

AD-A284 129



ARI Research Note 94-27

Building and Retaining the Career Force: New Procedures for Accessing and Assigning Army Enlisted Personnel

Annual Report, 1992 Fiscal Year

John P. Campbell and Lola M. Zook, Editors

Human Resources Research Organization



94-28916



238PX

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

Manpower and Personnel Research Division
Zita M. Simutis, Director

July 1994



DTIC QUALITY IMPROVED 5

94 9 06 005

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
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

Dale R. Palmer

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

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

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 views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 1994, July	3. REPORT TYPE AND DATES COVERED Final Oct 91 - Sep 92	
4. TITLE AND SUBTITLE Building and Retaining the Career Force: New Procedures for Accessing and Assigning Army Enlisted Personnel-- Annual Report, -1992 Fiscal Year			5. FUNDING NUMBERS MDA903-89-C-0202 63007A 792 2208 C1	
6. AUTHOR(S) Campbell, John P.; and Zook, Lola M., Editors (HumRRO)				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Suite 400 Alexandria, VA 22314			8. PERFORMING ORGANIZATION REPORT NUMBER ---	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-RS 5001 Eisenhower Avenue Alexandria, VA 22333-5600			10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARI Research Note 94-27	
11. SUPPLEMENTARY NOTES This report was prepared under the project Building the Career Force (Human Resources Research Organization, American Institutes for Research, Personnel Decisions (Continued)				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE ---	
13. ABSTRACT (Maximum 200 words) The Career Force research project is the second phase of an Army program to develop a selection and classification system for enlisted personnel based on expected future performance. In the first phase, Project A, a large and versatile data base was collected from a representative sample of Military Occupational Specialties (MOS) and used to (a) validate the Armed Services Vocational Aptitude Battery (ASVAB) and (b) develop and validate new predictor and criterion measures representing the entire domain of potential measures. Building on this foundation, Career Force research is finishing development of the selection/classification system and evaluating its effectiveness, with emphasis on assessing second-tour performance. This third year of the project completed data collection from the Longitudinal Validation cohort, conducted analyses of test results from the second-tour sample, and expanded development of a model of second-tour noncommissioned officer performance. Analyses of the test results are continuing.				
14. SUBJECT TERMS Career force Criterion measures Longitudinal validation			15. NUMBER OF PAGES 237	
			16. PRICE CODE ---	
			(Continued)	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

ARI Research Note 94-27

SUPPLEMENTARY NOTES (Continued)

Research Institute, U.S. Army Research Institute). Contracting Officer's Representative, Michael Rumsey.

SUBJECT TERMS (Continued)

Second-tour performance

PREFACE

This is the third annual report for work completed as part of the Building the Career Force project. It also constitutes the primary technical report of the work completed on several of the project's principal tasks. Consequently, it is a "stand alone" document for Fiscal Year 1992 and does not refer the reader to more detailed descriptions in supplementary reports for that period. The Career Force project extends the major work in selection and classification of Army enlisted personnel that was completed as part of Project A.

The Career Force project includes (1) a replication and extension of the Experimental Battery validities for the selection and classification of first-tour enlisted personnel; (2) validation of the Experimental Battery against end-of-training performance; (3) validation of training performance as a predictor of first-tour job performance; (4) measurement of second-tour performance; (5) validation of the Armed Services Vocational Aptitude Battery (ASVAB), the Experimental Battery, Advanced Individual Training (AIT) performance; and (6) identification of the optimal predictor battery for selection and classification, given certain specific sets of goals and constraints.

The annual report for year one described the results of a series of analyses directed at basic score development for (1) the Experimental Predictor Battery, (2) the End-of-Training performance measures, and (3) the second-tour job performance measures that were administered to the second-tour Concurrent Validation sample (CVII). The performance data from this initial sample of second-tour junior noncommissioned officers (NCO) were also used to develop a latent structure model of second-tour performance. The model hypothesizes six basic components for NCO performance.

The annual report for year two dealt with the analysis of performance data from the Longitudinal Validation I (LVI) sample, which is a sample of approximately 10,000 first-tour incumbents who entered the Army during 1986/87. It is the second of the two major cohorts of enlisted personnel that make up the total Project A/Career Force project data base. The criterion score development, data editing, and performance modeling analyses were each described in turn. The remainder of the report described the results of the basic Longitudinal Validation of the ASVAB and the Project A Experimental Predictor Battery against (1) training performance, (2) first-tour job performance, and (3) second-tour job performance (i.e., the second-tour performance factor scores developed during year one).

This third annual report covers the data collection procedures and criterion analyses for the longitudinal second-tour sample (LVII). It concludes with a confirmation and extension of the model of second-tour NCO performance that was originally developed in the concurrent sample of second-tour soldiers.

The remaining topics in the project are to (1) identify the "optimal" prediction equations, given constraints; (2) estimate the potential differential prediction/classification validity; and (3) analyze the predictability of alternative selection and classification goals. The results of these analyses will be the topics of subsequent reports.

As was the case for years one and two, the writing of this report was very much a collaborative effort by a lot of people. The primary authors for each chapter are indicated in the Table of Contents and also on the first page of each chapter. The editors, and the management, are deeply appreciative of their contributions.

FOREWORD

This document is a description of the research activities conducted during the third year of the project Building the Career Force. This project is the second phase of a research program of unprecedented scope and depth to provide the basis for improving the Army's selection and classification procedures and reenlistment and promotion decisions for soldiers up to the level of sergeant. The thrust for this program came from the practical, professional, and legal need to validate the Armed Services Vocational Aptitude Battery (ASVAB--the U.S. military selection/classification test battery) and other selection variables as predictors of training and performance. The authorization for the program was provided in a letter, Deputy Chief of Staff for Operations, "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, Manpower Reserve Affairs and Logistics (MRA&L), "Enlistment Standards," effective 11 September 1980.

The research program began in 1982 with an effort known as Project A. Project A not only validated the ASVAB against job performance; it further linked indicators of temperament (achievement, discipline, stress tolerance), psychomotor ability (e.g., eye-hand coordination), and spatial ability to job performance. Project A developed new tools for a variety of personnel decisions. Before these tools can be optimally used, however, two critical questions need to be answered: (1) What combinations of aptitude, temperament, psychomotor ability, and spatial ability, measured at or before entry into the Army, best predict later performance in individual military occupational specialties (MOS)? (2) Which indicators of first-tour performance best predict performance in the second tour? These questions will be answered in Building the Career Force.

The third-year Building the Career Force activities described in this report continued analyses of the combined set of initial entry predictor measures developed for selection and classification purposes and end-of-training and first-tour job performance measures to be linked to these predictor measures. Administration of second-tour measures to a sample already tested on initial entry, end-of-training, and first-tour measures was completed and analysis of the data was begun. These analyses are examining longitudinal linkages across the full set of measures, from initial entry into second tour. This will provide an unrivaled information base for setting selection, classification, reenlistment, and promotion policies.

The Director of Military Personnel Management (DMPM) actively sponsored this effort. The DMPM has been periodically briefed on the activities described in this report and has personally taken part in the execution of this project. To ensure that Building the Career Force research achieves its full scientific potential, an advisory group composed of experts in personnel measurement, selection, and classification was established to provide continuing guidance on technical aspects of the research. Members of this Scientific Advisory Group include Philip Bobko, Lloyd Bond, Milton HakeI (Chair), Lloyd Humphreys, Lawrence Johnson, Robert Linn, Mary Tenopyr, and Jay Uhlaner.

BUILDING AND RETAINING THE CAREER FORCE: NEW PROCEDURES FOR ACCESSING AND ASSIGNING ARMY ENLISTED PERSONNEL--ANNUAL REPORT, 1992 FISCAL YEAR

EXECUTIVE SUMMARY

Requirements:

The Career Force project is the second phase of a comprehensive, long-term research program sponsored by the Deputy Chief of Staff for Personnel to improve the selection and assignment of Army enlisted personnel. In the first phase, Project A, existing selection measures were validated against both existing and newly developed performance criteria, and new predictive measures were developed to aid in assignment and promotion decisions. The Career Force project extends the research to measure second-tour job performance and to examine how selection and classification tests administered before a soldier's enlistment can, with measures of performance during that enlistment, predict performance potential for second-tour duty.

Procedure:

In Task 1, measures adopted in Project A to assess the performance of second-tour soldiers have been revised and tested with the Longitudinal Validation (LV) sample first tested in Project A (the second-tour tests of these soldiers occurred when they have been in the Army from 41 to 63 months). The results of these tests are being analyzed to complete the predictive validation of the Armed Services Vocational Aptitude Battery (ASVAB) and the Project A Experimental Predictor Battery, measures of training success, and first-tour job performance tests against the criteria of successful second-tour performance.

Task 2 staff has established an integrated data base and is processing Project A and Career Force data and merging files with related military data. Task 3 covers all analyses being performed to develop the analytic framework needed to evaluate equations for predicting training performance, first-tour performance and attrition, reenlistment, and second-tour performance.

Findings:

The pattern of results from confirmatory analyses of Longitudinal Validation tests has been consistent with the results from earlier LV testing, as well as from the initial Concurrent Validation tests. The models for second-tour NCO job performance that have been developed and refined from the Longitudinal Validation data have strongly confirmed the earlier findings. The description of the latent structure of performance as individuals move from training through their first tour and into their second tour continued to be highly consistent as alternative ways of assessing development and leadership qualities are tested.

Utilization of Findings:

The findings from the validation and model development stages will provide a base for considering a variety of issues inherent in optimal prediction of performance. The long-term results from these analyses of performance potential will be applied in an improved system for selecting and assigning Army manpower in a changing military environment.

BUILDING AND RETAINING THE CAREER FORCE: NEW PROCEDURES FOR ACCESSING AND ASSIGNING ARMY ENLISTED PERSONNEL--ANNUAL REPORT, 1992 FISCAL YEAR

CONTENTS

	Page
INTRODUCTION	1
(John P. Campbell and James H. Harris)	
Building the Career Force: Objectives and Project Design	1
Summary of Project Efforts for Year One	8
Summary of Project Efforts for Year Two	18
Organization of the Current Report	60
LONGITUDINAL VALIDATION SECOND-TOUR DATA COLLECTION	63
(Deirdre Knapp)	
Description of the Measures	63
Obtaining and Scheduling the Required Troop Support	66
Site Coordination	68
Data Collection Procedures	69
ANALYSES OF LVII PERFORMANCE MEASURES	73
(Deirdre Knapp, Charlotte Campbell, Mary Ann Hanson, Ken Bruskiwicz, Cheryl Paullin, Carolyn Hill-Fotouhi, Chris Sager, and Leissa Nelson)	
Job Knowledge and Hands-On Tests	73
Performance Rating Scales	79
Administrative Measures: The Personnel File Form	97
Situational Judgment Test	101
Supervisory Simulation Exercises	114
Summary of Basic Criterion Scores	123
THE LVII DATA FILE	125
(Geoffrey Wilson, Charles T. Keil, Jr., Scott H. Oppler, and Deirdre Knapp)	
Initial Sample Sizes	125
LVII Performance Instruments	126
Extent of Missing Data	127
Treatment of Missing Data	129
Summary of Missing Data Treatment	135

CONTENTS (Continued)

	Page
DEVELOPMENT OF THE SECOND-TOUR PERFORMANCE MODEL FROM THE LONGITUDINAL VALIDATION SAMPLE	139
(Mary Ann Hanson, John P. Campbell, Amy Schwartz McKee, and Rodney A. McCloy)	
Introduction	139
The Modeling Analysis Procedure	143
Results and Discussion	151
Creating LVII Criterion Construct Scores for Validation Analyses	166
Concluding Comments	169
OVERALL SUMMARY AND FUTURE PLANS	173
(John P. Campbell)	
Summary of Year Three	173
Future Plans	175
REFERENCES	177
APPENDIX	
A. TASKS COMPRISING THE HANDS-ON AND JOB KNOWLEDGE COMPONENTS BY MOS (LVII)	A-1
B. TASK, FUNCTIONAL CATEGORY, TASK FACTOR, AND TASK CONSTRUCT SCORES DESCRIPTIVE STATISTICS BY MOS (LVII)	B-1
C. ARMY-WIDE AND MOS-SPECIFIC RATING SCALE CONTENTS	C-1

LIST OF TABLES

Table	1.1 ABLE Rational Composites and Corresponding Content Scales	19
	1.2 Distribution of ABLE Scale Items on ABLE-168 and ABLE-114 Factor Composites	21
	1.3 Mean of Multiple Correlations Computed Within- Job for End-of-Training Sample for ASVAB Factors, Spatial, Computer, JOB, ABLE Rational Composites, and AVOICE	23

CONTENTS (Continued)

	Page
Table 1.4	Mean of Incremental Correlations Over ASVAB Factors Computed Within-Job for End-of-Training Sample for Spatial, Computer, JOB, ABLE Rational Composites, and AVOICE 24
1.5	Measures Administered to Soldiers in LVI Sample 26
1.6	Comparison of LVI and CVI Army-Wide Factor Analysis Results: Pooled Peer/Supervisor Ratings 28
1.7	Composition and Definition of LVI Army-Wide Rating Composites 29
1.8	LVI Sample Sizes for Performance Measures for Batch A MOS 33
1.9	LVI Sample Sizes for Performance Measures for Batch Z MOS 34
1.10	LVI Combined Criteria Data: Percentage of Missing Data for Basic Scores by MOS 35
1.11	LVI Predictor Data: Amount of Missing Data for Paper-and-Pencil Scale Scores 36
1.12	LVI Predictor Data: Amount of Missing Data for Computer-Administered Scale Scores 37
1.13	Mapping of LVII Performance Measures Onto Latent Performance Factors 40
1.14	Mean Intercorrelations Among 13 Summary Criterion Scores for the Batch A MOS in the LVI Sample 41
1.15	Soldiers in CVI and LVI Data Sets With Complete Predictor and First-Tour Criterion Data by MOS 42
1.16	Mean of Multiple Correlations Computed Within-Job for LVI Listwise Deletion Samples for ASVAB Factors, Spatial, Computer, JOB, ABLE Composites, and AVOICE 44
1.17	Mean of Incremental Correlations Over ASVAB Factors Computed Within-Job for LVI Listwise Deletion Samples for Spatial, Computer, JOB, ABLE Composites, and AVOICE 45

CONTENTS (Continued)

		Page
Table	1.18 Mean of Multiple Correlations Computed Within-Job for LVI Setwise Deletion Samples for Spatial, Computer, JOB, ABLE Composites, and AVOICE	46
	1.19 Mean of Incremental Correlations Over ASVAB Factors Computed Within-Job for LVI Setwise Deletion Samples for Spatial, Computer, JOB, ABLE Composites, and AVOICE	47
	1.20 Comparison of Mean Multiple Correlations Computed Within-Job for LVI and CVI Listwise Deletion Samples for ASVAB Factors, Spatial, Computer, JOB, ABLE Composites, and AVOICE	48
	1.21 CVII Sample Sizes by MOS	51
	1.22 Multiple Correlations for ASVAB Factors, ASVAB Subtests, ABLE Composites, and ABLE-114 Scores Against 19 CVII Criterion Variables (All MOS), With Unit Weights	53
	1.23 Multiple Correlations for ASVAB Factors Plus ABLE Composites and Plus ABLE-114 Scores, and for ASVAB Subtests Plus ABLE Composites and Plus ABLE-114 Scores Against 19 CVII Criterion Variables, All MOS	54
	1.24 Multiple Correlations for 10 Sets of Criterion Composite Weights, All MOS	55
	1.25 Numbers of Soldiers With CVI and CVII Data by MOS	57
	1.26 Uncorrected Correlations Between CVI and CVII Raw Criterion Composites Computed Across Total Sample	58
	1.27 Correlations Between CVI Weighted Predictor Composites, CVI Criterion Composites, and CVII Criterion Composites for Raw Scores, Computed on Total Sample	59

CONTENTS (Continued)

	Page
Table 2.1 LVII Data Collection Instruments	64
2.2 LVII Data Collection Schedule	69
2.3 LVII Daily Testing Schedule	71
3.1 Number of LVII Job Knowledge Tests and Items by MOS	75
3.2 Number of LVII Hands-On Tests and Steps by MOS	76
3.3 Intercorrelations Among LVII Job Knowledge Task Factor Scores Across MOS	77
3.4 Intercorrelations Among LVII Hands-On Task Factor Scores Across MOS	78
3.5 Number of Raters Per LVII Ratee by MOS	81
3.6 Self-Reported Familiarity of LVII Raters With Rates	82
3.7 LVII Army-Wide Rating Distributions: Use of Scale Points	83
3.8 LVII Army-Wide Ratings: Dimension-Level Means and Standard Deviations	84
3.9 LVII Army-Wide Ratings: Dimension-Level Interrater Reliability Results	86
3.10 Comparison of LVI and LVII Factor Analysis Results: Non-Supervisory Dimensions	87
3.11 Comparison of LVII and CVII Army-Wide Factor Analysis Results: All Dimensions	88
3.12 Composition of LVII Army-Wide Rating Composites	89
3.13 Definitions of LVII Army-Wide Rating Composites	90
3.14 Interrater Reliability Results for CVII and LVII Army-Wide Rating Composites	91
3.15 Intercorrelations Among LVII and CVII Army-Wide Rating Composites	92

CONTENTS (Continued)

		Page
Table	3.16 MOS-Specific Ratings: LVII and CVII Means (Across Rating Dimensions) of Dimension Means and Standard Deviations	93
	3.17 LVII MOS-Specific Ratings: Dimension Interrater Reliability Results	94
	3.18 MOS-Specific Ratings: Composite Interrater Reliability Results for LVII and CVII	96
	3.19 Interrater Reliability Results for Combat Performance Prediction Scales Score for LVII and CVII	97
	3.20 Administrative Indices Descriptive Statistics for LVII and CVII	100
	3.21 Intercorrelations Among LVII and CVII Administrative Indices of Second-Tour Performance	100
	3.22 Comparison of LVII and CVII Situational Judgment Test Data: Means, Standard Deviations, and Internal Reliabilities	106
	3.23 Comparison of LVII 35-Item and 49-Item Situational Judgment Test Scores: Means, Standard Deviations, and Internal Reliabilities	108
	3.24 LVII 49-Item Situational Judgment Test: Score Intercorrelations for Various Scoring Methods	109
	3.25 LVII 49-Item Situational Judgment Test: Summary of Item Analysis Results	110
	3.26 Situational Judgment Test: Definitions of Factor- Based Subscales	111
	3.27 Situational Judgment Test: Score Intercorrelations for the Factor-Based Subscales and SJT Total Score	112
	3.28 Situational Judgment Test Scores by Combat/NonCombat and by MOS	113
	3.29 Descriptive Statistics for LVII Simulation Exercises	117

CONTENTS (Continued)

		Page
Table	3.30 Factor Analysis Summary Statistics for LVII Simulation Exercises	118
	3.31 LVII Personal Counseling Exercise Scales and Factor Analysis Results	119
	3.32 LVII Disciplinary Counseling Exercise Scales and Factor Analysis Results	120
	3.33 LVII Training Exercise Scales and Factor Analysis Results	121
	3.34 Correlations Among LVII Simulation Exercise Basic Scores	123
	4.1 LVII Sample by MOS	125
	4.2 LVII Sample by Gender	125
	4.3 LVII Sample by Race	126
	4.4 Number of LVII Soldiers With Complete or Partial Data by Criterion Instrument and MOS	128
	4.5 Number of LVII Soldiers With Data by Supplemental Instrument and MOS	129
	4.6 Number of LVII Soldiers With Incomplete Job Knowledge Data	130
	4.7 Percentage of LVII Soldiers With Missing Job Knowledge Scores by MOS	131
	4.8 Percentage of LVII Soldiers With Missing Hands-On Scores by MOS	132
	4.9 Percentage of LVII Soldiers With Missing Data for Performance Rating Composite Scores by MOS	133
	4.10 Percentage of LVII Soldiers With Missing Data for Personnel File Form Basic Scores by MOS	134
	4.11 Percentage of Soldiers With Missing Data for the Situational Judgment Test Total Score by MOS	135

