, a judge may have completed all the performance score sheets, but one or more of the ten resultant regression equations had constructs with a positive weight. This would mean that the higher a judge rank ordered the 15 hypothetical soldiers on the given score sheet, the lower were the soldiers' scores on one of the constructs. However, the scaling method employed (see Torgeson, op. cit.) required that both weights have the same sign and that a full set of weight ratios be available. Consequently, either the conjoint protocols with missing or positive weights could be eliminated or the missing weight ratios could be imputed by an appropriate estimation technique. As can be seen in Table 8, proportionately more NCOs than officers had one or more problems of this nature in their conjoint protocols.

Table 8

Conjoint Method Missing or Invalid Data by Grade

	Grade			
Amount of Missing Data	NCO[a]	Officers	Unknown	Total
1 data element missing or invalid	16	19	0	35
2 or more data elements missing or invalid	22	16	0	38
No data missing	193	436	10	639
Total	231	471	10	712

[a] Proportion of NCOs having missing data is higher than officers (χ^2 = 13.53) df = 2, significant at .01 level).

In order to keep at least some of these judges, the missing weight ratios for those judges who had only one conjoint performance score sheet uncompleted or who had one pair of weights of opposite sign were estimated by the technique described below. Judges with two or more problems in their conjoint data (38) were dropped from the conjoint data set.

The imputed estimates of the weight ratios were obtained by first correlating the judge's nonmissing ratios with the ratios of other judges within the MOS who had no missing data, and then computing a stepwise multiple regression equation to predict the missing ratios. No equation could be computed for seven of the 35 judges with one key data element missing because no other judge had values sufficiently correlated with these judges' ratios; these seven were dropped from the analysis.

The 28 judges for whom we imputed the missing regression weight ratios were then compared with the remaining judges on two indexes:

1) The correlation between each judge's set of weights produced by the direct estimation and the conjoint scaling methods.

2) The consistency with which a judge rank ordered the 15 hypothetical soldiers on the basis of their construct scores. For example, if a judge ranked a hypothetical soldier with scores of "5" and "3" on two performance scales lower than another soldier with scores of "3" and "3", the judge would be giving a higher rank to a poorer performing soldier. In the set of 10 conjoint performance score sheets, a maximum of 630 such rank order inversions was possible.

The judges with imputed conjoint scale values had somewhat lower direct estimation/conjoint correlations between their scale values than did judges with complete data and also had more inversions in the rank orders they assigned to hypothetical soldiers listed on the conjoint performance score sheets (see Table 9). Consequently, these judges were also dropped from the analyses of the conjoint data.

<u>Analyses of Outliers</u>

As was seen in Table 9, a number of the remaining 639 judges had a large number of inversions in the rank orders they assigned the 15 hypothetical soldiers. A within-MOS analysis was conducted in which judges with the highest number of inversions were progressively dropped from the sample. After each successive judge was dropped, the average 1-rater and n-rater intraclass correlations or reliability coefficients for the remaining pool of judges were calculated. The average n-rater reliabilities across the 20 MOS proved to be highest when the two judges with the largest number of inversions were eliminated.

Consequently, the two judges in each MOS who had the highest number of inversions were dropped, provided that they had at least 30 inversions in their protocol. In addition, any judge with 90 or more inversions was dropped even if that meant that more than two judges were eliminated for a given MOS. Altogether, 40 judges were dropped. The average 1-rater and n-rater reliabilities across the 20 MOS were .221 and .879, respectively, before the 40 judges were dropped, and .236 and .881 after they were eliminated.

While these gains in reliability for the conjoint judgments were not large, direct estimation reliabilities also improved with use of the reduced sample. The 1-rater and n-rater reliabilities for the direct estimation method averaged .186 and .854 when all the 712 judges were used but rose to .223 and .863 with the reduced sample of 599. Dropping judges who apparently had not accomplished the conjoint procedure carefully helped improve the reliability of the weights assigned the constructs under both methods. All remaining analyses were carried out on the reduced sample of 599.

Table 9

Frequency Distribution of Inversions Made by Judges With Imputed and Not-Imputed Conjoint Ratios[a]

No. of Inversions	Imputed	Not Imputed	All Judges
0	4	75	75
1-19	7	412	416
20-39	8	98	105
40-59	3	29	37
60-79	3	10	13
80-99	1	7	10
100-119		2	3
120-139		3	3
140-159	2		2
160-179		2	2
180-199		1	1
Total	28	639	667

[a] Median test results: $\chi^2 = 25.28$ (significant at .001 level).

RESULTS

Interjudge Reliability and Intermethod Agreement

Table 10 shows the intraclass reliabilities of the direct estimation and conjoint weights by grade and MOS. The average NCO 1-rater and n-rater reliabilities for the direct estimation and conjoint scaling methods were .132/.425 and .153/.509 respectively. The corresponding values for officers were .278/.864 and .287/.867.

As shown in Table 11, the correlations across the 20 MOS of the average weights derived from the direct estimation and conjoint scaling methods using officer judgments ranged from .836 to .996; the average intermethod agreement was .951. The corresponding range for the NCOs was .017 to .922 and their average MOS intermethod agreement was .653. These intermethod results reflect in part that lower 1-rater reliabilities were obtained for the NCOs under both methods and also that there were fewer NCO judges.

Another factor that may have played a role was the greater homogeneity of the weights that the NCOs assigned the five constructs. The average of the standard deviations of the weights assigned by the individual NCOs across both methods was 6.43, while the corresponding officer average standard deviation was 7.69 (see Table 12). The difference between these means was statistically significant (.001 level).