CONTENTS (Continued)

		Page
Table	4.12 Percentage of LVII Soldiers With Missing Data for Simulation Exercises Basic Scores by MOS	135
	4.13 Percentage of LVII Assigned Values by Type of Instrument and MOS	136
	4.14 LVII Combined Criteria Data: Percentage of Soldiers With Missing Data for Composite or Basic Scores by MOS	137
	4.15 Numbers of Soldiers With Complete Data (After Applying Scoring Rules) Across All Instruments and by Type of Instrument and MOS	138
	5.1 List of Basic Criterion Scores Used in LVII Performance Modeling Exercise	142
	5.2 Number of LVII Soldiers With Complete Array of Basic Criterion Scores (Excluding Combat Performance Prediction Scales) by MOS	143
	5.3 CVII Training and Counseling Model	144
	5.4 Consideration/Initiating Structure Model	146
	5.5 Correlations Among the LVII Basic Criterion Scores Based on All Soldiers With Complete Data	148
	5.6 Correlations Among the LVII Basic Criterion Scores With MOS 11B Excluded	149
	5.7 Correlations Between Situational Judgment Test Subscores and Other Selected LVII Basic Criterion Scores	150
	5.8 LISREL Results: Overall Fit Indices for the Training and Counseling Model in the LVII and CVII Samples	152
	5.9 LVII LISREL Results for the Training and Counseling Factor Model: Factor Loadings (Lambda X) and Unique Variance (Theta Delta) Parameter Estimates (Maximum Likelihood)	153
	5.10 LVII LISREL Results for the Training and Counseling Factor Model: Factor Correlations (Phi Estimates)	154

CONTENTS (Continued)

		Page
Table	5.11 Leadership Factor Model	156
	5.12 LVII LISREL Results: Overall Fit Indices for the Training and Counseling and the Leadership Factor Models	157
	5.13 LVII LISREL Results for the Leadership Factor Model: Factor Loadings (Lambda X) and Unique Variance (Theta Delta) Estimates (Maximum Likelihood)	158
	5.14 LVII LISREL Results for the Leadership Factor Model: Factor Correlations (Phi Estimates)	159
	5.15 LVII LISREL Results: Overall Fit Indices for the Leadership Factor Model With Combat Performance Prediction Scales Included	160
	5.16 LVII LISREL Results: Overall Fit Indices for a Series of Nested Models That Collapse the Substantive Factors in the Leadership Factor Model, Based on Total Sample Data	162
	5.17 LVII LISREL Results: Overall Fit Indices for a Series of Nested Models That Collapse the Substantive Factors in the Leadership Factor Model, for Sample Excluding MOS 11B	163
	5.18 CVII LISREL Results: Overall Fit Indices for a Series of Nested Models That Collapse the Substantive Factors in the Leadership Factor Model	164
	5.19 LVII LISREL Results: Overall Fit Indices for the Leadership Factor Model With One Factor Modified, for Clusters of MOS	165
	5.20 LVII LISREL Results: Overall Fit Indices for the Leadership Factor Model With Two Factors Modified, by Race	166
	5.21 Correlations of LVII Basic Criterion Scores With Proposed Construct Scores	168

CONTENTS (Continued)

		Page
Table	5.22 Correlations Between Two LVII Versions of the Achievement and Effort Construct Score (With and Without the Combat Prediction Score) and Other Proposed Construct Scores and the Combat Prediction Overall Composite Score	169

LIST OF FIGURES

Figure	1.1 Building the Career Force: Project management structure	4
	1.2 Project A/Career Force Military Occupational Specialties (MOS)	5
	1.3 Glossary of terms for Project A/Career Force research samples	6
	1.4 Career Force research flow and samples	7
	1.5 Experimental Predictor Battery tests and relevant constructs	10
	1.6 Longitudinal Validation Experimental Battery: Composite scores and constituent basic scores	11
	1.7 Composite scores that reflect End-of-Training performance factors	13
	1.8 Summary list of CVII basic criterion scores	16
	1.9 Relationship of specific variables to overall factors in the CVII performance model	17
	1.10 Hierarchical relationships among Functional Categories, Task Factors, and Task Constructs	31
	1.11 Summary list of LVI basic criterion scores	32
	3.1 Example of a Situational Judgment Test type of item	101
	3.2 Sample scales from LVII Personal Counseling Simulation Exercise	116
	3.3 Summary list of LVII basic criterion scores	124

CONTENTS (Continued)

	Page
Figure 5.1 Final LVII Criterion and Alternate Criterion Constructs based on more parsimonious models	170

BUILDING AND RETAINING THE CAREER FORCE: NEW PROCEDURES FOR ACCESSING AND ASSIGNING ARMY ENLISTED PERSONNEL--ANNUAL REPORT, 1992 FISCAL YEAR

**Chapter 1
INTRODUCTION**

James P. Campbell and James H. Harris

This report is a summary of the major activities undertaken during the third year of a Department of the Army research project entitled Building and Retaining the Career Force. The report covers the period of the 1992 fiscal year, beginning 1 October 1991. The research reported was conducted by a consortium comprised of Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), and Personnel Decisions Research Institute, Incorporated (PDRI, Inc.), under contract to and in collaboration with the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI).

The research effort is the second phase of a two-phase program to develop a selection and classification system for enlisted personnel based on expected future performance. Phase One was Project A (Campbell & Zook, 1991). Its goals were to validate the Armed Services Vocational Aptitude Battery (ASVAB) by collecting data from a representative sample of Military Occupational Specialties (MOS), and to build a large and versatile data base by developing and validating new predictors and criterion measures that represented the entire domain of potential measure.

The goals of Building the Career Force are to determine the longitudinal relationship between the new predictors and first-tour performance, to finalize and administer the measures of second-tour job performance, and to examine how selection and classification tests administered before a soldier's first enlistment, in conjunction with performance during that soldier's first enlistment, predict performance in a second enlistment.

The remainder of this chapter describes the objectives and organization of the project, summarizes the work completed during the first 27 months, and outlines the content to be included in this third annual report.

BUILDING THE CAREER FORCE: OBJECTIVES AND PROJECT DESIGN

The Project A data base, the predictor and criterion measures the project developed, the working models it provided, and its basic analytic work have provided a valuable foundation for the further production of scientific findings and operational products, and for the subsequent investigation of reenlistment decisions, noncommissioned officer (NCO) job performance, NCO promotion decisions, and the identification of NCO potential.

The work encompassed by the Career Force project is intended to accomplish several general goals relevant to building and retaining the career force. The goals may be summarized as follows.

- (1) Build the final pieces required for a complete selection/classification decision-making system for Army enlisted personnel.
- (2) Provide the analytic procedures and data necessary to maximize the system's performance and evaluate its effectiveness.
- (3) Build the foundation for its implementation.

The principal focus is on the greatest possible gains in overall individual performance, for both "can do" and "will do" components of performance, that can be obtained from enhancing the selection/classification system for first- and second-tour enlisted personnel. Maximizing the benefit from a more effective match of people and jobs has always been a goal of the Army personnel system. Given the population demographics for the United States during the coming decade, this goal becomes even more crucial. It is incumbent on virtually every organization to go as far as the state-of-the-art will allow.

This means that the information that is used to make personnel decisions must yield the maximum gain in terms of accuracy and fairness of predictions. It means that the models and procedures used to execute selection and classification decisions must both serve the goals of the organization and maximize the aggregate benefits that can be obtained from using the selection/classification measures (e.g., new computerized tests). It means that the implementation of the system, or any part of it, must serve the needs of the users and also maintain fidelity with the goals on which the system is based.

Specific Research Objectives

The specific scientific objectives of Building the Career Force are to

- (1) Develop a complete array of valid and reliable measures of second-tour performance as an Army NCO, using the Project A prototypes as a starting point.
- (2) Carry out a complete incremental predictive validation of (a) the ASVAB and the Project A Experimental Battery of predictors, (b) measures of training success, and (c) the full array of first-tour performance criteria developed as part of Project A. The criteria against which these three sets of predictors will be validated, both individually and incrementally for each major criterion component, are the second-tour job performance measures.
- (3) Develop a model of second-tour NCO performance that parallels the first-tour performance model from Project A and that identifies the major components of second-tour performance, provides information on their construct validity, and establishes how the major components of performance should be combined for specific prediction or interpretation purposes.
- (4) Develop the analytic framework needed to evaluate the optimal prediction equations for predicting (a) training performance; (b) first-tour performance; (c) first-tour attrition and the reenlistment decision; and (d) second-tour performance, under the

conditions when testing time is limited to a specified amount and when there must be a tradeoff among alternative selection/classification goals (e.g., maximizing aggregate performance vs. minimizing discipline and low-motivation problems vs. minimizing attrition).

- (5) Design and develop a fully functional and user-friendly research data base that includes all relevant personnel data on 1981/82, 1983/84, and 1986/87 accessions, including all Project A and Career Force Project data and all relevant Enlisted Master File (EMF), Accession File, and Army Training Requirements and Resources System (ATRRS) data.

Project Organization

To reflect the requirements of the research, the project is organized as shown in Figure 1.1. Management of the total project is the responsibility of the Project Director. The overall design, execution, and evaluation of the substantive tasks are the responsibility of the Principal Scientist. Oversight and scientific participation is provided by the U.S. Army Research Institute. Guidance is provided by the General Officers Steering Committee and the Scientific Advisory Group.

A brief summary of the work encompassed by the three substantive technical tasks follows:

Task 1 is to revise the measures developed in Project A to measure second-tour soldier performance. The second-tour performance measures were revised and were administered to the Project A Longitudinal Validation (LV) sample, beginning in June 1991. At that time, the soldiers in the sample were in their second tour and had been in the Army anywhere from 41 to 63 months. Once the data have been fully analyzed (under Task 3), it will be possible to complete the incremental predictive validation of the ASVAB and the Project A Experimental Battery, the measures of training success, and the full array of first-tour performance measures developed in Project A, against the second-tour criterion measures.

Task 2 has a single purpose--to establish, manage, and safeguard an integrated research data base (IRDB). As part of the establishment of the IRDB, Task 2 is integrating the Project A longitudinal research data base, extracting and merging data from other military data bases, processing data collected by Project A and this project, and creating workfiles for analyses.

Task 3 is responsible for all analyses performed under this project. The task is organized around the five major data sets to be analyzed: the Longitudinal Validation predictor data (LV), the Longitudinal Validation end-of-training (EOT) data, the Longitudinal Validation first-tour data (LVI), the Concurrent Validation second-tour data (CVII), and the Longitudinal Validation second-tour data (LVII). At the end of the project, Task 3 will have developed the analytic framework necessary to evaluate optimal prediction equations to predict training performance, first-tour performance and attrition, reenlistment, and second-tour performance.

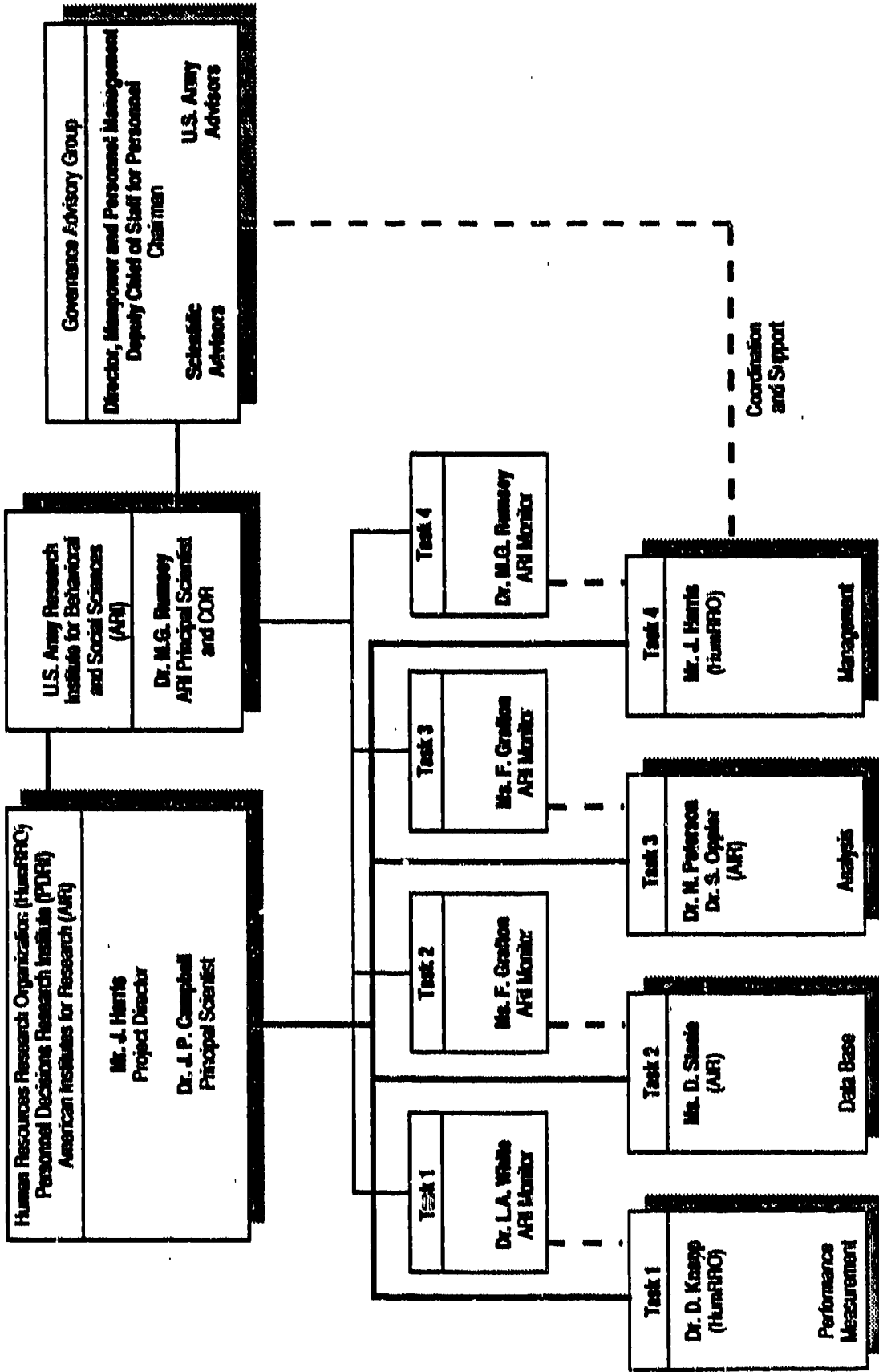


Figure 1.1. Building the Career Force: Project management structure.

Project Design

As will be explained in later sections of this chapter, the remaining chapters of this report all deal with the collection and analyses of data obtained at one major point in the total project design. To set the stage for these discussions, as well as for the summary of work done during years one and two, the basic overall project design is summarized below.

The Research Sample

In general, the combined design for Project A/Career Force encompasses two major cohorts of soldiers (new accessions for 1983/84 and for 1986/87), both of which were followed into their second tour of duty and which collectively have produced six major research samples. For each research sample there is a battery of predictor measures and an array of performance measures. For each of the six samples the predictor battery is composed of the ASVAB and either the Trial Battery or the Experimental Battery version of the new tests developed in Project A (see Campbell & Zook, 1991). There were three distinct arrays of performance measures corresponding to the need to assess (a) training performance, (b) first-tour job performance, and (c) second-tour job performance.

In each sample the individuals to be assessed were selected from two predetermined sets of MOS -- Batch A and Batch Z. They are listed in Figure 1.2. The Batch A MOS had been chosen in Project A to provide maximum coverage of high-density MOS, ASVAB aptitude areas, and Army career management fields; they were given time-intensive MOS-specific job performance and job knowledge tests as well as Army-wide measures. The additional 10 MOS in Batch Z were tested on Army-wide measures and on one MOS-specific test, measuring end-of-training accomplishment.

Batch A		Batch Z	
MOS		MOS	
11B	Infantryman	12B	Combat Engineer
13B	Cannon Crewmember	16S	MANPADS Crewman
19E	M60 Armor Crewman	27E	Tow/Dragon Repairer
19K	M1 Armor Crewman ^a	29E	Comm.-Electronics Radio Repairer
31C	Single Channel Radio Operator	51B	Carpentry/Masonry Specialist
63B	Light-Wheel Vehicle Mechanic	54B	NBC Specialist ^d
71L	Administrative Specialist	55B	Ammunition Specialist
88M	Motor Transport Operator ^b	67N	Utility Helicopter Repairer
91A/B	Medical Specialist/Medical NCO ^c	76Y	Unit Supply Specialist
95B	Military Police	94B	Food Service Specialist
		96B	Intelligence Analyst

^a Except for the type of tank used, this MOS is equivalent to the 19E MOS originally selected for Project A testing.
^b This MOS was formerly designated as 64C.
^c Although 91A was the MOS originally selected for Project A testing, second-tour medical specialists are usually reclassified as 91B.
^d This MOS was formerly designated as 54E.

Figure 1.2. Project A/Career Force Military Occupational Specialties (MOS).

The MOS in the two groups were carefully sampled to represent the variation in job content in the Army occupational structure. In addition, they were selected so as to overrepresent both the combat specialties and those MOS with the larger proportions of women and minority groups. The MOS selection procedure has been described in detail in previous Project A reports (e.g., Campbell, 1987).

A glossary of terms for the samples and for the different measurement batteries is given in Figure 1.3. The six major samples, their approximate size, and the predictor and/or performance batteries that were to be administered to each are shown in Figure 1.4.

Glossary of Terms	
CVI Sample (CVI)	Soldiers who entered the Army between 1 Jul 83 - 30 Jun 84 <u>and</u> were in 1985 Project A Concurrent Validation. They were administered the Trial Predictor Battery and the first-tour job performance measures.
CVII Sample (CVII)	Soldiers who entered the Army between 1 Jul 83 - 30 Jun 84 <u>and</u> were in the 1985 Project A Concurrent Validation (CVI) <u>and</u> the 1988 Second-Tour Concurrent Validation (CVII). They were administered the second-tour job performance measures and were re-administered the ABLE.
LV Sample (LV)	Soldiers in the Longitudinal Validation sample who entered the Army between 20 Aug - 30 Nov 87 <u>and</u> were administered the Experimental Predictor Battery and End-of-Training measures.
LV Training Sample (LVT)	Soldiers in the Longitudinal Validation sample who finished AIT and who were administered the End-of-Training measures.
LVI Sample (LVI)	Soldiers who entered the Army between 20 Aug 86 - 30 Nov 87 <u>and</u> were in the LV Sample <u>and</u> the 1988 First-Tour Longitudinal Validation Sample. They were administered the first-tour job performance measures.
LVII Sample (LVII)	Soldiers who entered the Army between 20 Aug 86 - 30 Nov 87 <u>and</u> were in the LV Sample <u>and</u> the LVI Sample <u>and</u> the Longitudinal Validation (LVII) sample. They were administered the second-tour job performance measures in LVII.
Note. Glossary definitions reflect the original research plan. In actuality, some CVII soldiers did not have CVI data, some LVI soldiers did not have LV data, and some LVII soldiers did not have both LV and LVI data.	

Figure 1.3. Glossary of terms for Project A/Career Force research samples.

Procedure

The data collection procedures for each sample have been described in detail in previous reports (e.g., see Campbell & Zook, 1990). Each data collection involved on-site administration by a trained data collection team headed by a team leader from the contractor staff who worked closely with a designated Army point-of-contact (POC) at the site. A brief characterization of each of the six samples in terms of the timing, location, and duration (per soldier) of the data collection follows.