Table 10

Intraclass Reliabilities of Direct Estimation and Conjoint Weights by Grade and MOS

	NCOs					Officers				
		Direct Estimation		Conjoint			Direct Estimation		Conjoint	
MOS	n	1-rater	n-rater	1-rater	n-rater	n	1-rater	n-rater	1-rater	n-rater
11B	9	.008	.070	-	-	28	.346	.937	.317	.929
12B	8	.337	.802	.171	.623	27	.259	.904	.398	.947
13B	8	-	-	.209	.679	26	.331	.928	.481	.960
16S	7	.160	.572	.154	.561	21	.437	.942	.358	.921
19E	9	.287	.784	.401	.858	21	.347	.918	.350	.919
27E	6	-	-	.114	.435	15	.253	.835	.252	.835
31C	7	.308	.757	.299	.749	24	.207	.862	.257	.893
51B	10	-	-	-	-	27	.184	.859	.215	.881
54E	11	.070	.454	.151	.662	20	.201	.835	.183	.818
55B	8	-	-	.218	.691	25	.369	.936	.312	.919
63B	10	-	-	.078	.459	22	.430	.943	.327	.915
64C	9	.184	.669	.023	.173	20	.224	.852	.185	.819
67N	10	.211	.728	.201	.715	27	.362	.939	.332	.931
71L	9	.015	.117	.031	.223	20	.239	.863	.229	.856
76W	5	.113	.389	.022	.103	9	.123	.558	.125	.561
76Y	6	.240	.654	.319	.737	20	.236	.861	.118	.727
91A	12	.112	.603	.158	.693	24	.352	.929	.386	.938
94B	9	.115	.539	.133	.581	18	.285	.878	.216	.833
95B	10	.247	.766	.188	.699	21	.188	.829	.347	.918
96B	5	.231	.600	.193	.545	9	.192	.682	.344	.825
Average		.132	.425	.153	.509		.278	.864	.287	.867

NOTE: Intraclass reliabilities that were negative are indicated in the table with -. In arriving at the average reliability across MOS, these values were treated as zeros.

Table 11

Correlations Between Construct Weights by Method, Grade, and MOS

	Direct Estimation With Conjoint			NCOs With Officers	
MOS	NCOs	Officers	Total	Direct Estimation	Conjoint
11B	.211	.963	.931	.590	.546
12B	.897	.973	.973	.865	.631
13B	.858	.980	.983	.607	.648
16S	.571	.957	.932	.707	.776
19E	.891	.935	.944	.695	.888
27E	.691	.836	.783	.820	.908
31C	.822	.989	.986	.649	.857
51B	.085	.983	.955	.515	.719
54E	.921	.965	.980	.563	.670
55B	.737	.866	.939	-.107	.100
63B	.551	.968	.987	.615	.837
64C	.017	.985	.961	.796	.364
67N	.922	.996	.991	.819	.965
71L	.772	.966	.962	.919	.968
76W	.575	.946	.956	-.379	.451
76Y	.780	.919	.942	.677	.965
91A	.805	.975	.964	.696	.940
94B	.685	.984	.981	.556	.810
95B	.731	.884	.918	.773	.737
96B	.542	.947	.924	.958	.329
Average	.653	.951	.950	.617	.705

Table 12

Average Standard Deviation of the Construct Weights Assigned by the Judges

Grade	n	Direct Estimation	Conjoint	Total
NCO	168	4.53	8.33	6.43
Officer	424	5.83	9.54	7.69
Total	592	5.46	9.20	7.33

Comparison of Direct Estimation and Conjoint Scaling Methods

To decide whether the final set of weights should be obtained from the direct estimation or the conjoint method, the two sets of weights were compared on several indexes. Though in general the differences were slight, they all favored the conjoint method. The 1-rater and n-rater intraclass reliabilities for the combined group of officers and NCOs tended to be slightly higher for the conjoint method across the 20 MOS (see Table 13). While the differences between the reliabilities for the two scaling methods were slightly greater for the NCOs than for the officers, the difference favored the conjoint method in each case.

Table 13

Intraclass Reliabilities of Direct Estimation and Conjoint Weights by MOS (Officers and NCOs Combined)

		Direct Estimation		Conjoint	
MOS	n	1-rater	n-rater	1-rater	n-rater
11B	37	.261	.929	.236	.920
12B	35	.273	.929	.324	.944
13B	34	.249	.918	.356	.949
16S	28	.359	.940	.307	.925
19E	30	.301	.928	.362	.944
27E	21	.164	.804	.237	.867
31C	31	.202	.887	.262	.917
51B	37	.136	.853	.157	.873
54E	31	.147	.842	.160	.855
55B	33	.247	.915	.188	.884
63B	33	.270	.924	.261	.921
64C	29	.208	.884	.123	.803
67N	37	.315	.945	.302	.941
71L	31	.205	.889	.207	.890
76W	14	.027	.283	.096	.597
76Y	29	.233	.898	.173	.858
91A	36	.247	.922	.295	.938
94B	28	.187	.865	.191	.869
95B	31	.194	.882	.256	.914
96B	14	.242	.818	.234	.801
Average		.223	.863	.236	.881

In general, the weights assigned the constructs by the NCOs correlated higher with those assigned by the officers when the conjoint scaling method was used (Table 11). Across the 20 MOS, the correlations between the NCO/officer mean conjoint weights ranged from .100 to .968 with an average of .705. The corresponding range for the direct estimation weights was -.379 to .958, with an average of .617.

The slight overall psychometric superiority of the conjoint weights may be due in part to the larger discriminability of the weights obtained from the conjoint method. The average standard deviation across all judges of the weights assigned by the conjoint method was 9.20; the corresponding average was 5.46 for the direct estimation method (Table 12).

Considering the above findings, the decision was made to favor the weights derived from the conjoint scaling method in combining the individual construct scores into an overall composite measure of performance.

Comparison of Mean Conjoint Weights by Construct, Grade, and MOS

The mean weights obtained through the conjoint scaling method are shown in Table 14 broken down by construct, grade, and MOS. It should be borne in mind that the weights are based on comparative judgments of the constructs within each MOS and should not be used for comparisons of importance across MOS. It is, however, interesting to note whether the relative pattern of weights differs across MOS and whether some constructs are fairly consistently given relatively higher weights than others.

To explore differences in the relative pattern of weights, an analysis of variance of the conjoint weights was conducted to test for mean construct differences and for any significant interactions with grade and MOS. The analysis also compared the mean weights assigned by judges drawn from MOS proponent posts with those of judges from USAREUR and FORSCOM posts. The means for grade (officer vs. NCO), type of unit (field vs. proponent), and MOS main effect were set at 20.0 by the scaling method and hence were not a source of variation. Table 15 shows the results of the overall analysis of variance. The construct means were significantly different. The interactions of constructs with grade and MOS were also highly significant, indicating that the relative weights were different for officers in comparison with NCOs and were also different across MOS. Finally, there was a significant three-way interaction among the constructs, MOS, and type of judge (field vs. proponent post).

Examination of the construct means in Table 14 shows that for all 20 MOS, military bearing/physical fitness received the lowest relative weight. In 13 of the 20 MOS, core technical skills received the highest relative weight, while the exercising leadership construct was second overall. The leadership component received the highest relative weight in 6 of the 20 MOS. For the most part, the MOS skills construct received the highest weight for the technical MOS in the sample and the exercising leadership construct received the highest weight for the combat MOS (the job of Armor Crewman is a notable exception). The general skills construct received the highest weight for only one MOS, Military Police (95B). These MOS

Table 14

Mean Construct Weights by Grade, and MOS: Conjoint Method

MOS	MOS Skills			General Skills			Ex. Leadership			Main. Discipline			Military Bearing		
	NCO	Off	Total	NCO	Off	Total	NCO	Off	Total	NCO	Off	Total	NCO	Off	Total
11B	19.5	22.9	22.0	18.7	18.5	18.5	21.7	29.1	27.3	22.2	17.2	18.4	18.0	12.3	13.7
12B	20.2	18.4	18.8	22.8	19.7	20.4	21.1	30.2	28.1	23.8	20.3	21.1	12.1	11.5	11.6
13B	28.6	22.7	24.1	16.2	19.2	18.5	21.1	27.7	26.2	20.7	18.3	18.9	13.3	12.1	12.4
16S	20.8	25.9	24.6	16.8	16.3	16.4	23.1	26.3	25.5	24.8	20.2	21.3	14.5	11.4	12.1
19E	30.0	29.4	29.6	20.2	21.1	20.8	16.5	20.5	19.3	23.3	17.9	19.5	10.0	11.0	10.7
27E	21.7	24.2	23.5	20.9	18.0	18.8	21.9	22.4	22.3	21.0	23.0	22.4	14.5	12.4	13.0
31C	29.4	29.0	29.1	20.0	20.3	20.3	16.7	22.0	20.8	19.0	17.3	17.6	14.9	11.4	12.2
51B	19.4	25.6	23.9	20.2	17.2	18.0	22.7	25.6	24.8	20.2	19.7	19.8	17.5	11.9	13.4
54E	28.6	25.4	26.5	16.7	21.5	19.8	17.3	20.7	19.5	20.6	19.8	20.1	16.8	12.6	14.1
55B	32.4	22.4	24.9	18.9	19.5	19.4	14.8	27.8	24.7	16.6	19.5	18.8	17.3	10.8	12.3
63B	21.4	27.5	25.6	21.4	18.1	19.1	21.8	23.5	23.0	19.4	21.1	20.5	16.1	9.9	11.8
64C	21.8	26.1	24.8	16.9	22.8	20.9	23.2	21.8	22.2	20.9	15.4	17.1	17.2	14.0	15.0
67N	28.7	25.9	26.7	15.8	15.9	15.9	23.7	25.3	24.9	20.7	22.2	21.8	11.1	10.6	10.8
71L	20.9	24.1	23.1	20.4	19.9	20.1	21.2	22.7	22.3	20.7	21.0	20.9	16.8	12.3	13.7
76W	29.4	23.6	25.7	19.1	17.2	17.9	17.3	25.0	22.2	18.5	22.9	21.3	15.7	11.4	12.9
76Y	26.3	25.7	25.8	23.3	21.7	22.1	18.7	19.8	19.5	15.8	17.5	17.1	15.8	15.3	15.4
91A	28.2	26.9	27.3	18.8	16.6	17.3	20.0	23.1	22.1	21.3	22.5	22.1	11.7	11.0	11.2
94B	19.9	24.5	23.0	15.0	17.4	16.6	26.7	26.2	26.4	23.1	20.9	21.6	15.3	11.0	12.4
95B	17.5	20.0	19.2	23.5	27.8	26.5	26.5	20.5	22.4	22.2	19.1	20.1	10.3	12.6	11.8
96B	18.9	28.7	25.2	22.2	18.2	19.6	19.0	23.3	21.8	25.2	19.8	21.7	14.7	10.0	11.7
AVG	24.2	24.9	24.7	19.4	19.3	19.3	20.8	24.2	23.3	21.0	19.8	20.1	14.7	11.8	12.6

Table 15

Results of Overall Analysis of Variance of Conjoint Weights

Source of Variation	df	Sum of Squares	Mean Square	F Value	PR>F
Between Subjects[a]					
Grade	1	0.0			
MOS	19	0.0			
Type of Unit	1	0.0			
Grade X MOS	19	0.0			
Grade X Type	1	0.0			
MOS X Type	14	0.0			
Grade X MOS X Type	13	0.0			
Error	523	0.0			
Within Subjects					
Constructs	4	52604.8	13151.2	161.9	.0000
Constructs X Grade	4	2694.3	673.6	8.3	.0001
Constructs X MOS	76	14133.1	186.0	2.3	.0001
Constructs X Type	4	432.3	108.1	1.3	.2562
Constructs X MOS X Grade	76	6930.5	91.2	1.1	.2227
Constructs X Grade X Type	4	60.1	15.0	.2	.9464
Constructs X MOS X Type	56	6373.9	113.8	1.4	.0276
C X G X MOS X T	52	3276.6	63.0	.8	.8781
Error	2092	169947.7			

[a]The between-subjects sum of squares are equal to zero since the weights for all subjects summed to 100.

differences in the constructs receiving the highest weights undoubtedly contributed to the significant Construct x MOS interaction.

Significant mean differences between the weights assigned by officers and NCOs were found for two constructs: Officers gave significantly higher relative weights to the exercising leadership construct than did NCOs, while NCOs gave higher weights to the military bearing/physical fitness construct than did officers. The NCOs may have been giving relatively more weight to aspects of first-tour soldiers' performance that were of more immediate concern to them. Although the mean differences were only significantly different at the .10 level, the NCOs gave the personal discipline construct weights that were higher on the average than those assigned by the officers.

The Impact of Scenario on Relative Construct Weights

Toward the end of the data collection, a field experiment was run to determine whether a change in scenario would affect the weighting judgments. Using the direct estimation scaling method, officers and NCOs in 13 MOS judged the relative weight of the five performance constructs under a wartime and a peacetime scenario, after they had completed judging the constructs under the heightened tension scenario using both the direct estimation and conjoint methods. The two additional scenarios were as described in Figure 2.

An analysis of variance was conducted on the data from 139 officers and 37 NCOs who judged the five constructs under all three scenarios. Of particular interest was whether the within-subject Scenario x Construct interaction term was significant, since that would indicate whether the judges changed the relative weights assigned one or more constructs as a function of the scenario.