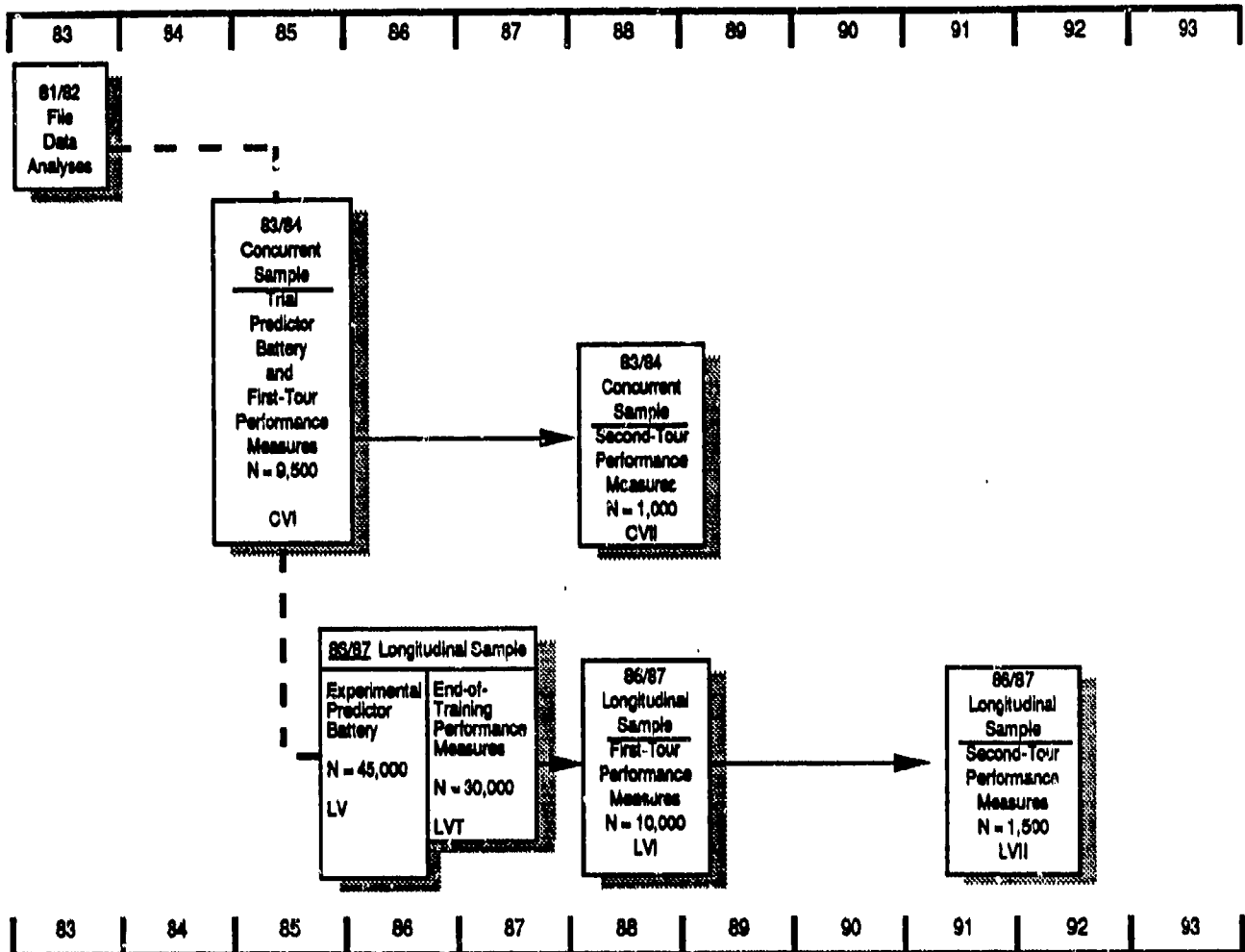


Figure 1.4. Career Force research flow and samples.

The Concurrent Validation (CVI) sample. The data were collected at 13 posts in the continental United States and at multiple locations in Germany. Each individual was assessed for 1 1/2 days on the project-developed first-tour job performance measures and for 1/2 day on the new predictor measures (the Trial Battery). The individuals in the sample had been in the Army for 18-24 months. Data analysis has been completed for this sample.

The Longitudinal Validation (LV) Sample. All individuals were assessed on the 4-hour Experimental Predictor Battery within 2 days of first arriving at their assigned Reception Battalion where they would undergo Basic/Advanced Individual training. Data were collected over a 14-month period at eight Reception Battalions by a permanent, on-site data collection team.

The Longitudinal Validation End-of-Training (LVT) Sample. The EOT performance measures were administered to those individuals in the LV sample who completed Advanced Individual Training (AIT), which could take from 2

months to 6 months, depending on the MOS. The training performance measures consisted of an MOS-specific training achievement test and a series of rating scales completed by peers and drill instructors. Data collection took place during the last three days of AIT.

The Longitudinal Performance Measurement (LVI) Sample. The individuals in the 86/87 cohort who were measured with the Experimental Predictor Battery, completed AIT, and remained in the Army were assessed with the full array of first-tour job performance measures when they were between 18 and 24 months of service. Data collections were conducted at 13 posts in the United States and multiple locations in Europe (primarily in Germany). The administration of the LVI first-tour criterion measures took one day per soldier.

The Concurrent Validation Second-Tour (CVII) Sample. The same data collection teams that administered the first-tour performance measures to the LVI sample also administered the second-tour performance measures at the same location and during the same time periods to a sample of junior NCOs from the 83/84 cohort in their second tour of duty (4-5 years of service). Every attempt was made to include second-tour personnel from the designated MOS who had been part of the first-tour Concurrent Validation sample (CVI). The CVII data collection took one day per soldier.

The Longitudinal Validation Second-Tour (LVII) Sample. The personnel in this sample are members of the 86/87 cohort from the designated MOS who were part of the LV (predictors and training performance measures) and LVI (first-tour job performance measures) samples and who reenlisted for a second tour of duty. The revised second-tour performance measures were administered at 15 U.S. posts, multiple locations in Germany, and two locations in Korea. The LVII performance assessment took one day per soldier.

Current Status

The LVII data collection was completed during the summer of 1992. The content of this third annual report is based on data from LVII samples.

SUMMARY OF PROJECT EFFORTS FOR YEAR ONE

As described in the first annual report (Campbell & Zook, 1990), the objectives of the project's first year were focused on developing a full design for the data base and on analyzing basic scores for (a) the final version of the Experimental Predictor Battery (EB), (b) the End-of-Training (EOT) performance measures, and (c) the second-tour criterion measures used to assess NCO performance in the second-tour Concurrent Validation (CVII) sample. The data from the End-of-Training (EOT) and second-tour Concurrent Validation (CVII) performance assessment were also used to formulate both a model of training performance and a model of second-tour (junior NCO) job performance. That is, the basic scores from the individual performance measures were aggregated into factor scores that represented, as well as possible, the major components, or latent structure, of training performance and second-tour job performance.

By the end of year one, the data collection for the Longitudinal Validation first-tour performance assessments had been completed, but the data cleaning and editing were still in progress and the analysis of the LVI performance measures had not yet begun.

Data Base Design

As described in the first-year annual report, the Career Force data base design allows access at any level of score aggregation. The report describes each variable and the amount of information that is available. The data are accessed via a secure system that requires prior approval by the Army.

The data base also includes data, for various periods relevant to the research, from the following operational files maintained by the Army:

- Applicant/Accessions Data
- Training Data
- Enlisted Master File Cohort Data
- World-Wide Locator Data

Continuous updates to the Career Force data base are made only for the Enlisted Master File. This file is updated on a quarterly basis with official Army information for each individual in all Project A and Career Force Project cohorts--in particular, current pay grade, reenlistment status, and separation status.

Basic Scores for the Experimental Battery

During year one, much effort was devoted to analyzing the data that had been obtained by administering the Experimental Predictor Battery to approximately 45,000 new accessions in the Longitudinal Validation sample. A number of data editing procedures were compared and evaluated, and great care was taken to maximize data quality for the information that was entered into the final data file. The psychometric properties and subgroup differences for each measure were analyzed, and a series of exploratory and confirmatory analyses were conducted to identify the basic predictor scores within each domain that would be used in the validation analyses.

The final array of tests in the Experimental Battery and the constructs they are intended to measure are shown in Figure 1.5. The 31 basic scores that are obtained from the specific test indicators are shown in Figure 1.6 (Campbell & Zook, 1990).

There was a very high degree of consistency between the Concurrent Validation and the Longitudinal Validation in terms of the factor structures of the various measures. The resulting definitions of the basic predictor scores to be used in the validation analyses were quite similar.

Basic Scores for the End-of-Training Measures

During year one, the data from the school knowledge test and seven training performance rating scales administered at the end of training were analyzed in terms of their psychometric properties and factor structure.

Test/Measure	Construct
<i>Paper-and-Pencil Spatial Tests</i>	
Assembling Objects	Spatial Visualization-Rotation
Object Rotation	Spatial Visualization-Rotation
Maze	Spatial Visualization-Scanning
Orientation	Spatial Orientation
Map	Spatial Orientation
Reasoning	Induction
<i>Computer-Administered Tests</i>	
Simple Reaction Time	Reaction Time (Processing Efficiency)
Choice Reaction Time	Reaction Time (Processing Efficiency)
Short-Term Memory	Short-Term Memory
Perceptual Speed and Accuracy	Perceptual Speed and Accuracy
Target Identification	Perceptual Speed and Accuracy
Target Tracking 1	Psychomotor Precision
Target Shoot	Psychomotor Precision
Target Tracking 2	Multilimb Coordination
Number Memory	Number Operations
Cannon Shoot	Movement Judgment
<i>Temperament, Interest, and Job Preference Measures</i>	
Assessment of Background and Life Experiences (ABLE)	Adjustment Dependability Achievement Physical Condition Leadership (Potency) Locus of Control Agreeableness/Likability
Army Vocational Interest Career Examination (AVOICE)	Realistic Interest Conventional Interest Social Interest Investigative Interest Enterprising Interest Artistic Interest
Job Orientation Blank (JOB)	Job Security Serving Others Autonomy Routine Work Ambition/Achievement

Figure 1.5. Experimental Predictor Battery tests and relevant constructs.

<u>ASVAB Factor Composites</u>	<u>Computer-Administered Test Composites*</u>	<u>ABLE Composites</u>	<u>AVOICE Composites</u>
Quantitative Mathematics Knowledge	Psychomotor	Achievement Orientation	Rugged/Outdoors
Arithmetic Reasoning	Target Tracking 1 Distance	Self-Esteem	Combat
Technical	Target Tracking 2 Distance	Work Orientation	Rugged Individualism
Auto/Shop Information	Cannon Shoot Time Score	Energy Level	Firearms Enthusiast
Mechanical Comprehension	Target Shoot Distance	Leadership Potential	Audiovisual Arts
Electronics Information	Movement Time	Dominance	Drafting
Speed	Pooled Movement Time	Dependability	Audiographics
Coding Speed	Perceptual Speed	Traditional Values	Aesthetics
Number Operations	Perceptual Speed & Accuracy (DT)	Conscientiousness	Interpersonal
Verbal	Target Identification (DT)	Modeling	Medical Services
Word Knowledge	Basic Speed	Adjustment	Leadership/Guidance
Paragraph Comprehension	Simple Reaction Time (DT)	Emotional Stability	Skilled/Technical
General Science	Choice Reaction Time (DT)	Cooperativeness	Science/Chemical
	Perceptual Accuracy	Cooperativeness	Computers
<u>Paper-and-Pencil Test Composite</u>	Perceptual Speed & Accuracy (PC)	Internal Control	Electronic Communication
Spatial	Target Identification (PC)	Internal Control	Administrative
Assembling Objects Test	Basic Accuracy	Physical Condition	Clerical/Administrative
Object Rotation Test	Simple Reaction Time (PC)	Physical Condition	Warehousing/Shipping
Maze Test	Choice Reaction Time (PC)	High Job Expectations	Food Service
Orientation Test	Number Speed and Accuracy	Pride	Food Service - Professional
Map Test	Number Memory (Operation DT)	Job Security	Food Service - Employee
Reasoning Test	Number Memory (PC)	Serving Others	Protective Services
	Short-Term Memory	Ambition	Fire Protection
	Short-Term Memory (PC)	Job Routine	Law Enforcement
	Short-Term Memory (DT)	Routine	Structural/Machines
		Job Autonomy	Mechanics
		Autonomy	Heavy Construction
			Electronics
			Vehicle Operator

*DT = Decision Time and PC = Proportion Correct

Figure 1.6. Longitudinal Validation Experimental Battery: Composite scores and constituent basic scores.

Confirmatory techniques were used to identify the "model" of training performance that best represented the covariances among the observed measures. That is, an a priori set of alternative models was proposed and evaluated in terms of the degree to which they fit the data. In the end six basic scores were proposed, two based on the knowledge test and four based on the rating scales. A brief characterization of the six scores is given in Figure 1-7.

These six scores serve both as criterion measures (for the Experimental Battery) and as predictors (of first-tour and second-tour job performance) in later validation analyses.

Development of Second-Tour Performance Scores (CVII)

The performance measures used in the CVII sample, and their development, have been described in detail in previous reports (Campbell, 1991; Campbell & Zook, 1991). First-tour measures were revised for use with second-tour personnel and new measures reflecting the unique components of second-tour jobs were added. A summary description of the specific measures is given below.

Rating Scales

On the basis of second-tour critical incident analyses, the Army-wide Behaviorally Anchored Ratings Scales (BARS) and MOS-specific BARS were revised and scales having to do with leadership and supervision were added. Further, based on job analysis data, seven new scales pertaining to supervision and leadership responsibilities were also added. A full list of the Army-wide rating scales is shown below. Not shown are the MOS BARS for each MOS, which were revised to reflect second-tour performance demands, and the Combat Performance Prediction Scales, which were the same as those used in LVI, and which were not administered to female NCOs during CVII.

Army-Wide Behavior Scales:

1. Demonstrating Technical Knowledge and Skill
2. Demonstrating Effort
3. Supervising Subordinates
4. Following Regulations and Orders
5. Demonstrating Integrity
6. Training and Development of Subordinates
7. Maintaining Equipment
8. Physical Fitness
9. Self-Development
10. Showing Consideration for Subordinates
11. Demonstrating Appropriate Military Bearing
12. Demonstrating Appropriate Self-Control

Additional Leadership Scales:

13. Serving as a Role Model
14. Communication With Subordinates
15. Personal Counseling
16. Monitoring Subordinate Performance
17. Organizing Missions/Operations

EOT RATING SCALE BASED SCORES

1) Effort and Technical Skill (ETS)

Technical Knowledge/Skill: How effective is each soldier in acquiring job/soldiering knowledge and skill?

Effort: How effective is each soldier in displaying extra effort?

2) Maintaining Personal Discipline (MPD)

Following Regulations and Orders: How effective is each soldier in adhering to regulations, orders, and SOP and displaying respect for superiors?

Self Control: How effective is each soldier in controlling own behavior related to aggressive acts?

3) Physical Fitness and Military Bearing (PFB)

Military Appearance: How effective is each soldier in maintaining proper military appearance?

Physical Fitness: How effective is each soldier in maintaining military standards of physical fitness?

4) Leadership Potential (LEAD):

Leadership Potential: Evaluate each soldier on his or her potential effectiveness as a leader. Do not necessarily rate on the basis of present performance.

EOT KNOWLEDGE TEST BASED SCORES

5) Basic Knowledge Score: Items measuring knowledge requirements common to all MOS.

6) Technical Knowledge Score: Items measuring technical knowledge requirements specific to each MOS.

Figure 1.7. Composite scores that reflect End-of-Training performance factors.

18. Personnel Administration
19. Performance Counseling

General Scales:

20. Overall Effectiveness
21. Senior NCO Potential

Situational Judgment Test (SJT)

A new paper-and-pencil measure of supervisory judgment was developed by describing prototypical judgment situations and asking the respondent to select the most appropriate and the least appropriate course of action. The situation descriptions and the scoring keys were refined through extensive subject matter expert (SME) judgments.

Supervisory Simulation Exercises

These measures were developed to assess NCO performance in job areas that were judged to be best assessed through the use of interactive exercises. The simulations were designed to evaluate performance in counseling and training subordinates. A trained evaluator (role player) played the part of a subordinate to be counseled or trained and the examinee assumed the role of a first-line supervisor who was to conduct the counseling or training. Evaluators also scored the examinee's performance, using a standard set of rating scales.

Here are brief descriptions of the three simulation exercises:

- **Personal Counseling Simulation:** A PFC is exhibiting declining job performance and personal appearance. Recently, the PFC's wall locker was left unsecured. The supervisor has decided to counsel the PFC about these matters.
- **Disciplinary Counseling Simulation:** There is convincing evidence that the PFC lied to get out of coming to work today. The PFC has arrived late to work on several occasions and has been counseled for lying in the past. The PFC has been instructed to report to the supervisor's office immediately.
- **Training Simulation:** The commander will be observing the unit practice formation in 30 minutes. The private, although highly motivated, is experiencing problems with the hand salute and about face.

For each exercise, examinee performance was evaluated on 3-point rating scales reflecting specific behaviors tapped by the exercises and a 5-point overall effectiveness rating scale.

Factor analyses of the ratings data suggested that each simulation could be scored in terms of the content of the NCO's behavior (i.e., did he or she do or say the right things) and the process, or style, with which the counseling steps were carried out.

Administrative Measures

The self-report Personnel File Form (PFF) used in LVI was modified for use with second tour and six administrative indices of performance were obtained.

Job Knowledge and Hands-On Measures

The content of each of these measures was revised on the basis of the second-tour job analyses and the revised instruments were subjected to extensive SME review. Analyses of alternative aggregations of item and scale scores from both of these measures resulted in the adoption of a general (Army-wide) and an MOS-specific score for each of them.

Final Array of Second-Tour Basic Performance Scores

After extensive analyses of their psychometric properties and factor structures, based on CVII data, the final array of basic second-tour performance scores was as shown in Figure 1.8. There were 22 basic scores. Scores from this array became the basis for the second-tour performance modeling analysis in CVII.

Development of the CVII Second-Tour Performance Model

The basic CVII performance scores served as input to the development of a latent structure model for second-tour performance. Based on a consensus of the project staff, three major alternatives could be used to explain the observed correlations. Consequently, the competing models that were evaluated for comparative goodness of fit, using the LISREL VI program (Jöreskog & Sörbom, 1985), were the following:

- (1) First-Tour Model: Included five substantive and two methods factors, with the SJT and Simulation variables all loading on the Effort and Leadership factor.
- (2) Leadership Factor Model: Included a sixth substantive factor with the SJT, Simulation, and Leadership Rating factor variables all loading on this factor. This model was evaluated with and without a separate simulation "methods" factor.
- (3) Training and Counseling Factor Model: Included a sixth substantive factor with just the Simulation variables. No separate simulation methods factor could be estimated under this model.

Of the three models, the Training and Counseling Factor Model provided the closest fit to the observed data. A result of considerable interest was that the SJT (a paper-and-pencil measure) fit best with the Effort and Leadership factor, in spite of the method variance involved.

The basic scores that have been used to represent the latent variables are as shown in Figure 1.9. For validation analysis purposes, the six substantive factor scores are obtained by standardizing and summing the basic scores within each factor.

Hands-On Performance Test

1. MOS-specific task performance score
2. General (common) task performance score

Job Knowledge Test

3. MOS-specific task knowledge score
4. General (common) task knowledge score

Army-Wide Rating Scales

5. Leadership/supervision composite
6. Technical skill and effort composite
7. Personal discipline composite
8. Physical fitness and military bearing composite

MOS-Specific Rating Scales

9. Overall MOS composite

Combat Performance Prediction Scales

10. Overall Combat Prediction scale composite (available for males only)

Personnel File Form

11. Awards and Certificates
12. Articles 15/Flag Actions (Disciplinary Actions)
13. Physical Readiness
14. M16/M19 Qualification
15. Military Training Courses
16. Promotion Rate

Situational Judgment Test

17. Total score obtained by subtracting the total "ineffectiveness" score from the total "effectiveness" score

Supervisory Simulation Exercises

18. Personal Counseling: Process
19. Personal Counseling: Content
20. Disciplinary Counseling: Process
21. Disciplinary Counseling: Content
22. Training: Total composite score

Figure 1.8. Summary list of CVII basic criterion scores.