The Scenario x Construct interaction was significant, and separate analyses of variance were conducted for each construct to help determine which construct weights were influenced the most by the different scenarios. These analyses indicated that the means of the MOS skills, general skills, and military bearing/physical fitness construct weights were significantly different across the scenarios (see Table 16). The military bearing/ physical fitness construct received relatively more weight under the peacetime scenario than it did under the heightened tensions and wartime scenarios. The general skills construct, on the other hand, received relatively more weight under the wartime scenario than under the heightened tensions and peacetime scenarios, while the MOS skills construct received its highest weights under the heightened tensions scenario.

Although these scenario differences were statistically significant, the actual mean differences were quite small and the rank ordering of the five components did not change across scenarios. Also, the correlations between the weights assigned under the three scenarios averaged about .85 across the 13 MOS. With weights correlated that highly, overall performance composites obtained through applying the separate sets of scenario weights to construct scores would most likely correlate between .95 and .99. As a consequence, we can predict with certainty that alternative criterion composites based on different scenario weights will <u>not</u> yield different predictor equations.

It is interesting to note that there was more discriminability in the weights assigned the constructs within MOS under the heightened tension scenario than under the peacetime and wartime scenarios. When the standard deviations of the mean (for n-judges) construct weights for each MOS were averaged across MOS, the means were 5.33, 4.76, and 4.80 respectively and these mean differences were significantly different at the .001 level. The reliabilities of the weights assigned under the heightened tension scenario were also higher. Across the 13 MOS the average 1-rater reliability for the heightened tension scenario was .224. The corresponding average reliabilities for the peacetime and wartime scenarios were .137 and .202.

Table 16

Mean Weights of Constructs by Scenario (Based on Data From 13 MOS)

Construct	Heightened Peacetime	Tensions	Wartime
MOS Skills[a]	21.6	22.3	21.7
General Skills[a]	19.9	20.4	21.3
Exercise Leadership	21.4	21.8	21.5
Personal Discipline	19.9	19.6	19.9
Military Bearing[a]	17.1	15.8	15.7

[a]Construct means significantly different across scenarios at .05 level.

DISCUSSION AND CONCLUSIONS

The five Project A performance constructs received significantly different patterns of weights in different MOS (e.g., see Table 14) and the different groups of experts agreed, in general, on the relative ranking of the weights. For example, the exercising peer leadership and effort construct tends to be rated highest among the combat MOS.

Multiple judges per MOS, about 30 on the average, produced n-rater reliabilities that are quite respectable (above .85 for most MOS). The high intermethod correlations (about .95 on the average) between the construct weights obtained by the direct estimation and conjoint methods for the separate MOS further document the reliability of the means of the scaled weights.

That different groups of judges may provide somewhat different MOS weights can be seen in the correlations between the officer and NCO weights, of .617 and .705 for the direct estimate and conjoint methods, respectively. The NCOs tended to give relatively higher weights to the military bearing/physical fitness construct, while the officers attached more importance to the leadership/effort construct. The NCOs could have been reacting more to the every-day problems of handling first-tour soldiers, while the officers could have been more concerned with performance characteristics required most under near or actual wartime conditions. The pattern of results obtained when the weights were evaluated under wartime and peacetime scenarios in part supports this hypothesis.

Though there were statistically significant differences in the mean weights assigned under the three scenarios, the very small differences will have little impact on the relative ranking of soldiers on the overall performance composites for an MOS. A more critical question is how much impact will the weights themselves have on recommended job assignments in an optimal selection and classification system? Would the same assignment recommendations be made were all weighted equally? Would a different set of predictors be selected using a weighted composite for validation than would have been selected if the constructs had been weighted equally?

The answers to these questions obviously depend not only on the set of weights used but on such factors as the intercorrelations among the construct performance scores, the validity of the predictor battery, the amount of differential prediction it affords across Army jobs, the MOS selection standards in effect, and the assignment algorithms employed.

Of particular interest is the difference between what might be called nominal and operative weights. Nominal weighting refers to multiplying the raw score on each component by the SME derived weight for the component and adding across components to get the total criterion composite score. However, a component's operative weight for determining the overall ranking of people on the total composite is also a function of its variance and its covariances with the other components. Components with higher variances carry more weight and differential weights have less differential effect as covariances become higher. The alternatives to cross-multiplying SME weights with raw component scores are to: a) standardize the component scores to control for variance differences, b) "assign" the total composite variance (which is the sum of all component variances and covariances) to components by adding a particular component's variance to its covariances with each of the other components; and choosing weights for the components which will make their proportion of the total variance equal to the SME-determined weight; or c) reconstitute the component scores as orthogonal vectors and assign weights to these variables. The most straightforward method would be to apply the SME weights to standardized component scores and to let the reality of the intercorrelations among the components have their influence. However, the most informative way to address these, and the other issues discussed above, is through a series of sensitivity analyses that portray the effects of these parameters on selection and classification and validity.

To the extent that the differential weights described here enhance the overall Army selection and classification process, the time and effort that have gone into developing them will be more than worthwhile. However, even if the weights' effect on the selection and classification process proves minimal, we will have developed defensible performance composites for the Project A sample MOS to use as overall criterion measures in validating the ASVAB and other selection instruments and procedures.

REFERENCES

Alpert, M. I., Betak, J. F., & Golden, L. L. (1978). Gathering issues in conjoint measurement. Working paper, Graduate School of Business, Stanford University.

Bernardin, H. J., & Beatty, R. W. (1984). Performance appraisal: Assessing human behavior at work. Belmont, Calif.: Kent Publishing Company.

Campbell, J. P. (Ed.). (1985). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year (ARI Technical Report 746). (AD A193 343)

Campbell, J. P. (Ed.). (1986a). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year (ARI Technical Report 792).

Campbell, J. P. (Ed.). (1986b, August). Project A: When the textbook goes operational. Paper presented at the Annual Convention of the American Psychological Association, Washington, DC.

Campbell, J. P. (Ed.). (1987). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1987 fiscal year (ARI Technical Report in preparation).

Eaton, N. K., & Goer, M. H. (Eds.). (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Technical appendix to the annual report (ARI Research Note 83-37). (AD A137 117)

Eaton, N. K., Goer, M. H., Harris, J. H., & Zook, L. M. (Eds.). (1985). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 fiscal year (ARI Technical Report 660). (AD A178 944)

Green, P. E. (1974). On the design of choice experiments involving multifactor alternatives. Journal of Consumer Research, 1, 61-73.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. Journal of Consumer Research, 5, 103-123.

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report (ARI Research Report 1347) (AD A141 807)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1985). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report synopsis, 1984 fiscal year (ARI Research Report 1393). (AD A173 824)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1985). Improving the selection, classification, and utilization of Army enlisted personnel: Appendices to annual report, 1984 fiscal year (ARI Research Note 85-14). (AD A170 475)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1987). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year--Supplement to ARI Technical Report 746 (ARI Research Note 87-54). (AD A188 267)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1986). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year--Supplement to ARI Technical Report 792 (ARI Research Note 88-36). (AD A196 274)

Jain, A. K., Acito, F. M., Malhotra, N., & Mahajan, V. (1978). A comparison of predictive validity of alternative methods for estimating parameters in preference models. Working paper, School of Management, State University of New York at Buffalo.

Johnson, R. M. (1974). Trade-off analysis of consumer values. Journal of Marketing Research, 11, 121-127.

Johnson, R. M., & VanDyk, G. J. (1975). A resistance analogy for efficiency of paired comparisons designs. Unpublished paper, Market Facts, Inc., Chicago.

Kane, J. S. (1980). Performance distribution assessment: A new framework for conceiving and appraising job performance. Unpublished paper, Kane and Associates.

Montgomery, D. B., Wittink, D. R., & Glaze, T. (1977). A predictive test of individual level concept evaluation and trade-off analysis (Research Paper No. 415). Graduate School of Business, Stanford University.

Oppedijk van Veen, W. M., & Beazley, D. (1977). An investigation of alternative methods of applying the trade-off model. Journal of Market Research Society, 19, 2-9.

Ross, R. T. (1934). Optimum orders for presentations of pairs in pair comparison. Journal of Educational Psychology, 25, 375-382.

Sadacca, R., deVera, M. V., DiFazio, A., & White, L. A. (1986). Weighting performance constructs in composite measures of job performance. Paper presented at the convention of the American Psychological Association, Washington, DC.

Schmidt, F. L. (1977). The measurement of job performance. Unpublished manuscript, U.S. Office of Personnel Management.

Torgerson, W. S. (1958). Theory and methods of scaling. New York: John Wiley & Sons.

Wise, L. L., Campbell, J. P., McHenry, J. J., & Hanser, L. M. (1986, August). A latent structure model of job performance factors. Paper presented at the convention of the American Psychological Association, Washington, DC.

Appendix A

INSTRUCTIONS AND MATERIALS USED TO COLLECT WEIGHTING JUDGMENTS FOR INFANTRYMEN (MOS 11B)

General Instructions

JUDGING THE IMPORTANCE OF PERFORMANCE FACTORS IN ARRIVING AT TOTAL SCORES

Background

A number of different kinds of performance factors are being considered by Project A to assess the effectiveness of first-tour enlisted personnel. These various performance factors must be combined into one overall measure of MOS performance. This overall measure should be the best that can be obtained given the available component performance factors. The overall measure will be used as the performance measure against which the ASVAB and other predictor performance factors will be validated. To obtain the best overall measure for each MOS in our sample, Project A staff will be asking knowledgeable officers and NCOs to assign weights to the various performance factors in a manner that reflects the relative importance that the component performance factors should have in forming an overall measure for the MOS.

Today we would like to get your judgments about the relative weights that the factors should receive in deriving an overall performance measure for first-tour Infantryman (11B). The performance factors are:

Task Proficiency: MOS specific technical skills -- This performance factor represents the proficiency with which the soldier performs the tasks which are "central" to MOS 11B. The tasks represent the core of the job and they are the primary definers of the MOS. For example, the first tour Infantryman engages enemy target with hand grenades; installs and fires/recovers an M18A1 claymore mine; selects hasty firing positions in urban terrain; zeros an AN/PVS-4 to an M16A1 rifle; and uses weapons and other equipment in offensive and defensive combat operations.

This performance factor does not include the individual's willingness to perform the task or the degree to which the individual can coordinate his efforts with others. It refers to how well the individual can execute the core technical tasks the job requires, given a willingness to do so.

Task Proficiency: General soldiering skills -- In addition to the core technical content specific to an MOS, individuals in every MOS are also responsible for being able to perform a variety of general soldiering tasks--for example, determines grid coordinates on military maps, puts on, wears and removes M17 series protective mask with hood, determines a magnetic azimuth using a compass, collects/reports information - SALUTE, recognizes and identifies friendly and threat aircraft. Performance on this factor represents overall proficiency on these general soldiering tasks. Again, it refers to how well the individual can execute general soldiering tasks, given a willingness to do so.

Exercise of Leadership, Effort, and Self Development--This performance factor reflects the degree to which the individual exerts effort over the full range of job tasks perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers. That is, can the individual be counted on to carry out assigned tasks, even under adverse conditions, to exercise good judgment, and to be generally dependable and proficient. While appropriate knowledges and skills are necessary for successful performance, this factor is only meant to reflect the individual's willingness to do the job required and to be cooperative and supportive with other soldiers.

Maintaining Personal Discipline--This performance factor reflects the degree to which the individual adheres to Army regulations and traditions, exercises personal self control, demonstrates integrity in day to day behavior, and does not create disciplinary problems. People who rank high on this factor show a commitment to high standards of personal conduct.

Military Bearing/Appearance and Physical Fitness--This performance factor represents the degree to which the individual maintains an appropriate military appearance and bearing and stays in good physical condition.

Please assume that a total score will be derived for each soldier from the separate scores obtained from each of these factors. These total scores will be our best estimate of the overall effectiveness of the troops whose performance will be measured. We need the assistance of experienced Army personnel in determining how much weight should be given each factor in arriving at the total effectiveness scores.

Purpose

The purpose of this workshop is to obtain the weights to be assigned each of the performance factors. Two methods of assigning weights will be used. The methods differ in the kinds of judgments you will be required to make:

Method A: You will be asked to rank order the performance factors and then assign weights to them, assuming that the top ranked factor has a weight of 100.

Method B: You will be given performance profiles on 10 sets of 15 soldiers each and asked to rank order them. (The profiles will give the scores of the soldiers on two of the five performance factors at a time.)

Assumptions for Both Methods

(1) The type of soldiers for whom performance factor weights are being derived is first tour Infantryman (11B).