Latent Variables in the CVII Performance Model

- **Core Technical Proficiency (CTP)**
 - MOS-Specific Hands-On
 - MOS-Specific Job Knowledge
- **General Soldiering Proficiency (GSP)**
 - General (Common) Hands-On
 - General (Common) Job Knowledge
- **Effort and Leadership (ELS)**
 - Awards and Certificates
 - Military Training Courses
 - Promotion Rate
 - Leadership/Supervision Rating Composite
 - Technical Skill/Effort Rating Composite
 - Overall MOS Rating Composite
 - Situational Judgment Test Total Score
- **Personal Discipline (MPD)**
 - Disciplinary Actions (reversed)
 - Personal Discipline Rating Composite
- **Physical Fitness/Military Bearing (PFB)**
 - Physical Readiness Score
 - Physical Fitness/Bearing Rating Composite
- **Training and Counseling Subordinates (TCS)**
 - Simulation Exercise - Personal Counseling Content
 - Simulation Exercise - Personal Counseling Process
 - Simulation Exercise - Disciplinary Content
 - Simulation Exercise - Disciplinary Process
 - Simulation Exercise - Training
- **Written Methods (WM)**
 - MOS-Specific Knowledge
 - Common Soldiering Knowledge
 - Situational Judgment Test
- **Ratings Methods (RM)**
 - Four Army-Wide Rating Composites
 - Overall MOS Rating Composite

Figure 1.9. Relationship of specific variables to overall factors in the CVII performance model.

SUMMARY OF PROJECT EFFORTS FOR YEAR TWO

As described in the second annual report (Campbell & Zook, 1994), year two was a period of score development, model building, and basic validation analyses for (a) training performance (EOT), (b) first-tour performance (LVI), and (c) second-tour performance (CVII). During year two, the second-tour longitudinal data collection (LVII) began and was ongoing.

Objectives

The specific objectives for the second-year annual report were as follows.

- (1) Describe the development of alternative scores for the Background and Life Experiences (ABLE) instrument.
- (2) Describe the basic validation analyses for the prediction of performance in training.
- (3) Describe the development of basic scores for the longitudinal sample first-tour performance measures.
- (4) Describe the replication/confirmation of the first-tour performance model and the basic Longitudinal Validation analyses for the Experimental Predictor Battery against first-tour performance.
- (5) Describe the basic validation analyses for the prediction of second-tour performance, using the CVII sample.
- (6) Report the results of a preliminary analysis of the prediction of second-tour performance from first-tour predictors and performance.

Development of Alternative ABLE Factor Composites

As part of Project A, and based on the results of an extensive review of the literature, 10 temperament scales had been developed to form the ABLE. These constructs were selected as the most promising for predicting performance in Army enlisted occupational specialties. In addition, four validity scales were added to detect inaccuracies in self-reports of temperament and a self-report measure of physical condition was also included (see Hough, Eaton, Dunnette, Kamp, & McCloy, 1990, for more information on the development of ABLE). To develop a set of conceptually meaningful construct (composite) scores, Peterson et al. (1992) carried out both exploratory and confirmatory factor analyses on the correlation among the content scale scores.

The resulting seven temperament constructs (composites) and associated ABLE scales are shown in Table 1.1. The constructs of Dependability, Dominance (Surgency), Adjustment, and Cooperativeness have counterparts in the Big Five personality dimensions described by Norman (1963) and Goldberg (1981). Conversely, Achievement and Internal Control are not in the Big Five taxonomy, but were among the strongest predictors of job performance in the Project A review of the temperament domain (see Hough, 1992 for more details on the relationship of ABLE constructs to the Big Five).

Table 1.1

ABLE Rational Composites and Corresponding Content Scales

Composite	ABLE Scale
Achievement Orientation	Self-Esteem Work Orientation Energy Level
Leadership Potential	Dominance
Dependability	Traditional Values Conscientiousness Nondelinquency
Adjustment	Emotional Stability
Cooperativeness	Cooperativeness
Internal Control	Internal Control
Physical Condition	Physical Condition

As noted previously, a rational/theoretical approach was the primary method used in developing ABLE. An alternative empirical procedure emphasizes the internal covariance structure of a set of items and uses factor analytic methods. Consequently, during year two, internal scale construction methods were used to increase, through homogeneous keying, the internal consistency of ABLE composites and to decrease their intercorrelations.

Results from factor analyses of 199 items were used to form seven preliminary composites. These composites contained 99 items. Next, correlations between the remaining content-type items (excluding the validity scale items) and the preliminary factor composites were examined and each remaining item was assigned to the composite with which it had the highest correlation. The seven factor composites resulting from this procedure used 168 items and are called the ABLE-168 composites. In all, 125 items were assigned in the same way on the ABLE-168 composites and ABLE rational composites.

As a second alternative, an item was retained only if it correlated at least .33 with the scale for which it was assigned and had a higher correlation with its own composite (by .03) than any other. In addition, several items that added only minimally to internal consistency were dropped. The resulting set of composites had a total of 114 items and are called the ABLE-114 composites. Eighty-nine of these items were assigned in the same way on ABLE-114 and the ABLE rational composites.

The three scoring methods converged to yield seven similar temperament constructs. The composites measuring the same constructs were very highly correlated ($r = .88$ to 1.0).

ABLE-114 composites had greater discriminant validity than either the ABLE-168 factor composites or the ABLE rational composites. The average correlation among the composites (off-diagonal elements) was .40 for ABLE-114, and .47 for the ABLE rational composites and ABLE-168.

Table 1.2 shows the distribution of items on ABLE-168 and ABLE-114 for each of the ABLE content scales. Items outside the shaded areas were assigned differently on the rational and factor composites.

As shown in Table 1.2, there is much overlap between the rational and factor composites. However, approximately 25 percent of item assignments for the factor composites were different from those used for the rational composites. Most of these are consistent with results from previous research and/or can be understood on the basis of item content.

In sum, there are three alternative ABLE composites measuring seven temperament constructs. The 114-item form is shorter and has higher discriminant validity than the other two sets of composites, with little apparent loss of reliability. Subsequent analyses in the Career Force Project examine the criterion-related validities of these alternative sets of composites.

Prediction of Performance in Training

The objectives of analyses of the end-of-training (EOT) data were to:

- (1) Compute the validities for ASVAB and Experimental Battery predictors against rating measures and also paper-and-pencil test measures of training performance.
- (2) Compare the validities of four alternative sets of ASVAB scores.
- (3) Compare the validities of three alternative sets of ABLE scores.
- (4) Assess the incremental validities for the Experimental Battery predictors over ASVAB.

Procedure

The EOT validation analysis consisted of the following steps:

- A) Multiple correlations between each set of predictor scores and each set of criterion scores were computed separately by MOS and then averaged across the Batch A MOS and across all MOS.

Table 1.2
Distribution of ABLE Scale Items on ABLE-168 and ABLE-114 Factor Composites

ABLE Scale	No. of Items	ABLE Factor Composite									
		Achievement Orientation	Leadership Potential	Dependability	Adjustment	Cooperativeness	Internal Control	Physical Condition			
Self-Esteem	12 (6)		10 (6)		2 (0)						
Work Orientation	19 (15)	18 (14)	1 (1)								
Energy Level	21 (9)	13 (6)			6 (1)					2 (2)	
Dominance	12 (12)		12 (12)								
Traditional Values	11 (7)			5 (4)			5 (3)				
Conscientiousness	15 (11)	9 (8)		6 (3)							
Nondelinquency	26 (13)			20 (13)							
Emotional Stability	17 (11)				17 (11)						
Cooperativeness	18 (11)			2 (1)		16 (10)					
Internal Control	16 (11)	2 (0)			2 (1)		12 (10)				
Physical Condition	6 (5)									6 (5)	
Poor Impression	2 (2)				2 (2)						
Total	169 (114)	42 (28)	73 (19)	33 (21)	29 (15)	16 (10)	17 (13)			8 (8)	

Note. ABLE-114 items are shown in parentheses. Shaded areas indicate convergence between the rational and factor composites.

- 1) The ASVAB predictor set was represented by:
 - a) The nine ASVAB subtest scores
 - b) The four ASVAB factor scores
 - c) The Armed Forces Qualification Test (AFQT)
 - d) The MOS-appropriate Aptitude Area composite score
- 2) The ABLE predictor set was represented by three sets of scores:
 - a) The seven rational scales
 - b) Seven empirical scales that retained 168 items
 - c) Seven empirical scales that retained only 114 items
- 3) Each of the other predictor sets (i.e., spatial, computer, AVOICE, JOB) was represented as in previous analyses.

All results were adjusted for shrinkage and corrected for multivariate range restriction.

- B) Incremental validity was computed for each set of Experimental Battery predictors over the ASVAB.
- C) Multiple correlations were computed between each set of predictor scores and a "Peer 1" rating, a "Peer 2" rating, a supervisor rating, and various combinations.

Results

To summarize the principal findings, multiple correlations for six predictor sets are shown in Table 1.3; the incremental validities are summarized in Table 1.4. In general, ASVAB shows high validity against the school knowledge measures and the relative validities for the four ratings factors are as would be expected on the basis of the factors. The ABLE does not predict the "will do" factors quite as well as it did in CVI but it predicts the "can do" factors somewhat better.

These results indicate that the level of validity of the ASVAB factors for predicting the School Knowledge (SK) test scores was extremely high, especially for the Technical (SK-Tech) and Total (SK-Total) scores. Likewise, the spatial composite and the computer battery produced high validities for these criteria.

Results from other analyses indicate that peer ratings of training performance are more accurately predicted than supervisor ratings of training performance. This suggests that peer ratings may be more valid training measures than supervisor ratings, presumably because, in training, peers generally have greater opportunity to observe ratees than do supervisors. This comparison is confounded, however, by the greater reliability of the peer ratings that is, at least in part, due to the fact that they are based on more raters per ratee than are the supervisor ratings. Yet analyses at the 1-rater level corroborate the notion that the peer ratings have more utility than the supervisor ratings for assessing training performance.

Table 1.3

Mean of Multiple Correlations Computed Within-Job for End-of-Training Sample for ASVAB Factors, Spatial, Computer, JOB, ABLE Rational Composites, and AVOICE

Criterion ^a	MOS	No. of MOS ^b	ASVAB Factors [4]	Spatial [1]	Computer [8]	JOB [3]	ABLE Comp. [7]	AVOICE [8]
Peer-ETS	Batch A	11	41 (07)	35 (05)	36 (05)	24 (06)	19 (09)	22 (07)
	All MOS	22	43 (13)	37 (10)	33 (14)	23 (11)	23 (12)	23 (10)
Peer-MPD	Batch A	11	25 (04)	22 (05)	21 (05)	09 (07)	19 (05)	11 (07)
	All MOS	22	26 (11)	22 (08)	15 (10)	12 (10)	22 (10)	09 (09)
Peer-PFB	Batch A	11	14 (09)	05 (06)	11 (05)	05 (05)	29 (06)	07 (07)
	All MOS	22	19 (14)	10 (11)	12 (09)	09 (12)	26 (11)	10 (10)
Peer-LEAD	Batch A	11	30 (10)	24 (07)	28 (07)	18 (09)	22 (09)	17 (10)
	All MOS	22	30 (16)	26 (12)	25 (16)	20 (14)	22 (12)	16 (14)
Supv-ETS	Batch A	11	21 (06)	18 (05)	17 (10)	10 (08)	09 (10)	11 (10)
	All MOS	22	27 (15)	22 (11)	18 (13)	10 (10)	11 (12)	10 (10)
Supv-MPD	Batch A	11	13 (09)	12 (07)	11 (08)	06 (06)	05 (06)	06 (06)
	All MOS	22	16 (16)	14 (11)	10 (13)	06 (08)	05 (07)	04 (06)
Supv-PFB	Batch A	11	11 (07)	09 (05)	09 (08)	06 (05)	11 (09)	07 (07)
	All MOS	22	16 (15)	13 (12)	11 (15)	05 (07)	11 (11)	05 (06)
Supv-LEAD	Batch A	11	15 (10)	14 (09)	13 (10)	08 (08)	10 (11)	08 (09)
	All MOS	22	19 (17)	17 (11)	12 (12)	11 (09)	11 (12)	07 (09)
SK-Basic	Batch A	9	68 (06)	57 (06)	57 (06)	38 (05)	30 (07)	37 (05)
	All MOS	20	67 (08)	58 (07)	55 (14)	36 (10)	31 (14)	37 (11)
SK-Tech	Batch A	11	76 (05)	63 (05)	61 (05)	41 (07)	33 (05)	44 (07)
	All MOS	22	75 (06)	62 (08)	59 (06)	38 (11)	33 (13)	40 (12)
SK-Total	Batch A	11	78 (03)	65 (04)	64 (03)	43 (07)	34 (05)	45 (06)
	All MOS	22	77 (05)	65 (07)	62 (07)	40 (11)	35 (14)	42 (13)

Note: Corrected for range restriction and adjusted for shrinkage (Rozeboom formula B). Numbers in parentheses are standard deviations. Numbers in brackets are the numbers of predictor scores entering prediction equations. Decimals omitted.

^a ETS = Effort and Technical Skill; MPD = Maintaining Personal Discipline; PFB = Physical Fitness and Military Bearing; LEAD = Leadership Potential; SK = School Knowledge.
^b Number of MOS for which validities were computed.

Table 1.4

Mean of Incremental Correlations Over ASVAB Factors Computed Within-Job for End-of-Training Sample for Spatial, Computer, JOB, ABLE Rational Composites, and AVOICE

Criterion ^a	MOS	No. of MOS ^b	A4 ASVAB Factors [4]	A4+ Spatial [5]	A4+ Computer [12]	A4+ JOB [7]	A4+ ABLE Comp. [11]	A4+ AVOICE [12]
Peer-ETS	Batch A	11	<i>.41</i> (.07)	<i>.42</i> (.07)	<i>.42</i> (.06)	<i>.41</i> (.07)	<i>.44</i> (.06)	<i>.41</i> (.07)
	All MOS	22	<i>.43</i> (.13)	<i>.42</i> (.14)	<i>.40</i> (.16)	<i>.42</i> (.13)	<i>.45</i> (.11)	<i>.41</i> (.14)
Peer-MPD	Batch A	11	<i>.25</i> (.04)	<i>.25</i> (.05)	<i>.24</i> (.05)	<i>.25</i> (.05)	<i>.34</i> (.06)	<i>.24</i> (.07)
	All MOS	22	<i>.26</i> (.11)	<i>.25</i> (.11)	<i>.22</i> (.12)	<i>.25</i> (.12)	<i>.33</i> (.11)	<i>.22</i> (.11)
Peer-PFB	Batch A	11	<i>.14</i> (.09)	<i>.13</i> (.09)	<i>.17</i> (.07)	<i>.15</i> (.09)	<i>.31</i> (.09)	<i>.15</i> (.09)
	All MOS	22	<i>.19</i> (.14)	<i>.18</i> (.14)	<i>.16</i> (.12)	<i>.19</i> (.17)	<i>.30</i> (.14)	<i>.18</i> (.11)
Peer-LEAD	Batch A	11	<i>.30</i> (.10)	<i>.30</i> (.10)	<i>.31</i> (.08)	<i>.30</i> (.11)	<i>.35</i> (.09)	<i>.29</i> (.13)
	All MOS	22	<i>.30</i> (.16)	<i>.30</i> (.17)	<i>.28</i> (.18)	<i>.31</i> (.18)	<i>.34</i> (.15)	<i>.28</i> (.18)
Supv-ETS	Batch A	11	<i>.21</i> (.06)	<i>.21</i> (.07)	<i>.19</i> (.09)	<i>.20</i> (.06)	<i>.19</i> (.12)	<i>.17</i> (.12)
	All MOS	22	<i>.27</i> (.15)	<i>.26</i> (.15)	<i>.24</i> (.15)	<i>.25</i> (.15)	<i>.25</i> (.19)	<i>.22</i> (.16)
Supv-MPD	Batch A	11	<i>.13</i> (.09)	<i>.12</i> (.09)	<i>.11</i> (.09)	<i>.11</i> (.09)	<i>.13</i> (.11)	<i>.11</i> (.10)
	All MOS	22	<i>.16</i> (.16)	<i>.16</i> (.16)	<i>.12</i> (.17)	<i>.14</i> (.17)	<i>.16</i> (.16)	<i>.11</i> (.14)
Supv-PFB	Batch A	11	<i>.11</i> (.07)	<i>.11</i> (.07)	<i>.10</i> (.08)	<i>.10</i> (.07)	<i>.16</i> (.09)	<i>.10</i> (.09)
	All MOS	22	<i>.16</i> (.15)	<i>.15</i> (.14)	<i>.12</i> (.15)	<i>.14</i> (.13)	<i>.18</i> (.13)	<i>.11</i> (.13)
Supv-LEAD	Batch A	11	<i>.15</i> (.10)	<i>.14</i> (.10)	<i>.14</i> (.11)	<i>.14</i> (.10)	<i>.16</i> (.13)	<i>.13</i> (.12)
	All MOS	22	<i>.19</i> (.17)	<i>.19</i> (.17)	<i>.15</i> (.15)	<i>.19</i> (.16)	<i>.20</i> (.17)	<i>.15</i> (.15)
SK-Basic	Batch A	9	<i>.68</i> (.06)	<i>.69</i> (.06)	<i>.68</i> (.06)	<i>.68</i> (.06)	<i>.68</i> (.07)	<i>.68</i> (.06)
	All MOS	20	<i>.67</i> (.08)	<i>.68</i> (.08)	<i>.65</i> (.16)	<i>.67</i> (.09)	<i>.66</i> (.11)	<i>.66</i> (.10)
SK-Tech	Batch A	11	<i>.76</i> (.05)	<i>.77</i> (.05)	<i>.77</i> (.05)	<i>.76</i> (.05)	<i>.76</i> (.05)	<i>.76</i> (.05)
	All MOS	22	<i>.75</i> (.06)	<i>.75</i> (.06)	<i>.75</i> (.05)	<i>.75</i> (.06)	<i>.75</i> (.07)	<i>.74</i> (.07)
SK-Total	Batch A	11	<i>.78</i> (.03)	<i>.79</i> (.03)	<i>.79</i> (.03)	<i>.78</i> (.03)	<i>.79</i> (.03)	<i>.78</i> (.04)
	All MOS	22	<i>.77</i> (.05)	<i>.77</i> (.05)	<i>.77</i> (.05)	<i>.77</i> (.05)	<i>.77</i> (.06)	<i>.76</i> (.06)

Note: Corrected for range restriction and adjusted for shrinkage (Rozeboom formula B). Numbers in parentheses are standard deviations. Numbers in brackets are the numbers of predictor scores entering prediction equations. Multiple R_s for ASVAB Factors alone are in italics. Underlined numbers denote multiple R_s greater than for ASVAB Factors alone. Decimals omitted.

^a ETS = Effort and Technical Skill; MPD = Maintaining Personal Discipline; PFB = Physical Fitness and Military Bearing; LEAD = Leadership Potential; SK = School Knowledge.

^b Number of MOS for which validities were computed.

Further analysis showed that the average multiple correlations for the four different sets of ASVAB scores differed only slightly in validity, except that the peer ratings of Physical Fitness (PFB) were better predicted by the nine subtests and the four factors. However, the school knowledge test scores were predicted somewhat better (about three to five points) by the ASVAB subtests and factors than by the AFQT or Aptitude Area composites.

Both ABLE and AVOICE predicted the knowledge-based scores quite well. The largest incremental validities were for ABLE over ASVAB when predicting Personal Discipline, Fitness and Bearing, and Leadership.

Finally, there were virtually no differences in validities for the three alternative sets of ABLE scores although the ABLE-114 validities were consistently slightly higher.

Development of Basic Scores for the Longitudinal Validation (LVI) Performance Measures

In 1988 and 1989, first-tour criterion measures were administered to the Longitudinal Validation sample (LVI). This data collection was conducted concurrently with the administration of second-tour criterion measures to the Concurrent Validation sample (CVII). Before the LVI performance model development and subsequent validation analyses could begin, it was necessary to derive basic scores for each of the individual first-tour job performance measures. Dealing with all the individual scores from each task test, each rating scale, and each administrative index was simply not feasible or desirable. There were too many, and the reliabilities of the individual items or scales preserved too much measurement error with very little gain in total information. Consequently, the full array of scale scores was aggregated into a smaller set of basic scores for each measure.

Table 1.5 lists the individual measures that were administered.

Differences Between CVI and LVI Performance Measures

The 3-year time period between CVI and LVI raised the issue that for the job knowledge and hands-on measures, equipment and/or procedural changes would require test revisions, and changes in MOS responsibilities had the potential of making some tasks obsolete.

Project staff identified relevant changes so that the appropriate revisions could be made. In a few cases where an entire task was obsolete, the task was dropped without replacement. In many cases, revisions were simply a matter of replacing outdated terminology. Updated criterion measures were forwarded to the MOS proponents for a currency review and additional revisions were made on the basis of this review.

While there was considerable interest in keeping the Combat Performance Prediction Scales, project staff and the Scientific Advisory Group agreed that the version used in CVI was too lengthy. New scales were field tested in conjunction with the second-tour criterion measure field tests. The decision was made to retain the original summated scale format, but the total number of items was reduced from 40 to 19.

Table 1.5

Measures Administered to Soldiers in LVI Sample

MOS in

Batch A: Background Information Form
Job Knowledge Tests
Hands-On Tests
Army-Wide Rating Scales
MOS-Specific Rating Scales
Combat Performance Prediction Scales (males only)
Personnel File Form
Army Job Satisfaction Questionnaire
Job History Questionnaire
Physical Requirements Survey

MOS in

Batch Z: Background Information Form
School Knowledge Test
Army-Wide Rating Scales
Combat Performance Prediction Scales (males only)
Personnel File Form
Army Job Satisfaction Questionnaire
Physical Requirements Survey

Note. Rating scale data were collected from both supervisors and peers. The Physical Requirements Survey is not a Career Force or Project A measure.

The self-report form for gathering information on administrative records was updated by reviewing its contents with officers and NCOs representing the Army Personnel Command (PERSCOM). The form was altered to allow soldiers to report an M19 qualification in the event that an M16 qualification was not applicable. Also, three awards were dropped per guidance from PERSCOM.

Task-level ratings were deleted from the array of Batch A first-tour criterion measures used in CVI. The Army-wide and MOS-specific rating scales were retained in their original form.

The development of the basic scores for each measure was based on the performance data collected from individuals in the Batch A and Batch Z MOS that were included in the administration of first-tour criterion measures in 1988 and 1989. The Batch A MOS were the same as those studied in the Concurrent Validation, except for the addition of 19K (M1 Armor Crewman).

As in CVI, the Batch A MOS differed from the Batch Z MOS in the comprehensiveness of the MOS-specific criterion measures that were available for administration. MOS-specific rating scales, hands-on tests, and job knowledge tests were administered to Batch A soldiers. The only MOS-specific measure available for administration to the Batch Z soldiers was the school

knowledge test that had been developed for administration at the end of training. The school knowledge test was administered to the Batch Z examinees as a surrogate for a job knowledge test.

Score Development for Administrative Indices

Five scores were computed from the LVI Personnel File Form: (a) awards and memoranda/certificates of achievement, (b) Physical Readiness Test, (c) M16 qualification, (d) Articles 15 and flag actions (disciplinary actions), and (e) promotion rate.

The first score was a composite of (a) awards and decorations; (b) memoranda of appreciation, commendation, or achievement; and (c) certificates of appreciation, commendation, or achievement. The last score, promotion rate, was derived from data available in the Army's computerized personnel records. It was the residual of pay grade regressed on time in service, adjusted by MOS.

A Basic Score for the Combat Performance Prediction Ratings

Principal components analyses of the LVI/CVII Combat Scale data indicated the presence of two factors. The second factor, however, was defined by the three negatively worded items. Given that the second factor was probably not substantively distinct from the first, the calculation of a single total score (with the negatively worded items reverse-scored) for the Combat Scale ratings appeared appropriate. Note that the two factors found in the LVI/CVII data were essentially the same as those found in CVI and used to derive the two Combat Scale scores at that time.

Development of Basic Scores for the First-Tour Performance Rating Scales

The Army-wide rating scales include 12 dimensions of soldier effectiveness that are important regardless of soldiers' MOS. MOS-specific rating scales were developed for each of the nine Batch A MOS, and these rating scales include between 7 and 13 dimensions of MOS-specific performance.

Principal factor analyses with varimax rotation were conducted on the Army-wide ratings (across all MOS), for supervisor and peer ratings separately and pooled together. The pooled ratings were computed by averaging the mean peer rating and one supervisor rating for those soldiers who had at least one peer rating and one supervisor rating. Because previous analyses (using the CVI sample) showed that a single factor was sufficient to account for the majority of the variance in the MOS-specific ratings, factor analyses were not conducted for the MOS-specific rating data.

Table 1.6 shows the three-factor, rotated solutions for the pooled peer/supervisor ratings. These data demonstrate the remarkable similarity of the rotated factor structures for the CVI and LVI samples. It is worth noting that these same three factors were also obtained in factor analyses of performance rating data for a sample of 950 second-tour soldiers, which was collected using a set of rating scales very similar to those used to collect the present data (Campbell & Zook, 1990).

Table 1.6

Comparison of LVI and CVI Army-Wide Factor Analysis^a Results: Pooled Peer/Supervisor Ratings^b

Dimension	Factor Loadings (LVI/CVI)		
	1	2	3
Technical Knowledge/Skill	<u>.67/.71</u>	.30/.28	.38/.30
Leadership	<u>.65/.69</u>	.34/.30	.44/.37
Effort	<u>.66/.69</u>	.47/.43	.32/.26
Self-Development	<u>.52/.57</u>	.42/.33	.46/.38
Maintaining Equipment	<u>.50/.54</u>	.41/.34	.41/.35
Following Regulations	.39/.41	<u>.73/.69</u>	.31/.30
Self-Control	.19/.22	<u>.65/.63</u>	.20/.20
Integrity	.44/.50	<u>.66/.59</u>	.30/.28
Military Bearing	.31/.32	.35/.32	<u>.57/.57</u>
Physical Fitness	.24/.21	.16/.15	<u>.49/.49</u>
Percent Common Variance	37.7/44.9	36.6/32.7	25.6/22.4

Note. Sample size is 7,919 for LVI and 8,642 for CVI.

^a Principal factor analysis, varimax rotation.

^b Computed by averaging the mean peer rating and the mean supervisor rating.

For both the Army-wide and MOS-specific rating scales, the mean, variability, and reliability of the peer, supervisor, and pooled peer/supervisor ratings appear quite acceptable and are comparable to what was found in the CVI research. Factor analyses of the Army-wide ratings showed that the three-factor CVI solution was replicated in the present data. Accordingly, the three composites shown in Table 1.7, along with the overall effectiveness rating, were used as the basic scores for the Army-wide rating data.

Table 1.7

Composition and Definition of LVI Army-Wide Rating Composites

Factor Name and Definition	Percent Common Variance Accounted For by Relevant Factor ^a (LVI/CVI)	Dimensions Included
<p>1. Technical Skills and Job Effort:</p> <p>Exerting effort over the full range of job tasks; engaging in training or other development activities to increase proficiency; persevering under dangerous or adverse conditions; and demonstrating leadership and support toward peers.</p>	37.8/44.9	<p>Technical Knowledge/ Skill Leadership Effort Self-Development Maintaining Equipment</p>
<p>2. Personal Discipline:</p> <p>Adhering to Army rules and regulations; exercising self-control; demonstrating integrity in day-to-day behavior; and not causing disciplinary problems.</p>	36.6/32.7	<p>Following Regulations Self-Control Integrity</p>
<p>3. Physical Fitness/Military Bearing:</p> <p>Maintaining an appropriate military appearance and bearing and staying in good physical condition.</p>	25.6/22.4	<p>Military Bearing Physical Fitness</p>

^a Factor analysis of pooled peer/supervisor ratings.

Development of Basic Scores for Hands-On Performance and Job Knowledge Measures

As the first step in replicating the CVI procedures for constructing the basic scores, tasks were clustered into Functional Categories as described in the Project A annual report for 1986 (Campbell, 1988).

Following the procedures developed with the CVI data, tasks were also sorted into six higher level groups referred to as Task Factors (Communication, Vehicles, Basic Techniques, Identify Targets, Technical, and Safety/Survival (CVBITS)). Tasks were also combined into just two groups: General (i.e., Army-wide) and MOS-specific.

In general, the grouping schemes are hierarchical: Tasks (the lowest level) are placed in Functional Categories, the Functional Categories (level two) are aggregated to form the six Task Factors (level three), and Task Factors are then aggregated to form the two Task Constructs (level four), as diagrammed in Figure 1.10.

For the LVI data, confirmatory factor analyses were conducted to assess the fit of alternative levels of score aggregation. These analyses served two purposes: They were used to assess the relative merits of each model and to corroborate the CVI decision to use the six task factor scores (CVBITS). The analysis required the computation of separate tests of goodness of fit for hands-on and job knowledge test data, for each of the 10 MOS, on each of three competing models. The three models tested were: a one-factor model, postulating the existence of a single factor in the data; a two-factor model, proposing the Basic and the Technical Task Constructs; and a three-to-six-factor model (the number of factors varying among MOS and test method), using the Task Factors. Examination of the results from LVI argues for the retention of the six Task Factor scores for both the Hands-On and Job Knowledge measures.

Final Array of LVI Basic Performance Scores

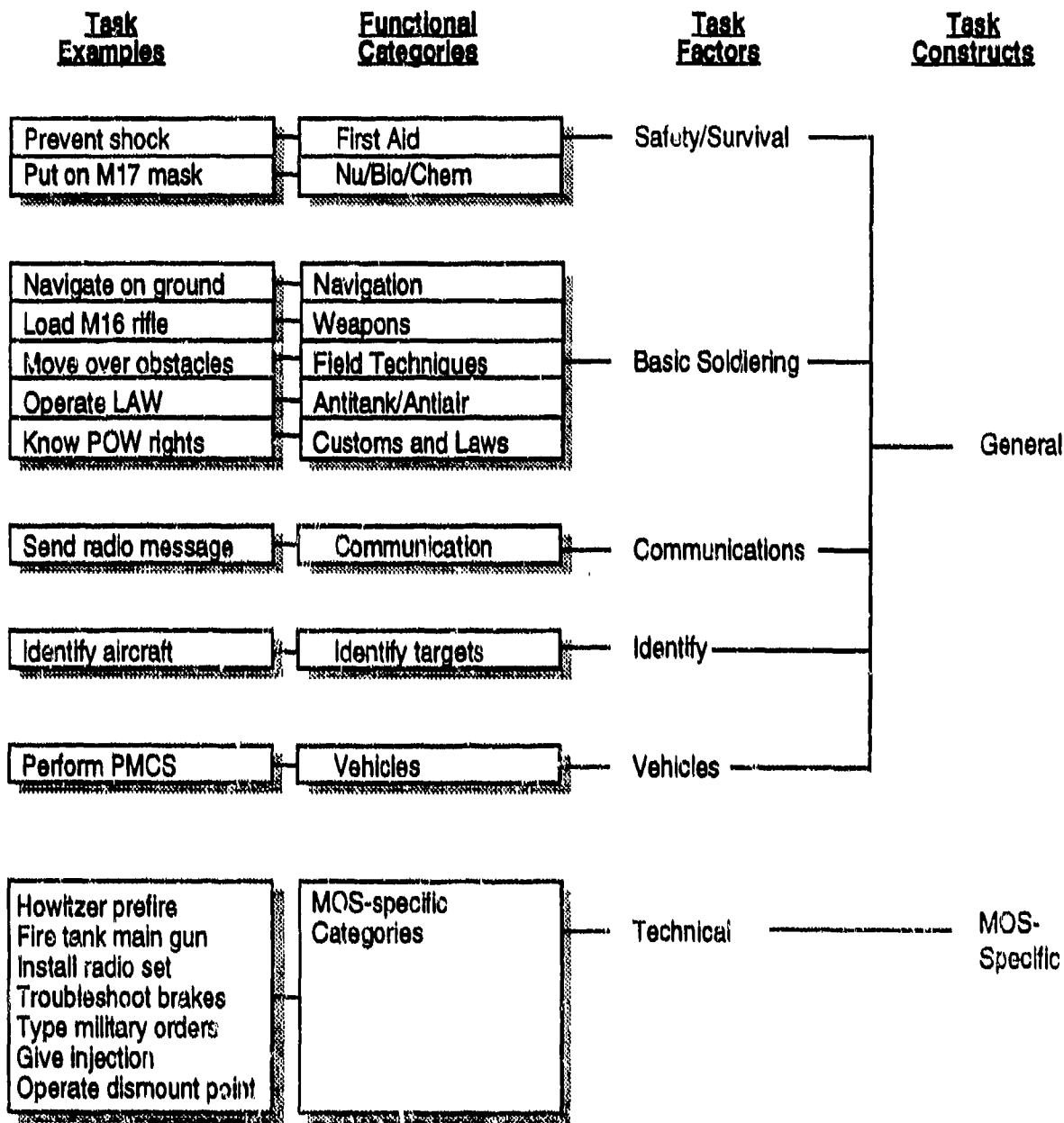
A summary list of the basic performance scores produced by the analyses summarized above is given in Figure 1.11. These are the scores that were put through the final editing and score imputation procedures for the LVI data file. The scores that formed the basis for the confirmatory tests of the LVI model of first-tour job performance were also drawn from this array.

The LVI Data File: Final Data Editing and Score Imputation

The Longitudinal Validation First-Tour (LVI) data were collected from 11,266 soldiers in 21 MOS. There were 6,815 Batch A examinees and 4,451 Batch Z examinees. Extensive efforts were made to collect complete information from each examinee for all instruments. However, as with all data collection exercises, circumstances precluded complete success. The final counts of soldiers for whom data were analyzed for each instrument are given in Tables 1.8 and 1.9 for Batch A and Batch Z MOS, respectively.

Data for each performance measure were processed individually. After processing was completed for these individual measures, they were combined so that all LVI data for each examinee were included in a single file. The data were combined separately by MOS. When the data were combined, basic scores were calculated for the individual performance measures. Table 1.10 shows the amount of missing data for the final set of basic criterion scores.

In addition to the performance data, missing Longitudinal Validation predictor data were also imputed. For a complete description of the editing process used on the predictor data, see the 1990 annual report. The bulk of the editing process was accomplished during FY90, but additional work was done during FY91. The amounts of missing data for each score on each paper-and-pencil and each computerized measure are shown in Tables 1.11 and 1.12.



Note. The Task Factors correspond to the six task groups known as CVBITS. The Task Constructs termed General and MOS-Specific refer to the same constructs that have previously been called Basic and Technical, or Common and Technical.

Figure 1.10. Hierarchical relationships among Functional Categories, Task Factors, and Task Constructs.

Hands-On Performance Test

1. Safety/survival performance score
2. General (common) task performance score
3. Communication performance score
4. Vehicles performance score
5. MOS-specific task performance score

Job Knowledge Test

6. Safety/survival knowledge score
7. General (common) task knowledge score
8. Communication knowledge score
9. Identify targets knowledge score
10. Vehicles knowledge score
11. MOS-specific task knowledge score

Army-Wide Rating Scales

12. Overall effectiveness rating
13. Technical skill and effort composite
14. Personal discipline composite
15. Physical fitness/military bearing composite

MOS-Specific Rating Scales

16. Overall MOS composite

Combat Performance Prediction Scales

17. Overall Combat Prediction scale composite (available for males only)

Personnel File Form

18. Awards and Certificates
19. Disciplinary Actions (Articles 15 and Flag Actions)
20. Physical Readiness
21. M16 Qualification
22. Promotion Rate

Figure 1.11. Summary list of LVI basic criterion scores.

Table 1.8

LVI Sample Sizes for Performance Measures for Batch A MOS

MOS	N	Hands-On	Job Knowledge	Army-Wide Ratings	MOS Ratings	Combat Ratings	Personnel File	Combined Criteria ^a
11B Infantryman	909	890	895	899	899	898	906	907
13B Cannon Crewmember	916	773	810	897	897	897	916	916
19E M60 Armor Crewman	249	243	248	241	241	241	249	249
19K M1 Armor Crewman	824	749	812	782	778	782	819	825
31C Single Channel Radio Operator	529	446	504	497	481	442	527	529
63B Light Wheel Vehicle Mechanic	752	624	723	728	719	666	750	752
71L Administrative Specialist	678	641	664	634	626	199	675	678
88M Motor Transport Operator	682	588	674	666	663	479	680	682
91A Medical Specialist	824	794	798	807	797	670	818	824
95B Military Police	452	444	446	451	450	366	452	452
Total	6,815	6,192	6,574	6,602	6,547	5,640	6,792	6,814

^a Combined Criteria include Hands-On, Job Knowledge, Army-Wide Ratings, MOS Ratings, and Personnel File Form.

Table 1.9

LVI Sample Sizes for Performance Measures for Batch Z MOS

MOS	N	Job Knowledge	Army-Wide Ratings	Combat Ratings	Personnel File
12B Combat Engineer	841	840	827	827	838
16S MANPADS Crewman	472	471	468	468	472
27E Tow/Dragon Repairer	90	90	89	84	90
29E Comm.-Electronics Radio Repairer	112	111	106	101	111
51B Carpentry/Masonry Specialist	213	212	193	190	212
54B NBC Specialist	499	498	492	462	498
55B Ammunition Specialist	279	279	269	243	279
67N Utility Helicopter Repairer	197	194	193	192	197
76Y Unit Supply Specialist	788	788	734	616	787
94B Food Service Specialist	832	932	818	717	931
96B Intelligence Analyst	128	128	122	103	128
Total	4,451	4,443	4,311	4,003	4,443

An imputation procedure known as PROC IMPUTE was developed that used existing data to estimate values for missing data. This procedure was also used in the CVI analyses (Wise, McHenry, & Young, 1986). The decision rules used in the CVI analyses were replicated in the LVI analyses as closely as possible.

PROC IMPUTE uses regression estimates to predict missing values. Each missing value is predicted from other values for the subject in question so that individual differences are retained. The regression coefficient and intercept vary from item to item so that differences in item difficulty are also reflected in the predicted values. PROC IMPUTE also adds a random variable with variance equal to the error of estimate for predicting the missing value.

The results of the imputation were examined at two levels. First, after each PROC IMPUTE run, the program output was inspected. Second, the pre-imputed and the post-imputed data sets were compared for each MOS (a) after the hands-on score level imputation, and (b) after the criterion construct level imputation.

The means and variances of the pre- and post-imputation results for the hands-on data for each MOS were found to be virtually identical. Imputation also made virtually no difference in the magnitude of the intercorrelations among the criterion scores that were used to create the performance factor scores in the validation analyses. These results are similar to those obtained earlier from the CVI imputation (Wise et al., 1986).