(2) As the weights you assign may be a function of the particular context in which the soldiers' performance is being evaluated, please assume the following military situation prevails:

> The world is in a period of heightened tensions. There is an increasing probability that hostilities will break out in Europe, Asia, the Caribbean, Latin America, and Africa. The Army's mission is to support U.S. treaty obligations and to help defend the borders of allied and friendly nations. Some of the potential enemies have nuclear and chemical capability. Air parity does exist between allied forces and potential hostile nations. U.S. Army training and other preparatory activities have been substantially increased. Most combat and associated support units are participating in frequent field exercises. Most units are being actively resupplied.

(3) Performance factor scores are available <u>only</u> on the factors given. Although there may be other factors that comprise overall performance, no scores are available for them at this time.

Materials for Method A

DIRECTIONS FOR METHOD A

Under this weighting method, the procedure for assigning weights to the performance factors is as follows:

1. Rank order the set of performance factors to be weighted by assigning a "1" to the most important, a "2" to the next most important, etc. Please refer to the "PERFORMANCE FACTORS FOR MOS 11B" handout for a complete description of the 5 performance factors.

2. After you have recorded the rank orders on the weighting sheet, assign 100 points to the factor you ranked as most important. Then ask yourself, "If I'm assigning 100 points to this performance factor, how many points should I assign to the next most important one." If, for example, you think that the second most important one should receive half the weight of the first, assign it 50 points. Continue assigning points in this manner until all the factors have been weighted.

3. In assigning the points, please keep in mind that the points represent how many times more (or less) important one performance factor is than another. For example, if you assign 30 points to one factor and 5 points to another, that means that you believe that the 30-point factor should receive 6 times the weight in the total score as the 5-point factor.

4. If you feel that two or more factors should be weighted equally, you may assign them equal weights. For example, if you feel that the factors ranked first and second are really tied in importance, then you can assign them both 100 points.

5. If you believe that a particular performance factor should not be used at all in arriving at the total score, you should assign it zero points.

6. When you are finished assigning points to all performance factors, please make sure that they are in the "right" ratio to one another. That is, the points assigned to all factors are in correct proportion to one another.

   Thank you for your cooperation.

Name ______________________ Workshop __________

MOS 11B Performance Factor Weighting Sheet

Performance Factor*	Rank Order	Weight
1. Task proficiency -- MOS specific technical skills.	____	____
2. Task proficiency -- general soldiering skills.	____	____
3. Exercise of leadership, effort, and self development.	____	____
4. Maintaining personal discipline.	____	____
5. Military bearing/appearance and physical fitness.	____	____

* Please refer to the "PERFORMANCE FACTORS FOR MOS 11B" handout for a complete description of the 5 performance factors.

PERFORMANCE FACTORS FOR MOS 11B

1) Task Proficiency: MOS specific technical skills

This performance factor represents the proficiency with which the soldier performs the tasks which are "central" to MOS 11B. The tasks represent the core of the job and they are the primary definers of the MOS. For example, the first tour Infantryman engages enemy target with hand grenades; installs and fires/recovers an M18A1 claymore mine; selects hasty firing positions in urban terrain; zeros an AN/PVS-4 to an M16A1 rifle; and uses weapons and other equipment in offensive and defensive combat operations.

This performance factor does not include the individual's willingness to perform the task or the degree to which the individual can coordinate his efforts with others. It refers to how well the individual can execute the core technical tasks the job requires, given a willingness to do so.

2) Task Proficiency: General soldiering skills

In addition to the core technical content specific to an MOS, individuals in every MOS are also responsible for being able to perform a variety of general soldiering tasks--for example, determines grid coordinates on military maps, puts on, wears and removes M17 series protective mask with hood, determines a magnetic azimuth using a compass, collects/reports information - SALUTE, recognizes and identifies friendly and threat aircraft. Performance on this factor represents overall proficiency on these general soldiering tasks. Again, it refers to how well the individual can execute general soldiering tasks, given a willingness to do so.

3) Exercise of Leadership, Effort, and Self Development

This performance factor reflects the degree to which the individual exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers. That is, can the individual be counted on to carry out assigned tasks, even under adverse conditions, to exercise good judgment, and to be generally dependable and proficient. While appropriate knowledges and skills are necessary for successful performance, this factor is only meant to reflect the individual's willingness to do the job required and to be cooperative and supportive with other soldiers.

4) Maintaining Personal Discipline

This performance factor reflects the degree to which the individual adheres to Army regulations and traditions, exercises personal self control, demonstrates integrity in day-to-day behavior, and does not create disciplinary problems. People who rank high on this factor show a commitment to high standards of personal conduct.

5) Military Rearing/Appearance and Physical Fitness

This performance factor represents the degree to which the individual maintains an appropriate military appearance and bearing and stays in good physical condition.

Materials for Method B

DIRECTIONS FOR METHOD B

Under this method, judgments of the overall performance scores for 10 sets of Infantrymen will be obtained. Each set will contain 15 Infantrymen. The performance scores of each of the 15 first tour Infantrymen have been recorded on 2 performance factor scales. (A different pair of performance factor scales are provided for each of the 10 sets). For each scale there is a description of high, medium and low levels of performance. Each of the 15 soldiers is rated on a 7-point scale that ranges from the lowest level of performance to the highest. Please refer to the "PERFORMANCE FACTORS FOR MOS 11B" handout for a complete description of the 5 performance factors. Also, please review the assumptions given in the General Instructions.

Specific Instructions

1. Rank the 15 Infantrymen in the first set in order of their overall performance. Give the "best" soldier a rank of "1", the second best soldier a rank of "2" and so on. Make comparisons between the soldiers on the basis of their overall performance as Infantrymen; do not consider how they might be used in other capacities.

2. When you are finished, please go over the rank order carefully making sure that, in your judgment, the ranks reflect the relative overall performance of the soldiers. Feel free to change any ranks.

3. When satisfied with your rank ordering, proceed to the next set of 15 Infantrymen.

   Thank you for your cooperation.

   (A sample sheet for one pair of performance factor scales follows.)