Table 1.10

LVI Combined Criteria Data: Percentage of Missing Data for Basic Scores by MOS

Criteria	11B	13B	19E	19K	31C	63B	71L	88M	91A	95B
Hands-On - Task Factors										
C - Communications	1.87	15.61	2.41	9.21	15.69	--	--	--	--	--
V - Vehicles	--	--	--	--	15.69	17.02	--	13.78	--	1.77
B - Basic Soldiering	1.87	15.61	2.41	9.21	15.69	17.02	5.46	13.78	3.64	1.77
I - Identify Targets	--	--	--	--	--	--	--	--	--	--
T - Technical	--	15.61	2.41	9.21	15.59	17.02	5.46	--	3.64	1.77
S - Safety/Survival	1.87	15.61	2.41	9.21	15.69	17.02	5.46	13.78	3.64	1.77
Job Knowledge - Task Factors										
C - Communications	2.65	12.01	.40	1.94	6.62	--	--	--	--	1.55
V - Vehicles	--	--	--	--	6.62	5.05	--	1.91	4.98	1.55
B - Basic Soldiering	2.65	12.31	.40	1.94	6.62	5.05	2.36	1.91	4.98	1.55
I - Identify Targets	2.65	12.01	.40	1.94	6.62	--	--	1.91	4.98	1.55
T - Technical	--	12.01	.40	1.94	6.62	5.05	2.36	--	4.98	1.55
S - Safety/Survival	2.65	12.01	.40	1.94	6.62	5.05	2.36	1.91	4.98	1.55
Army-Wide Ratings										
Overall Effectiveness	1.10	2.95	3.21	5.33	6.99	3.32	7.67	2.79	2.31	22
Technical Skill and Effort	.88	2.07	3.21	5.21	6.05	3.19	6.93	2.35	2.06	22
Personal Discipline	.88	2.07	3.21	5.21	6.05	3.19	6.95	2.35	2.06	22
Physical Fitness/Bearing	.88	2.07	3.21	5.21	6.05	3.19	6.93	2.35	2.06	22
MOS Ratings										
MOS Composite Rating	1.32	5.35	3.21	6.67	9.83	4.65	9.73	4.55	6.43	1.55
Personnel File Form										
Awards and Certificates	2.43	3.60	2.01	3.76	4.91	3.19	2.65	2.79	2.91	1.77
Articles 15 and Flag Actions	1.21	1.53	.00	1.82	1.51	1.06	1.18	.73	1.94	.44
Physical Readiness Score	4.63	5.46	3.21	5.21	9.45	11.44	9.00	9.09	6.55	5.31
M16 Qualification	2.65	4.04	29.32	18.30	2.65	3.19	1.77	2.93	3.88	3.98
Promotion Rate	1.76	1.86	.80	4.00	5.10	4.79	5.01	3.96	2.67	0.88

Note. -- indicates that the particular score was not calculated for that MOS.

Table 1.11

LVI Predictor Data: Amount of Missing Data for Paper-and-Pencil Scale Scores

Score	Not Missing	Missing
Assembling Objects - Number Correct	49,042	366
Map - Number Correct	49,047	361
Maze - Number Correct	49,052	356
Object Rotation - Number Correct	49,103	305
Orientation - Number Correct	49,072	336
Reasoning - Number Correct	49,103	305
JOB Scale 1 - Pride	46,525	2,883
JOB Scale 2 - Job Security/Comfort	46,634	2,774
JOB Scale 3 - Serving Others	46,295	3,113
JOB Scale 4 - Job Autonomy	46,037	3,371
JOB Scale 5 - Routine	45,975	3,433
JOB Scale 6 - Ambition	46,058	3,350
ABLE Scale 1 - Emotional Stability	44,264	5,144
ABLE Scale 2 - Self-Esteem	44,247	5,161
ABLE Scale 3 - Cooperativeness	44,258	5,150
ABLE Scale 4 - Conscientiousness	44,199	5,209
ABLE Scale 5 - Nondelinquency	44,228	5,180
ABLE Scale 6 - Traditional Values	44,190	5,218
ABLE Scale 7 - Work Orientation	44,260	5,148
ABLE Scale 8 - Internal Control	44,254	5,154
ABLE Scale 9 - Energy Level	44,217	5,191
ABLE Scale 10 - Dominance	44,246	5,162
ABLE Scale 11 - Physical Condition	44,264	5,144
AVOICE Scale 1 - Clerical/Administrative	45,477	3,931
AVOICE Scale 2 - Mechanics	45,941	3,467
AVOICE Scale 3 - Heavy Construction	45,851	3,557
AVOICE Scale 4 - Electronics	45,922	3,486
AVOICE Scale 5 - Combat	45,939	3,469
AVOICE Scale 6 - Medical Services	45,545	3,863
AVOICE Scale 7 - Rugged Individualism	45,944	3,464
AVOICE Scale 8 - Leadership/Guidance	45,508	3,900
AVOICE Scale 9 - Law Enforcement	45,958	3,450
AVOICE Scale 10 - Food Service Professional	45,916	3,492
AVOICE Scale 11 - Firearms Enthusiast	45,942	3,466
AVOICE Scale 12 - Science/Chemical	45,970	3,438
AVOICE Scale 13 - Drafting	45,976	3,432
AVOICE Scale 14 - Audiographics	45,452	3,956
AVOICE Scale 15 - Aesthetics	45,279	4,129
AVOICE Scale 16 - Computers	45,554	3,854
AVOICE Scale 17 - Food Service Employee	45,965	3,443
AVOICE Scale 18 - Mathematics	45,691	3,717
AVOICE Scale 19 - Electronic Communications	45,602	3,806
AVOICE Scale 20 - Warehousing/Shipping	45,963	3,445
AVOICE Scale 21 - Fire Protection	45,972	3,436
AVOICE Scale 22 - Vehicle Operator	45,971	3,437

Table 1.12

LVI Predictor Data: Amount of Missing Data for Computer-Administered Scale Scores

Score	Not Missing	Missing
Target Identification - Mean of Clipped Decision Time	38,401	513
Target Identification - Proportion Correct	38,404	510
Number Memory - Mean of Clipped Operation Means	38,324	590
Number Memory - Proportion Correct	38,353	561
Target Track 1 - Mean Log (Distance+1)	38,825	89
Target Track 2 - Mean Log (Distance+1)	38,793	121
Cannon Shoot - Mean Absolute Time Discrepancy	38,603	311
Target Shoot - Mean Log (Distance+1)	37,477	1,437
Mean of Median Movement Times across 5 tests	37,863	1,051
Simple Reaction Time - Median Decision Time	38,747	167
Simple Reaction Time - Proportion Correct	38,747	167
Choice Reaction Time - Median Decision Time	38,856	58
Choice Reaction Time - Proportion Correct	38,856	58
Perceptual Speed/Accuracy - Mean of Clipped Decision Time	38,703	211
Perceptual Speed/Accuracy - Proportion Correct	38,734	180
Short-Term Memory - Mean of Clipped Decision Time	38,483	431
Short-Term Memory - Proportion Correct	38,490	424

Development of the LVI First-Tour Performance Model

A latent factor model of first-tour performance, developed using data from the Project A Concurrent Validation (CVI) sample, has been described by J. P. Campbell, McHenry, and Wise (1990). This model included the now familiar five performance factors--Core Technical Proficiency (CTP), General Soldiering Proficiency (GSP), Effort and Leadership (ELS), Maintaining Personal Discipline (MPD), and Physical Fitness and Military Bearing (PFB)--and two measurement method factors, a Ratings method factor and a Paper-and-Pencil Test method factor. During year two, the CVI model was subjected to a confirmatory analysis, using first-tour performance data collected from the Longitudinal Validation (LVI) sample. Additionally, comparative analyses aimed at evaluating more parsimonious models of first-tour performance were carried out.

An earlier section summarized how each of the major sets of performance measures was reduced from a large number of item, task, or individual scale scores to a smaller set of factor or category scores. The results of this first level of aggregation have been referred to as the "basic" array of criterion scores, summarized in Figure 1.11. These included the scores that were used in the modeling analyses described below.

Altogether, the LVI first-tour performance measures were reduced to 20 basic scores. However, because MOS differ in their task content, not all 20 variables were scored in each MOS, and there was some slight variation in the number of variables used in the subsequent analyses.

To test the fit of the different models to the LVI data, confirmatory factor-analytic techniques were applied to each MOS individually, using LISREL 7 (Jöreskog & Sörbom, 1989a). The first alternative five-factor model was developed using CVI data. After the fit of the five-factor model was assessed in each MOS, four reduced models (all nested within the five-factor model) were examined. Finally, as had been done in the original CVI analyses, the five-factor model was applied to the Batch A MOS simultaneously (using LISREL's multigroups option). The fit statistics (e.g., root mean-square residuals [RMSRs]) of the five-factor model for each MOS in the LVI and CVI samples were very similar. In fact, for three of the MOS (11B, 13B, and 71L), the RMSRs for the LVI data were smaller than those for the CVI data. These results indicate that the model developed using the CVI data does fit the LVI data quite well.

Four reduced models were also examined using the LVI data. For the four-factor model, the Core Technical Proficiency and General Soldiering Proficiency performance factors were collapsed into a single "can do" performance factor. The three-factor model retained the "can do" performance factor of the four-factor model, but also collapsed the Effort and Leadership and Maintaining Personal Discipline performance factors into a "will do" performance factor. For the two-factor model, the "can do" performance factor was retained; however, the Physical Fitness and Military Bearing performance factor became part of the "will do" performance factor. Finally, for the one-factor model, the "can do" and "will do" performance factors, or equivalently, the five original performance factors, were collapsed into a single performance factor.

The chi-square statistics and RMSRs, respectively, for the four reduced models, as well as for the five-factor model, indicate that the four- and five-factor models fit the LVI data well, while the one-, two-, and three-factor models fit less well. The results also indicated that the parameter estimates for the five-factor model were generally similar across the 10 MOS. The final step was to determine whether the variation in some of these parameters could be attributed to sampling variation. To do this (as described earlier), the following were specified to be invariant across jobs: (a) the correlations among performance factors, (b) the loadings of all the Army-wide measures on the performance factors and on the rating method factor, (c) the loadings of the MOS-specific score on the rating method factor, and (d) the uniqueness coefficients for the Army-wide measures.

The results indicated that the fit of the five-factor model is not as good when the parameters listed above are constrained to be equal across the 10 jobs. Still, the root mean-square residuals associated with the across-MOS model are not substantially greater than those for the within-job analyses. (The average RMSR for the across-MOS model is .0676; the average for the within-MOS models is .0585.)

To create criterion construct scores for use in validation analyses, the scoring procedures were based on the five-factor model. Although the four-factor model has the advantage of greater parsimony, the five-factor model offered the advantage of corresponding to the criterion constructs generated in the CVI validation analyses. Table 1.13 shows the mapping of the basic scores on the five performance factors. As with the CVI data, five residual scores, corresponding to the five criterion constructs, were also created.

The five "raw" criterion construct scores, the five residual criterion construct scores, the total rating and job knowledge scores, and the total score derived from the hands-on test were used to generate a 13 x 13 matrix of criterion intercorrelations for each MOS in Batch A. The averages of these correlations are reported in Table 1.14. These results are very similar to the correlations that were reported by Campbell et al. (1990) for the CVI sample.

Basic Validation Results for the LVI Sample

The LVI validation results were based on two different sample editing strategies. The first required complete data for all predictor composites, as well as for the ASVAB, and for each performance factor; this sample is referred to as the "listwise deletion" sample. In the alternative strategy, called setwise deletion, a separate validation sample was identified for each set of predictors in the Experimental Battery.

The number of soldiers with complete predictor and criterion data in each MOS is reported in Table 1.15 for both the CVI and LVI data sets.

The analysis procedure consisted of the following major steps:

- A) Using the listwise deletion sample, multiple correlations between each set of predictor scores and the five substantive factor scores, their five residual factor scores, the two method factor scores, and the total scores from the hands-on and job knowledge tests were computed separately by MOS and then averaged.
- B) Using the listwise deletion sample, incremental validities for each set of Experimental Battery predictors (e.g., AVOICE composites or computer composites) over the four ASVAB factor composites were computed against the same criteria used to compute the validities in Step A. Once again, the results were computed separately by MOS and then averaged.

Table 1.13

Mapping of LVII Performance Measures Onto Latent Performance Factors

Criterion Score*	Performance Factors					Method Factors	
	Core Technical Proficiency	General Soldiering	Maintaining Effort and Leadership	Personal Discipline	Physical Fitness/Military Bearing	Written Knowledge Tests	Rating Scales
HO Technical	X						
HO Communication		X					
HO Vehicles		X					
HO General Soldier		X					
HO Safety/Survival		X					
JK Technical	X					X	
JK Communication		X				X	
JK Vehicles		X				X	
JK General Soldier		X				X	
JK ID Threat/Target		X				X	
JK Safety/Survival		X				X	
AMB Skill/Effort Composite			X				X
AMB Discipline Composite				X			X
AMB Fitness Composite					X		X
AMB Overall Composite			X				X
MOS Rating Composite			X				
PFF Awards/Certificates			X				
PFF Physical Readiness					X		
PFF Articles 15/Flags				X			
PFF Promotion Rate				X			

* AMB = Army-Wide Rating Scales; HO = Hands-On; JK = Job Knowledge; PFF = Personnel File Form.

Table 1.14

Mean Intercorrelations Among 13 Summary Criterion Scores for the Batch A MOS in the LVI Sample

Summary Criterion Score ^a	CTP Raw	GSP Raw	ELS Raw	MPD Raw	PFB Raw	CTP Res	GSP Res	ELS Res	MPD Res	PFB Res	PRT	HOT	JKT
CTP (raw)	1.00												
GSP (raw)	.57	1.00											
ELS (raw)	.25	.26	1.00										
MPD (raw)	.16	.18	.58	1.00									
PFB (raw)	.06	.06	.48	.36	1.00								
CTP (residual)	.88	.41	.30	.20	.07	1.00							
GSP (residual)	.40	.88	.32	.23	.06	.45	1.00						
ELS (residual)	.41	.42	.70	.43	.26	.40	.42	1.00					
MPD (residual)	.20	.22	.28	.88	.17	.20	.23	.46	1.00				
PFB (residual)	.07	.07	.20	.21	.90	.04	.03	.29	.21	1.00			
Perf. Rating Total	.22	.24	.88	.72	.58	.27	.28	.40	.35	.24	1.00		
Hands-On Total	.72	.76	.26	.15	.08	.81	.85	.41	.18	.09	.23	1.00	
Job Knowledge Total	.74	.80	.25	.19	.04	.40	.46	.40	.23	.04	.22	.47	1.00

^a CTP = Core Technical Proficiency; GSP = General Soldiering Proficiency; ELS = Effort and Leadership; MPD = Maintaining Personal Discipline; PFB = Physical Fitness and Military Bearing.

Table 1.15

Soldiers in CVI and LVI Data Sets With Complete Predictor and First-Tour
Criterion Data by MOS

MOS		CVI	LVI (Listwise Deletion Sample)
11B	Infantryman	491	235
13B	Cannon Crewmember	464	553
19E ^a	M60 Armor Crewman	394	73
19K	M1 Armor Crewman	---	446
31C	Single Channel Radio Operator	289	172
63B	Light-Wheel Vehicle Mechanic	478	406
71L	Administrative Specialist	427	252
88M	Motor Transport Operator	507	221
91A	Medical Specialist	392	535
95B	Military Police	597	270
Total		4,039	3,163

^a MOS 19E not included in LVI validity analyses.

- C) Using the setwise deletion samples, multiple correlations and incremental validities (over the four ASVAB factor composites) between each set of Experimental Battery predictors and the criteria used in the first two steps were computed separately by MOS and then averaged. All results to this point were corrected for range restriction and adjusted for shrinkage using the Rozeboom formula.
- D) Finally, once again using the listwise deletion sample, multiple correlations and incremental validities (over the four ASVAB factors) were computed for each set of predictors in the Experimental Battery, this time adjusting the results for shrinkage with the Claudy (1978) instead of the Rozeboom formula. This step was conducted to allow comparisons between the first-tour validity results associated with the longitudinal sample and those that had been reported for the concurrent sample (for which only the Claudy formula was used, e.g., McHenry, Hough, Toquam, Hanson, & Ashworth, 1990).

Multiple Correlations and Incremental Validities Based on Listwise Deletion Samples

Multiple correlations for the four ASVAB factor composites, the single spatial composite, the eight computer composites, the three JOB composites, the seven ABLE composites, and the eight AVOICE composites are reported in Table 1.16.

Incremental validity results for the Experimental Battery predictors over the ASVAB factors are reported in Table 1.17. The results indicate that the spatial composite added slightly to the prediction of the raw and residual Core Technical and General Soldiering performance factors, as well as to the written method factor and the hands-on and job knowledge total scores. They also show that the seven ABLE composites contributed substantially to the prediction of the raw and residual Personal Discipline and Physical Fitness performance factors.

Multiple Correlations and Incremental Validities Based on the Setwise Deletion Samples

Multiple correlations for the spatial composite, the eight computer composites, the three JOB composites, the seven ABLE composites, and the eight AVOICE composites based on the setwise deletion samples described above are reported in Table 1.18. These multiple correlations were very similar to those computed with the listwise sample. However, there was a consistent difference between the two sets of results; specifically, the multiple correlations based on the setwise samples were generally one to three validity points higher.

Incremental validity results associated with the setwise deletion samples can be found in Table 1.19. The incremental validity results based on the setwise samples were practically identical to those based on the listwise sample. Again, the primary difference between the two sets of results was that the level of validities was sometimes one or two points lower for the listwise sample than for the setwise samples.

Comparison of Validity Research in LVI and CVI Samples

The final set of results concern the comparison between the validity estimates associated with the longitudinal data (i.e., LVI) and those reported for the concurrent validation data (CVI). Table 1.20 reports the multiple correlations for the ASVAB factors and each set of experimental predictors as computed for the listwise sample in both data sets.

The results in Table 1.20 demonstrate that the patterns and levels of validities are very similar across the two sets of analyses. Still, there are some differences worth pointing out. Specifically, in comparison to the results of the CVI analyses: (a) The LVI validities of the "cognitive" predictors (i.e., ASVAB, spatial, computer) for predicting the "will do" performance factors (ELS, MPD, and PFB) are higher; (b) the LVI validities of the ABLE composites for predicting the "will do" performance factors are somewhat lower; and (c) the LVI validities of the AVOICE composites for predicting the "can do" performance factors (CTP and GSP) are higher.

Table 1.16

Mean of Multiple Correlations Computed Within-Job for LVI Listwise Deletion Sampler for ASVAB Factors, Spatial, Computer, JOB, ABLE Composites, and AVOICE

Criterion ^a	No. of MOS ^b	ASVAB Factors [4]	Spatial [1]	Computer [8]	JOB [3]	ABLE Comp. [7]	AVOICE [8]
CTP (Raw)	9	62 (13)	57 (11)	47 (16)	29 (13)	21 (09)	38 (08)
GSP (Raw)	8	66 (07)	64 (06)	55 (08)	29 (13)	23 (14)	37 (07)
ELS (Raw)	9	37 (12)	32 (08)	29 (15)	18 (14)	13 (11)	17 (15)
MPD (Raw)	9	17 (13)	14 (11)	10 (16)	06 (13)	14 (11)	05 (10)
PFB (Raw)	9	16 (06)	10 (04)	07 (07)	06 (06)	27 (07)	05 (09)
CTP (Res)	9	46 (17)	42 (15)	29 (22)	17 (12)	08 (11)	28 (12)
GSP (Res)	8	51 (10)	51 (08)	41 (10)	18 (11)	12 (12)	26 (09)
ELS (Res)	9	46 (18)	41 (13)	37 (20)	23 (15)	21 (15)	24 (16)
MPD (Res)	9	18 (13)	14 (12)	08 (16)	07 (11)	13 (11)	06 (10)
PFB (Res)	9	20 (10)	12 (08)	09 (11)	07 (06)	28 (10)	09 (11)
Written Ratings	9	54 (13)	49 (12)	43 (18)	29 (16)	23 (12)	29 (14)
	9	12 (09)	09 (07)	07 (09)	06 (09)	03 (05)	02 (07)
HO-Total	9	50 (14)	48 (11)	38 (15)	18 (13)	11 (11)	28 (09)
JK-Total	9	71 (08)	65 (07)	58 (10)	36 (14)	31 (08)	41 (08)

Note. Corrected for range restriction, and adjusted for shrinkage (Rozeboom formula 8). Numbers in parentheses are standard deviations. Numbers in brackets are the numbers of predictor scores entering prediction equations. Decimals omitted.