Name______________________ Sheet No. 1

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Task Proficiency--MOS Specific Technical Skills	Task Proficiency--General Soldiering Skills	Rank Order
1	6	2	___
2	5	5	___
3	2	6	___
4	5	3	___
5	2	3	___
6	6	5	___
7	4	7	___
8	1	4	___
9	5	6	___
10	7	4	___
11	3	5	___
12	3	3	___
13	4	4	___
14	4	1	___
15	3	2	___

Performance Scales:

TASK PROFICIENCY--MOS SPECIFIC TECHNICAL SKILLS

Does not display the knowledge/skill required to perform many core technical skills.	Displays the knowledge/skill required to perform most core technical tasks properly, but may need help for harder tasks.	Displays the knowledge/skill to perform all core technical tasks properly.
1 2	3 4 5	6 7

TASK PROFICIENCY--GENERAL SOLDIERING SKILLS

Does not display the knowledge/skill required to perform many general soldiering tasks.	Displays the knowledge/skill required to perform most general soldiering tasks, but may need help for harder tasks.	Displays the knowledge/skill to perform all general soldiering skills.
1 2	3 4 5	6 7

Name ______________________ Sheet No. 2

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Military Bearing/ Appearance and Physical Fitness	Exercise of Leadership, Effort and Self Development	Rank Order
1	6	2	___
2	5	5	___
3	2	6	___
4	5	3	___
5	2	3	___
6	6	5	___
7	4	7	___
8	1	4	___
9	5	6	___
10	7	4	___
11	3	5	___
12	3	3	___
13	4	4	___
14	4	1	___
15	3	2	___

Performance Scales:

MILITARY BEARING/APPEARANCE AND PHYSICAL FITNESS

Maintains self in poor physical condition. Fails to meet military standards for dress or personal hygiene.		Meets Army standards of physical fitness. Dresses neatly and meets Army standards of personal hygiene.			Exceeds Army standards and expectations set for physical fitness. Maintains excellent personal hygiene and proper appearance.	
1	2	3	4	5	6	7

EXERCISE OF LEADERSHIP, EFFORT AND SELF DEVELOPMENT

Fails to take charge when leadership is required in unit. Provides little or no assistance to other unit members. Seldom exerts effort in accomplishing many job assignments and tasks. Gives up easily under adverse conditions.		Performs satisfactorily in leadership situations where what is expected is well known. When asked, gives help and support to fellow soldiers. Usually exerts effort to perform most job assignments and tasks.			Takes charge when necessary to lead unit; leads the squad to outstanding performance. Does everything possible to assist other soldiers. Always exerts considerable effort in performing all job assignments and tasks.	
1	2	3	4	5	6	7

Name____________________ Sheet No. 3

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Maintaining Personal Discipline	Task Proficiency--MOS Specific Technical Skills	Rank Order
1	6	2	____
2	5	5	____
3	2	6	____
4	5	3	____
5	2	3	____
6	6	5	____
7	4	7	____
8	1	4	____
9	5	6	____
10	7	4	____
11	3	5	____
12	3	3	____
13	4	4	____
14	4	1	____
15	3	2	____

Performance Scales:

MAINTAINING PERSONAL DISCIPLINE

Occasionally shows disrespect towards superiors; often fails to follow Army/unit rules, regulations or orders. Creates disciplinary problems.	Rarely exhibits disrespectful behavior towards superiors. Almost always follows Army/unit rules, regulations or orders.	Always treats superiors with respect. Maintains high level of personal integrity. Obeys orders quickly and with enthusiasm.
1 2	3 4 5	6 7

TASK PROFICIENCY--MOS SPECIFIC TECHNICAL SKILLS

Does not display the knowledge/skill required to perform many core technical skills.	Displays the knowledge/skill required to perform most core technical tasks properly, but may need help for harder tasks.	Displays the knowledge/skill to perform all core technical tasks properly.
1 2	3 4 5	6 7

Name ______________________________ Sheet No. 4

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Exercise of Leadership, Effort, and Self Development	Task Proficiency--General Soldiering Skills	Rank Order
1	6	2	____
2	5	5	____
3	2	6	____
4	5	3	____
5	2	3	____
6	6	5	____
7	4	7	____
8	1	4	____
9	5	6	____
10	7	4	____
11	3	5	____
12	3	3	____
13	4	4	____
14	4	1	____
15	3	2	____

Performance Scales:

EXERCISE OF LEADERSHIP, EFFORT AND SELF DEVELOPMENT

Fails to take charge when leadership is required in unit. Provides little or no assistance to other unit members. Seldom exerts effort in accomplishing many job assignments and tasks. Gives up easily under adverse conditions.	Performs satisfactorily in leadership situations where what is expected is well known. When asked, gives help and support to fellow soldiers. Usually exerts effort to perform most job assignments and tasks.	Takes charge when necessary to lead unit; leads the squad to outstanding performance. Does everything possible to assist other soldiers. Always exerts considerable effort in performing all job assignments and tasks.
1 2	3 4 5	6 7

TASK PROFICIENCY--GENERAL SOLDIERING SKILLS

Does not display the knowledge/skill required to perform many general soldiering tasks.	Displays the knowledge/skill required to perform most general soldiering tasks, but may need help for harder tasks.	Displays the knowledge/skill to perform all general soldiering skills.
1 2	3 4 5	6 7

Name ______________________________ Sheet No. 5

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Maintaining Personal Discipline	Military Bearing/ Appearance and Physical Fitness	Rank Order
1	6	2	____
2	5	5	____
3	2	6	____
4	5	3	____
5	2	3	____
6	6	5	____
7	4	7	____
8	1	4	____
9	5	6	____
10	7	4	____
11	3	5	____
12	3	3	____
13	4	4	____
14	4	1	____
15	3	2	____

Performance Scales:

MAINTAINING PERSONAL DISCIPLINE

Occasionally shows disrespect towards superiors; often fails to follow Army/unit rules, regulations or orders. Creates disciplinary problems.	Rarely exhibits disrespectful behavior towards superiors. Almost always follows Army/unit rules, regulations or orders.	Always treats superiors with respect. Maintains high level of personal integrity. Obeys orders quickly and with enthusiasm.
1 2	3 4 5	6 7

MILITARY BEARING/APPEARANCE AND PHYSICAL FITNESS

Maintains self in poor physical condition. Fails to meet military standards for dress or personal hygiene.	Meets Army standards of physical fitness. Dresses neatly and meets Army standards of personal hygiene.	Exceeds Army standards and expectations set for physical fitness. Maintains excellent personal hygiene and proper appearance.
1 2	3 4 5	6 7

Name______________________ Sheet No. 6

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Task Proficiency--MOS Specific Technical Skills	Exercise of Leadership, Effort, and Self Development	Rank Order
1	6	2	____
2	5	5	____
3	2	6	____
4	5	3	____
5	2	3	____
6	6	5	____
7	4	7	____
8	1	4	____
9	5	6	____
10	7	4	____
11	3	5	____
12	3	3	____
13	4	4	____
14	4	1	____
15	3	2	____