^a CTP = Core Technical Proficiency; GSP = General Soldiering Proficiency; ELS = Effort and Leadership; MPD = Maintaining Personal Discipline; PFB = Physical Fitness and Military Bearing; HO = Hands-On; JK = Job Knowledge.

^b Number of MOS for which validities were computed.

Table 1.17

Mean of Incremental Correlations Over ASVAB Factors Computed Within-Job for LVI Listwise Deletion Samples for Spatial, Computer, JOB, ABLE Composites, and AVOICE

Criterion	No. of MOS ^a	ASVAB Factors (A4) [4]	A4+ Spatial [5]	A4+ Computer [12]	A4+ JOB [7]	A4+ ABLE Comp. [11]	A4+ AVOICE [12]
CTP (Raw)	9	62 (13)	<u>63</u> (13)	61 (14)	61 (13)	61 (13)	62 (13)
GSP (Raw)	8	66 (07)	<u>68</u> (07)	66 (07)	66 (07)	66 (07)	66 (07)
ELS (Raw)	9	37 (12)	<u>36</u> (13)	35 (13)	36 (13)	34 (17)	33 (16)
MPD (Raw)	9	17 (13)	16 (14)	16 (15)	14 (15)	<u>23</u> (14)	10 (15)
PFB (Raw)	9	16 (06)	13 (08)	09 (08)	<u>17</u> (08)	<u>30</u> (06)	12 (10)
CTP (Res)	9	46 (17)	<u>47</u> (17)	44 (18)	45 (18)	43 (19)	46 (19)
GSP (Res)	8	51 (10)	<u>53</u> (09)	51 (10)	50 (10)	50 (10)	50 (10)
ELS (Res)	9	46 (18)	<u>47</u> (18)	44 (21)	45 (21)	45 (22)	44 (21)
MPD (Res)	9	18 (13)	<u>15</u> (14)	15 (14)	14 (14)	<u>22</u> (14)	12 (13)
PFB (Res)	9	20 (10)	18 (12)	13 (11)	20 (11)	<u>34</u> (10)	18 (13)
Written Ratings	9	54 (13)	<u>55</u> (13)	51 (18)	54 (13)	54 (12)	52 (17)
	9	12 (09)	11 (08)	09 (10)	09 (10)	09 (08)	05 (08)
HO-Total	9	50 (14)	<u>52</u> (13)	49 (14)	49 (15)	48 (14)	49 (15)
JK-Total	9	71 (08)	<u>72</u> (08)	71 (09)	71 (08)	71 (08)	71 (08)

Note. Corrected for range restriction, and adjusted for shrinkage (Rozeboom formula 8). Numbers in parentheses are standard deviations. Numbers in brackets are the numbers of predictor scores entering prediction equations. Multiple Rs for ASVAB Factors alone are in italics. Underlined numbers denote multiple Rs greater than for ASVAB Factors alone. Decimals omitted.

^a Number of MOS for which validities were computed.

Table 1.18

Mean of Multiple Correlations Computed Within-Job for LVI Setwise Deletion Samples for Spatial, Computer, JOB, ABLE Composites, and AVOICE

Criterion	No. of MOS ^a	Spatial [1]	Computer [8]	JOB [3]	ABLE Composites [7]	AVOICE [8]
CTP (Raw)	9	58 (11)	49 (16)	31 (13)	21 (09)	39 (07)
GSP (Raw)	8	65 (06)	55 (08)	32 (13)	24 (14)	38 (07)
ELS (Raw)	9	33 (08)	30 (15)	19 (14)	12 (11)	20 (12)
MPD (Raw)	9	14 (11)	10 (16)	06 (13)	15 (11)	05 (11)
PFB (Raw)	9	08 (04)	13 (07)	07 (06)	28 (07)	09 (09)
CTP (Res)	9	43 (15)	31 (22)	17 (12)	10 (11)	29 (09)
GSP (Res)	8	51 (08)	40 (10)	21 (11)	14 (12)	28 (09)
ELS (Res)	9	41 (13)	36 (20)	24 (15)	21 (15)	26 (06)
MPD (Res)	9	13 (12)	10 (16)	06 (11)	15 (11)	07 (13)
PFB (Res)	9	11 (08)	10 (11)	09 (06)	30 (10)	12 (10)
Written Ratings	9	51 (11)	46 (16)	31 (17)	25 (11)	32 (15)
	9	09 (08)	09 (09)	07 (08)	04 (06)	03 (07)
HO-Total	9	50 (11)	38 (15)	20 (13)	13 (11)	30 (07)
JK-Total	9	66 (07)	60 (10)	38 (14)	30 (08)	43 (08)

Note. Corrected for range restriction and adjusted for shrinkage (Rozeboom formula 8). Numbers in parentheses are standard deviations. Numbers in brackets are the numbers of predictor scores entering prediction equations. Decimals omitted.

^a Number of MOS for which validities were computed.

Table 1.19

Mean of Incremental Correlations Over ASVAB Factors Computed Within-Job for LVI Setwise Deletion Samples for Spatial, Computer, JOB, ABLE Composites, and AVOICE

Criterion	No. of MOS ^a	ASVAB Factors (A4) + Spatial [5]	A4+ Computer [12]	A4+ JOB [7]	A4+ ABLE Composites [11]	A4+ AVOICE [12]
CTP (Raw)	9	<u>64</u> (10)	61 (11)	63 (11)	61 (12)	64 (11)
GSP (Raw)	8	<u>69</u> (06)	<u>66</u> (07)	67 (07)	66 (08)	66 (07)
ELS (Raw)	9	37 (10)	36 (14)	37 (11)	36 (13)	36 (11)
MPD (Raw)	9	15 (13)	15 (15)	12 (13)	<u>24</u> (13)	11 (14)
PFB (Raw)	9	15 (08)	17 (05)	<u>17</u> (07)	<u>32</u> (04)	15 (10)
CTP (Res)	9	<u>48</u> (12)	45 (14)	46 (14)	45 (14)	47 (14)
GSP (Res)	8	<u>54</u> (06)	50 (08)	51 (08)	50 (07)	50 (07)
ELS (Res)	9	47 (12)	43 (20)	46 (15)	46 (15)	46 (14)
MPD (Res)	9	14 (13)	13 (15)	13 (13)	<u>22</u> (12)	11 (14)
PFB (Res)	9	20 (11)	18 (11)	20 (10)	<u>36</u> (08)	21 (11)
Written Ratings	9	<u>57</u> (13)	53 (17)	58 (12)	55 (13)	54 (18)
	9	10 (09)	<u>11</u> (11)	11 (09)	<u>11</u> (07)	06 (09)
HO-Total	9	<u>53</u> (09)	49 (11)	50 (12)	49 (11)	50 (11)
JK-Total	9	<u>73</u> (08)	71 (09)	72 (08)	71 (09)	71 (09)

Note. Corrected for range restriction and adjusted for shrinkage (Rozeboom formula 8). Numbers in parentheses are standard deviations. Numbers in brackets are the numbers of predictor scores entering prediction equations. Underlined numbers denote multiple Rs greater than for ASVAB Factors alone. Decimals omitted.

^a Number of MOS for which validities were computed.

Table 1.20

Comparison of Mean Multiple Correlations Computed Within-Job for LVI and CVI Listwise Deletion Samples for ASVAB Factors, Spatial, Computer, JOB, ABLE Composites, and AVOICE

Criterion	No. of MOS ^a	ASVAB Factors		Spatial		Computer		JOB		ABLE Comp.		AVOICE	
		LV [4]	CV [4]	LV [1]	CV [1]	LV [8]	CV [6]	LV [3]	CV [3]	LV [7]	CV [4]	LV [8]	CV [6]
CTP (Raw)	9	63	63	57	56	50	53	31	29	27	26	41	35
GSP (Raw)	8	67	65	64	63	57	57	32	30	29	25	40	34
ELS (Raw)	9	39	31	32	25	34	26	22	19	20	33	25	24
MPD (Raw)	9	22	16	14	12	15	12	11	11	22	32	11	13
PFB (Raw)	9	21	20	10	10	17	11	12	11	31	37	15	12
CTP (Res)	9	48	47	42	37	35	37	20	21	18	22	33	28
GSP (Res)	8	53	49	51	48	44	41	22	22	19	21	31	26
ELS (Res)	9	48	46	41	41	40	38	25	27	26	31	29	32
MPD (Res)	9	23	19	14	15	14	13	12	10	21	28	13	15
PFB (Res)	9	24	21	12	11	17	14	11	10	32	35	16	14
Written Ratings	9	56	62	49	55	47	54	31	28	29	21	33	32
	9	16	15	09	07	17	08	10	08	09	18	09	09

Note. Corrected for range restriction and adjusted for shrinkage (Claudy formula). Numbers in brackets are the numbers of predictor scores entering prediction equations. Decimals omitted.

^a Number of MOS for which validities were computed.

Further Exploration of ELS and ABLE

As shown in the data reported above, the largest difference between the CVI and LVI validation results was in the prediction of the Effort and Leadership (ELS) performance factors with the ABLE basic scores. Corrected for restriction of range and for shrinkage, the validity of the four ABLE composite scores in CVI was .33 for ELS and the validity of the seven ABLE factor scores in LVI was .20. When cast against the variability in results across studies in the extant literature, such a difference may not seem all that large or very unusual. However, since the obtained results from CVI, CVII, and LVI have been so consistent, in terms of the expected convergent and divergent results, we subjected this particular difference to a series of additional analyses in an attempt to determine the reason for the discrepancy.

First, the discrepancy does not seem to arise from any general deterioration in the measurement properties of either the ABLE or the ELS composite in the LVI sample. For example, while the correlation of the ABLE with ELS and MPD went down, the ABLE's correlations with CTP and GSP went up slightly. Similarly, a decrease in the validity with which ELS was predicted was characteristic only of the ABLE. The validities of the cognitive measures, the JOB, and AVOICE for predicting ELS actually increased by varying amounts. Consequently, the decrease in validity seems to be specific to the ABLE/ELS correlation and, to a lesser extent, the ABLE/MPD correlation.

The followup analyses were also able to rule out two possible additional sources of the CVI/LVI validity differences. First, differences in the composition and number of ABLE basic scores from CVI to LVI did not account for the differences in patterns of validity. Second, differences in the composition of the Effort/Leadership factor score from CVI to LVI did not account for differences in validity.

Rather, the somewhat lower correlation of ABLE with Effort/Leadership in LVI seems due to the joint effects of two influences. First, the determinants of ELS scores seem to favor ability slightly more and motivation slightly less in LVI versus CVI, perhaps because their true score variances were different across the two cohorts. Second, the greater influence of the social desirability response tendency in LVI seems to produce more positive manifold (i.e., higher intercorrelations for the LVI ABLE basic scores), as contrasted with CVI. This could also lower the correlation of the regression-weighted ABLE composite with ELS, whereas it might not have the same effect with the Core Technical and General Soldiering factors.

Yet another component of the explanation is the negative correlation between the Social Desirability scale and AFQT. AFQT and Social Desirability correlated $-.22$ in the CV sample and $-.20$ in the LV sample. This would tend to lower the correlation between ABLE and ELS if the correlations between ABLE and ASVAB and between ASVAB and ELS were positive, which they were.

Summary of LVI Validation

Generally speaking, the ASVAB was the best predictor of performance. However, the composite of spatial tests provided a small amount of incremental validity for the "can do" criteria (1-3 points), and the ABLE provided larger increments (7-20 points) for two of the three "will do" criteria (Maintaining Personal Discipline, and Physical Fitness and Bearing). Estimates of incremental validity were somewhat higher when the results were not corrected for range restriction.

With regard to ASVAB scoring options, results indicate a very slight edge for using multiple regression equations based on the four ASVAB unit-weighted factor scores. In the test of ABLE scoring options, the method using factor scores computed from a subset of all the ABLE items (ABLE-114) proved to have consistently slightly higher validities.

Perhaps the most interesting finding is derived from the comparisons between the Longitudinal Validation results and those from the Concurrent Validation. Generally speaking, the pattern and level of the validity coefficients were highly similar across the two samples. The correlation between the CV and LV coefficients in Table 1.20 was $.962$ and the root mean

squared difference between the two sets of coefficients was .046. However, the correlation is not 1.00. As noted above, the longitudinal validities were higher for cognitive predictors against "will do" criteria and lower for ABLE composites against "will do" criteria. Some of the possible explanations for those differences include changes in the nature of predictor scores when administered in a longitudinal versus concurrent design, changes in criterion or predictor scores due to cohort differences, and changes in the true relationship between abilities and performance as persons gain more experience and training in an organization and job. These and other possible explanations will be explored in future analyses.

Results of the Concurrent Sample Second-Tour Validation (CVII)

The CVII validation results are based on the CVII sample which was assessed on the criterion measures of second-tour performance at the same time that the LVI performance data were collected from the first-tour longitudinal sample. The predictor set is limited to ASVAB and ABLE because only a small proportion (approximately 12%) of the CVII sample had been assessed with the Experimental Predictor Battery. ASVAB scores, taken 5-6 years earlier, were available from the Enlisted Master File. The ABLE was administered concurrently during the CVII data collection to approximately 45 percent of the total sample (i.e. those individuals who had no peers in the sample to rate and thus had time to take the ABLE). Everyone in the sample was assessed on the full set of second-tour performance measures. By design, the MOS in the CVII sample were limited to the MOS in Batch A. Because of the generally small samples for individual MOS, results for most analyses are reported for the combined sample.

The CVII data collection and data presentation are described in the first annual report for Building the Career Force (Campbell & Zook, 1990; see Chapters 5 and 6). After final editing, the total N for CVII was 1,053. The total sample was distributed across the Batch A MOS as shown in Table 1.21.

Because of some missing data, the sample sizes varied depending on the specific analysis being reported. For example, for the reasons cited above, ABLE scores were available only for 477 individuals. All the analyses that require a common covariance matrix for ABLE and ASVAB were based on this reduced sample.

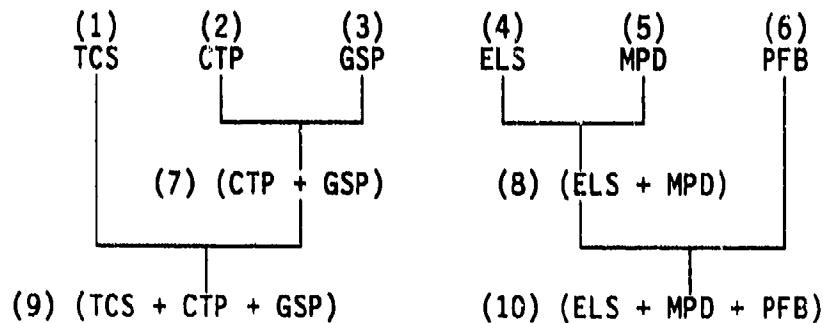
The development of the CVII performance measures, and the analysis and modeling of CVII performance, all have been described previously (Campbell & Zook, 1990 and are summarized in a previous section of the present chapter. The solution that yielded the best fit consisted of six substantive factors and two methods factors. The two methods factors were defined to be orthogonal to the substantive factors, but the correlations among the substantive factors were not so constrained. The six substantive factors and two methods factors, and the variables that are scored on each, were shown in Figure 1.9.

Table 1.21

CVII Sample Sizes by MOS

MOS	N
11B Infantryman	127
13B Cannon Crewmember	162
19E M60 Armor Crewman	33
19K M1 Armor Crewman	10
31C Single Channel Radio Operator	103
63B Light-Wheel Vehicle Mechanic	116
71L Administrative Specialist	112
88M Motor Transport Operator	144
91A Medical Specialist	146
95B Military Police	141
Total	1,053

The complete basic validation analyses utilized a total of 10 scores for the performance factors, as shown below.



TCS = Training/Counseling Subordinates; CTP = Core Technical Proficiency; GSP = General Soldiering Proficiency; ELS = Effort/Leadership; MPD = Maintaining Personal Discipline; PFB = Physical Fitness Bearing

That is, all 10 scores were used as criterion measures. All higher order composite scores were obtained by standardizing the component scores and then taking the simple sum.

Procedure

The CVII validation analysis procedure consisted of the following steps.

- (1) The ASVAB and ABLE were correlated with the six performance factor scores, their five residual scores (there was no residual for TCS), the higher order factor composites, the two methods factor scores, and the total score from the

hands-on tests, the job knowledge tests, and the Situational Judgment Test. ASVAB was represented by the AFQT, a regression-weighted composite of the four factors, and a regression-weighted composite of the nine subtests. ABLE was represented by the three alternative sets of scores described previously. Both corrected (for multivariate restriction of range) and uncorrected estimates were computed, and both regression weights and unit weights (applied to standardized scores) were used. When multiple regression weights were used, the Rozeboom correction (Rozeboom, 1978) was used to account for the fitting of error.

- (2) As in CVI, incremental validities for the ABLE composites over the ASVAB composites were also computed against each criterion score.
- (3) A hierarchical regression analysis, stopping at six predictors, was run against each performance factor, factor composite, and individual criterion score (i.e., hands-on, job knowledge, and Situational Judgment Test).
- (4) A hierarchical regression analysis was also carried out on selected criterion variables for the combined samples from three MOS clusters. The clusters were based on the results of an MOS clustering within the Synthetic Validation Project (Wise, Peterson, Hoffman, Campbell, & Arabian, 1991) and on the results of the validity generalization analysis for the Batch A MOS in the CVI sample (Wise, McHenry, & Campbell, 1990).
- (5) The final step consisted of using the optimal six variable equations from the hierarchical regression analyses described above to develop a picture of the degree of differential prediction across performance factors and across the three MOS clusters.

Results

The basic multiple correlations for ASVAB (four factors vs. nine subtests) and ABLE (seven theoretically based composites vs. seven "purified" empirical factors) are given in Table 1.22. Several things are worth noting. ASVAB, taken at time of entry, is still a highly valid predictor of Core Technical and General Soldiering Proficiency and has respectable validity for Effort/Leadership. For ASVAB, the four factors and the nine subtests provide virtually the same level of predictive accuracy. However, for ABLE the reduced factor scores (114 items) are consistently the best predictor set. ABLE predicts Effort/Leadership and Physical Fitness very well and has reasonable correlations with General Soldiering and Training/Counseling.

In general, after adjustments, regression weights and unit weights for ASVAB yield about the same level of validity. However, regression weights are somewhat better than unit weights for the seven empirical ABLE factors. There is not as much positive manifold among the ABLE factors as there is among the ASVAB subtests.