Performance Scales:

TASK PROFICIENCY--MOS SPECIFIC TECHNICAL SKILLS

Does not display the knowledge/skill required to perform many core technical skills.	Displays the knowledge/skill required to perform most core technical tasks properly, but may need help for harder tasks.	Displays the knowledge/skill to perform all core technical tasks properly.
1 2	3 4 5	6 7

EXERCISE OF LEADERSHIP, EFFORT AND SELF DEVELOPMENT

Fails to take charge when leadership is required in unit. Provides little or no assistance to other unit members. Seldom exerts effort in accomplishing many job assignments and tasks. Gives up easily under adverse conditions.	Performs satisfactorily in leadership situations where what is expected is well known. When asked, gives help and support to fellow soldiers. Usually exerts effort to perform most job assignments and tasks.	Takes charge when necessary to lead unit; leads the squad to outstanding performance. Does everything possible to assist other soldiers. Always exerts considerable effort in performing all job assignments and tasks.
1 2	3 4 5	6 7

Name ____________________ Sheet No. 7

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Task Proficiency--General Soldiering Skills	Maintaining Personal Discipline	Rank Order
1	6	2	___
2	5	5	___
3	2	6	___
4	5	3	___
5	2	3	___
6	6	5	___
7	4	7	___
8	1	4	___
9	5	6	___
10	7	4	___
11	3	5	___
12	3	3	___
13	4	4	___
14	4	1	___
15	3	2	___

Performance Scales:

TASK PROFICIENCY--GENERAL SOLDIERING SKILLS

Does not display the knowledge/skill required to perform many general soldiering tasks.		Displays the knowledge/skill required to perform most general soldiering tasks, but may need help for harder tasks.			Displays the knowledge/skill to perform all general soldiering skills.	
1	2	3	4	5	6	7

MAINTAINING PERSONAL DISCIPLINE

Occasionally shows disrespect towards superiors; often fails to follow Army/unit rules, regulations or orders. Creates disciplinary problems.		Rarely exhibits disrespectful behavior towards superiors. Almost always follows Army/unit rules, regulations or orders.			Always treats superiors with respect. Maintains high level of personal integrity. Obeys orders quickly and with enthusiasm.	
1	2	3	4	5	6	7

Name________________________ Sheet No. 8

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Soldier No.	Performance Level: Military Bearing/ Appearance and Physical Fitness	Performance Level: Task Proficiency-- MOS Specific Technical Skills	Rank Order
1	6	2	____
2	5	5	____
3	2	6	____
4	5	3	____
5	2	3	____
6	6	5	____
7	4	7	____
8	1	4	____
9	5	6	____
10	7	4	____
11	3	5	____
12	3	3	____
13	4	4	____
14	4	1	____
15	3	2	____

Performance Scales:

MILITARY BEARING/APPEARANCE AND PHYSICAL FITNESS

Maintains self in poor physical condition. Fails to meet military standards for dress or personal hygiene.		Meets Army standards of physical fitness. Dresses neatly and meets Army standards of personal hygiene.			Exceeds Army standards and expectations set for physical fitness. Maintains excellent personal hygiene and proper appearance.	
1	2	3	4	5	6	7

TASK PROFICIENCY--MOS SPECIFIC TECHNICAL SKILLS

Does not display the knowledge/skill required to perform many core technical skills.		Displays the knowledge/ skill required to perform most core technical tasks properly, but may need help for harder tasks.			Displays the knowledge/ skill to perform all core technical tasks properly.	
1	2	3	4	5	6	7

Name______________________________ Sheet No. 9

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Exercise of Leadership, Effort, and Self Development	Maintaining Personal Discipline	Rank Order
1	6	2	____
2	5	5	____
3	2	6	____
4	5	3	____
5	2	3	____
6	6	5	____
7	4	7	____
8	1	4	____
9	5	6	____
10	7	4	____
11	3	5	____
12	3	3	____
13	4	4	____
14	4	1	____
15	3	2	____

Performance Scales:

EXERCISE OF LEADERSHIP, EFFORT AND SELF DEVELOPMENT

Fails to take charge when leadership is required in unit. Provides little or no assistance to other unit members. Seldom exerts effort in accomplishing many job assignments and tasks. Gives up easily under adverse conditions.	Performs satisfactorily in leadership situations where what is expected is well known. When asked, gives help and support to fellow soldiers. Usually exerts effort to perform most job assignments and tasks.	Takes charge when necessary to lead unit; leads the squad to outstanding performance. Does everything possible to assist other soldiers. Always exerts considerable effort in performing all job assignments and tasks.
1 2	3 4 5	6 7

MAINTAINING PERSONAL DISCIPLINE

Occasionally shows disrespect towards superiors; often fails to follow Army/unit rules, regulations or orders. Creates disciplinary problems.	Rarely exhibits disrespectful behavior towards superiors. Almost always follows Army/unit rules, regulations or orders.	Always treats superiors with respect. Maintains high level of personal integrity. Obeys orders quickly and with enthusiasm.
1 2	3 4 5	6 7

Name________________________ Sheet No. 10

MOS 11B OVERALL PERFORMANCE SCORE SHEET

Performance Level

Soldier No.	Task Proficiency--General Soldiering Skills	Military Bearing/Appearance and Physical Fitness	Rank Order
1	6	2	____
2	5	5	____
3	2	6	____
4	5	3	____
5	2	3	____
6	6	5	____
7	4	7	____
8	1	4	____
9	5	6	____
10	7	4	____
11	3	5	____
12	3	3	____
13	4	4	____
14	4	1	____
15	3	2	____

Performance Scales:

TASK PROFICIENCY--GENERAL SOLDIERING SKILLS

Does not display the knowledge/skill required to perform many general soldiering tasks.	Displays the knowledge/skill required to perform most general soldiering tasks, but may need help for harder tasks.	Displays the knowledge/skill to perform all general soldiering skills.
1 2	3 4 5	6 7

MILITARY BEARING/APPEARANCE AND PHYSICAL FITNESS

Maintains self in poor physical condition. Fails to meet military standards for dress or personal hygiene.	Meets Army standards of physical fitness. Dresses neatly and meets Army standards of personal hygiene.	Exceeds Army standards and expectations set for physical fitness. Maintains excellent personal hygiene and proper appearance.
1 2	3 4 5	6 7