Table 1.22

Multiple Correlations for ASVAB Factors, ASVAB Subtests, ABLE Composites, and ABLE-114 Scores Against 19 CVII Criterion Variables (All MOS), With Unit Weights

Variable	ASVAB Factors [4]	ASVAB Subtests [9]	ABLE Composites [7]	ABLE-114 [7]
Core Technical (Raw)	43 (42)	43 (43)	15 (14)	20 (15)
General Soldiering (Raw)	56 (54)	57 (55)	22 (16)	26 (18)
Effort/Leadership (Raw)	38 (38)	39 (38)	37 (32)	41 (32)
Personal Discipline (Raw)	00 (11)	00 (11)	20 (21)	18 (22)
Physical Fitness (Raw)	13 (16)	06 (16)	32 (23)	34 (21)
Training/Counseling (Raw)	06 (13)	00 (12)	27 (19)	23 (18)
Core Technical (Res)	29 (29)	28 (30)	00 (12)	07 (13)
General Soldiering (Res)	42 (42)	43 (42)	14 (15)	18 (16)
Effort/Leadership (Res)	25 (26)	27 (25)	38 (31)	41 (30)
Personal Discipline (Res)	00 (09)	00 (09)	16 (20)	15 (19)
Physical Fitness (Res)	16 (20)	09 (20)	34 (21)	35 (18)
ELS - No Situational Judgment	24 (22)	23 (22)	34 (31)	38 (30)
Criterion Composite CTP/GSP	57 (55)	58 (56)	22 (17)	27 (19)
Criterion Composite ELS/MPD	29 (30)	29 (29)	34 (32)	37 (32)
Criterion Factor 1 CTP+GSP+TCS	50 (50)	50 (50)	29 (22)	32 (23)
Criterion Factor 2 ELS+MPD+PFB	14 (16)	12 (15)	34 (35)	35 (34)
Hands-On Average	39 (40)	38 (40)	12 (12)	18 (13)
Job Knowledge Total	59 (56)	59 (57)	25 (14)	28 (16)
Situational Judgment	42 (43)	42 (43)	27 (20)	31 (21)

Note. N = 412. Adjusted (Rozeboom formula). Validities of unit-weighted composites are in parentheses. Numbers in brackets are the number of predictor scores entering prediction equations. Decimals omitted.

Table 1.23 contains the same type of incremental analyses that was done in CVI (Campbell & Zook, 1991). ABLE does not add to the prediction of Core Technical and General Soldiering Proficiency, but it adds about the same amount to the prediction of Effort/Leadership as it did in CVI. However, the overall level of prediction for ELS is higher in CVII than it was in CVI ($R = .50$ vs. $.43$).

The hierarchical procedure asked for the optimal six-variable equation. For any specific criterion measure the first four variables were never all from ASVAB or all from ABLE. It appears that ABLE, most frequently the Dependability scale, does play a role in predicting CTP and GSP. This contribution is masked when the non-hierarchical procedure is used.

Table 1.23

Multiple Correlations for ASVAB Factors Plus ABLE Composites and Plus ABLE-114 Scores, and for ASVAB Subtests Plus ABLE Composites and Plus ABLE-114 Scores Against 19 CVII Criterion Variables, All MOS

Variable	4 ASVAB Factors + 7 ABLE Comp (K=11)	4 ASVAB Factors + 7 ABLE-114 (K=11)	9 ASVAB Subtests + 7 ABLE Comp (K=16)	9 ASVAB Subtests + 7 ABLE-114 (K=16)
Core Technical (Raw)	.42	.43	.42	.43
General Soldiering (Raw)	.56	.57	.58	.58
Effort/Leadership (Raw)	.49	.49	.49	.50
Personal Discipline (Raw)	.16	.13	.09	.03
Physical Fitness (Raw)	.34	.35	.32	.33
Training/Counseling (Raw)	.26	.20	.24	.17
Core Technical (Res)	.24	.26	.24	.25
General Soldiering (Res)	.42	.42	.44	.44
Effort/Leadership (Res)	.43	.43	.43	.43
Personal Discipline (Res)	.09	.07	.00	.00
Physical Fitness (Res)	.36	.37	.34	.34
ELS - No Situational Judgment	.39	.41	.38	.41
Criterion Composite CTP/GSP	.57	.57	.58	.58
Criterion Composite ELS/MPD	.40	.40	.40	.40
Criterion Factor 1: CTP+GSP+TCS	.54	.54	.54	.54
Criterion Factor 2: ELS+MPD+PFB	.35	.35	.34	.34
Hands-On Average	.37	.37	.37	.37
Job Knowledge Total	.60	.60	.60	.60
Situational Judgment	.45	.44	.45	.44

Note. N = 412. Corrected for range restriction and adjusted (Rozeboom formula).

Generalizability

A descriptive picture of the generalizability of prediction equations across performance factors (for the combined sample) is shown in Table 1.24. All entries are multiple correlations and the diagonals represent estimates based on optimal weights. Estimates of what happens when less than optimal weights are used to predict the same criterion are obtained by looking across the rows. Estimates of what happens when a particular set of weights is applied to other criterion measures or other MOS are obtained by looking down the columns. All estimates are based on the corrected covariance matrix. The diagonals are adjusted for shrinkage using the Rozeboom formula with $k = 6$. The off-diagonals are not adjusted because the weights were not computed against that particular dependent variable.

Table 1.24

Multiple Correlations for 10 Sets of Criterion Composite Weights, All MOS

	Raw CTP Weights	Raw GSP Weights	Raw ELS Weights	Raw MPD Weights	Raw PFB Weights	Raw TCS Weights	CTP+GSP Weights	ELS+MPC Weights	Criterion 1 Weights	Criterion 2 Weights
Core Technical (Raw CTP)	.451 (.429)	.436	.331	.165	.193	.195	.446	.298	.434	.154
General Soldiering (Raw GSP)	.553	.571 (.557)	.422	.173	.232	.288	.568	.367	.561	.192
Effort/Leadership (Raw ELS)	.368	.370	.500 (.482)	.375	.046	.358	.372	.485	.404	.422
Personal Discipline (Raw MPD)	.083	.069	.169	.226 (.169)	.057	.130	.075	.197	.694	.188
Physical Fitness (Raw PFB)	.171	.163	.037	.100	.401 (.375)	.055	.168	.059	.135	.235
Training/Counseling (Raw TCS)	.119	.139	.197	.155	.038	.275 (.231)	.131	.196	.175	.178
CTP+GSP	.572	.574	.429	.193	.242	.276	.578 (.564)	.379	.567	.197
ELS+MPC	.272	.265	.403	.359	.061	.293	.270	.412 (.387)	.301	.366
Criterion 1 ^a	.514	.524	.431	.223	.180	.339	.524	.390	.533 (.517)	.235
Criterion 2 ^b	.129	.128	.319	.314	.222	.245	.129	.336	.167	.378 (.350)

Note. Rows are criteria; columns are weights corrected for range restriction; multiple R for optimal weights in bold; Rozeboom adjustments in parentheses.

^a Criterion Factor 1 = CTP+GSP+TCS.

^b Criterion Factor 2 = ELS+MPD+PFB.

As shown in Table 1.24, within MOS there is very little differential validity for Core Technical vs. General Soldiering Proficiency. Either set of weights works about as well. However, the same is not the case for the other four performance factors. Better prediction is always achieved by using the equation developed for each factor.

The greatest degree of differential validity across MOS groups is for General Soldiering and Training/Counseling, not Core Technical Proficiency. The smallest difference is for Effort/Leadership.

Summary of LVII Validity Estimates

In general, in spite of the small samples for each MOS and the necessity of regarding all mean criterion differences as error (i.e., standardizing criterion scores within MOS), the validities for ASVAB and ABLE were as high, or higher, for predicting second-tour performance as for predicting first-tour performance. While unit weights did not weaken the validities for ASVAB, they did constrain the predictive accuracy of ABLE.

A consistent finding from the hierarchical analysis is that for Core Technical Proficiency, General Soldiering Proficiency, and Effort/Leadership criteria, the optimal predictor battery is never composed of only ASVAB or only ABLE factor scores. For example, the Dependability factor from the ABLE is a consistent predictor of the "can do" component of performance.

Finally, based on the above analyses, there appears to be more differential validity across MOS for the second-tour samples than was found during the analyses of the first-tour data in CVI.

All of these issues can be analyzed more rigorously when the larger samples and fuller set of predictor measures from the second-tour longitudinal (LVII) validation are analyzed.

Prediction of Second-Tour Performance From the Trial Battery and From First-Tour Performance

The original research designs for Project A and Career Force include the concept of combining successive pieces of information from (a) predictor tests administered at entry, (b) measures of performance during training, and (c) measures of first-tour job performance to predict individual performance in the second tour of duty.

These analyses of CVI and CVII data examine the relationship of ASVAB scores (given at the time recruits entered the Army), the CVI predictor scores (i.e. the Project A CVI Trial Battery, the preliminary version of the Experimental Predictor Battery, given during the first tour), and first-tour job performance scores to second-tour CVII job performance scores. Two complications with these initial analyses were that available sample sizes for this preliminary exploration were extremely small, and it was unclear exactly how to account for range restriction for a sample of this type.

There were 121 soldiers in Batch A MOS who had been assessed on at least a subset of measures during the CVI and CVII data collections. Not all 121 soldiers had complete CVI and CVII data. The minimum number of soldiers

for a given combination of CVI and CVII measures was 102. Table 1.25 shows the maximum number of soldiers who had CVI and CVII data, by MOS.

Table 1.25

Numbers of Soldiers With CVI and CVII Data by MOS

<u>MOS</u>	<u>N</u>
11B Infantryman	8
13B Cannon Crewmember	26
19E M60 Armor Crewman	4
31C Single Channel Radio Operator	8
63B Light-Wheel Vehicle Mechanic	25
71L Administrative Specialist	15
88M Motor Transport Operator	7
91A/B Medical Specialist	15
95A Military Police	13
Total	121

Measures

The second-tour performance criterion CVII measures used in the analysis were the raw and residual scores for the five constructs first identified during the first-tour Concurrent Validation, and confirmed by the CVII modeling analysis.

Predictor measures came from the ASVAB, from the Project A CVI Trial Battery, and from first-tour job performance measures. The least-squares weights developed for the CVI criterion constructs were used rather than developing new weights for CVII criterion constructs because of the extremely limited sample sizes.

Analysis and Results

CVI predictor scores were correlated with the CVII criterion scores in two ways: (a) Correlations were computed within each MOS and these values were averaged (weighted by N), and (b) correlations were computed across the total sample. Correlations with CVII criteria were computed separately for the ASVAB, Spatial, Computer-administered, ABLE, AVOICE, and JOB composites and for the CVI criterion scores. Correlations were also computed for the ASVAB plus each of the other predictor sets from the Trial Battery and the CVI criteria. When the CVI criteria were combined with any of the other predictor scores, they were standardized within MOS (using the larger CVI samples to compute standard scores) and summed to achieve equal weighting between ASVAB/Trial Battery and CVI criterion scores.

Because of the number of different points at which additional range restriction could occur, there are a number of different "populations" to which the CVII sample could be corrected. If the problem is to select second-tour soldiers from experienced first-tour personnel, then the set of all persons who are nearing completion of the first tour seems the most appropriate population.

The correlations of scores on the first-tour criteria with scores on second-tour criteria in the combined sample are shown in Table 1.26. The correlations are not corrected for restriction of range. The note for the table shows the mean of the diagonal correlations, which contains the correlations of the same criteria across first and second tour--that is, the correlation of Core Technical between first and second tour, and so on. This mean is an index of convergent validity for the set of criterion constructs. The note also shows the mean of the off-diagonal correlations--that is, the correlations between different criterion constructs across first and second tour. The difference between the mean diagonal and mean off-diagonal correlation can be thought of as an indicator of discriminant validity.

Table 1.26

Uncorrected Correlations Between CVI and CVII Raw Criterion Composites
Computed Across Total Sample

CVI Criterion Composite	CVII Criterion Composite				
	CTP	GSP	ELS	MPD	PFB
Core Technical Proficiency	<u>.47</u>	.48	.22	.10	.08
General Soldiering Proficiency	.47	<u>.43</u>	.36	.13	.17
Effort and Leadership	.19	.07	<u>.30</u>	.19	.13
Maintaining Personal Discipline	.06	.14	.16	<u>.26</u>	.19
Physical Fitness and Military Bearing	.00	-.04	.15	.15	<u>.48</u>

Note. Ns = 102-121. Mean diagonal value = .39; mean off-diagonal value = .17.

Table 1.27 shows the correlations, in the combined sample, of predicted scores based on CVI weights for ASVAB and Trial Battery composites and CVI criterion scores with CVII criteria.

On the whole, of all the predictors, the CVI criterion scores have the highest correlations with CVII criterion scores. However, adding the ASVAB and the ASVAB plus Trial Battery composite scores to CVI scores does increment the CVI validity coefficients.

Table 1.27

Correlations Between CVI Weighted Predictor Composites, CVI Criterion Composites, and CVII Criterion Composites for Raw Scores, Computed on Total Sample

Predictor and CVI Criterion Composites and Combinations	CVII Criterion Composite				
	CTP	GSP	ELS	MPD	PFB
ASVAB	.33	.42	.11	-.05	.11
CVI Performance	.47	.43	.30	.26	.48
ASVAB+CVI Performance	.51	.51	.33	.26	.47
Computer Tests	.23	.13	-.01	-.04	.10
ASVAB+Computer Tests	.37	.41	.13	.05	.12
ASVAB+Comp. Tests+CVI Performance	.52	.51	.33	.27	.46
AVOICE	.15	.16	.06	-.02	.06
ASVAB+AVOICE	.43	.44	.14	.00	.13
ASVAB+AVOICE+CVI Performance	.54	.52	.33	.27	.46
JOB	.12	.00	.19	.30	.12
ASVAB+JOB	.33	.41	.16	.20	.16
ASVAB+JOB+CVI Performance	.51	.51	.34	.31	.48
Spatial	.47	.41	.14	-.01	.04
ASVAB+Spatial	.41	.43	.10	-.06	.11
ASVAB+Spatial+CVI Performance	.52	.51	.33	.26	.46
ABLE	.10	.01	.21	.15	.29
ASVAB+ABLE	.34	.41	.22	.12	.25
ASVAB+ABLE+CVI Performance	.51	.52	.36	.30	.47

Note. Ns = 102-121. Correlations are uncorrected for range restriction. Coefficients do not require shrinkage adjustments. CVI criterion scores and predictor composites were summed.

The ASVAB validities follow the familiar pattern of predicting the two "can do" criteria, but not predicting the "will do" criteria very well. The JOB unexpectedly did the best job of predicting Maintaining Personal Discipline.

In sum, these results provide evidence that ASVAB scores, weighted on the basis of regression estimates for predicting first-tour performance, predict second-tour "can do" performance with substantial validity. The results also provide impressive evidence of convergent and discriminant validity of the first-tour job performance for predicting second-tour job performance criteria.

Future analyses of the LVI Experimental Predictor Battery and LVII criterion scores will provide better indications of the new predictors' relationships with second-tour performance.

ORGANIZATION OF THE CURRENT REPORT

This third annual report for the Career Force Project deals exclusively with the second-tour Longitudinal Validation (LVII) data collection and the development of basic criterion scores and performance factor scores for second-tour performance. It replicates much of the work that was done using the CVII data file, but with larger samples and more complete data. In addition, the LVII sample provided a true confirmatory test of the Career Force model of second-tour performance and includes a much higher percentage of individuals who took the Experimental Predictor Battery at the start of their first-term enlistment and who were assessed on the first-tour performance measures.

The objectives of this report are to describe the LVII data collection and data file editing procedures, the development of the basic criterion scores, and the development of the LVII performance model. The chapter organization is as follows.

Chapter 2 describes the steps taken to specify the nature of the sample, obtain the cooperation of the data collection sites, train the data collection team, and administer the second-tour performance measures. Based on experiences with CVII, a number of improvements were made in these procedures.

Chapter 3 describes the way in which the individual task and scale scores from the performance measures were aggregated into a set of basic criterion scores for each measure. The general strategy was the same as for CVII; however, the LVII data file provided a somewhat different array of scores, in comparison to CVII, for the Situational Judgment Test and the Supervisory Simulation Exercises.

Chapter 4 summarizes the content of the LVII data file in terms of sample sizes by MOS, by instrument, and by basic score. It also describes the extent of missing data and outlines the procedures used to deal with the various types of missing observations.

Chapter 5 reports the results of the confirmatory analysis obtained when the CVII performance model was fit to the data from LVII. It also describes a revised model of second-tour performance based on further analyses of the LVII data. For example, the improvements made for LVII allowed for much better measurement of the leadership factor. In a retrospective analysis, the revised model (from LVII) fits the CVII data as well as the original CVII model.

Chapter 6 presents an overall summary of the current report and sets the stage for the fourth annual report.

In sum, the Career Force Project third annual report will describe the collection and analyses of the LVII sample data, up to and including the development of specifications for the revised model of second-tour performance. The factors that comprise the LVII performance model will be

used as criterion measures in the LVII validity analyses. The LVII estimates of the validity of ASVAB, the validity of the Experimental Predictor Battery, and the validity of the first-tour performance measures for predicting second-tour performance will be topics of subsequent reports, as will further considerations of differential prediction and classification efficiency.

Chapter 2 LONGITUDINAL VALIDATION SECOND-TOUR DATA COLLECTION

Deirdre Knapp

The purpose of the LVII data collection was to administer second-tour criterion measures to soldiers in the longitudinal validation sample. Although this data collection involved substantially fewer soldiers than the CVI or LVI data collections, it posed a number of challenges. Having to locate and test individual soldiers, especially when there were relatively few to begin with, made it difficult to meet the project's sample size goals. This problem was made more critical as a result of a major deployment of U.S. troops to Southwest Asia (Operation Desert Shield/Storm) that occurred shortly before the project's data collection activities were initially scheduled to begin.

Despite these and other difficulties, LVII data were collected from 1,577 soldiers. Details regarding final sample sizes, by MOS, after data editing are provided in Chapter 4. The purpose of this chapter is to describe the data collection instruments, test site coordination activities, staffing, and data collection procedures.

DESCRIPTION OF THE MEASURES

A list of the instruments administered in the LVII data collection is provided in Table 2.1. Most of the instruments served as second-tour performance criterion measures, and several other instruments (e.g., the Background Information Form) provided supplemental data for the project. All of the instruments are briefly described below, with more detailed descriptions provided in the next chapter.

Performance Criterion Instruments

Job Knowledge Tests

The job knowledge tests consisted of 100-145 written, performance-based, multiple-choice test items that covered from 27 to 30 technical tasks per MOS. The performance-based test items required examinees to indicate what should be done to accomplish a given task step rather than recalling why a task step should be done in a particular fashion. The job knowledge test items also made liberal use of pictures, drawings, and other aids to depict actual job stimuli. Although the specific tasks covered varied across MOS, soldiers in each MOS were tested on both job-specific and general soldiering tasks.

Hands-On Performance Tests

Approximately half of the technical tasks covered on the written job knowledge tests were also tested using a hands-on format. The hands-on tests required soldiers to perform each task under standardized conditions. Hands-on test performance was scored by breaking down each task into a checklist of discrete, observable steps that were then rated go or no-go (C. H. Campbell et

Table 2.1

LVII Data Collection Instruments

Performance Criterion Instruments

- Job Knowledge Tests
- Hands-On Tests
- Performance Rating Scales (completed by supervisors)
 - Army-Wide Booklet
 - MOS-Specific Booklet
 - Combat Performance Prediction Scales
 - Combat Performance Questionnaire (Operation Desert Shield/Storm), administered if applicable
- Personnel File Form
- Situational Judgment Test (SJT)
- Supervisory Simulation Exercises
 - Personal Counseling
 - Disciplinary Counseling
 - Training

Supplemental Instruments

- Background Information Form
 - MOS-Specific Job History Questionnaire
 - Supervisory Experience Questionnaire
 - Army Job Satisfaction Questionnaire (AJSQ)
 - Assessment of Background and Life Experiences (ABLE)
 - Leader and Unit Attitudes Questionnaire
-

al., 1990). The tests were administered and scored by senior NCOs under the supervision of civilian project personnel.

Performance Rating Scales

In previous criterion data collections conducted as part of Project A, performance ratings were collected from both supervisors and peers. However, because of the relative autonomy of second-tour soldiers and the increased administrative difficulty of identifying and tasking sufficient numbers of second-tour peers to participate in the data collection, performance ratings in LVII were collected from supervisors only.

Supervisors were asked to complete three rating booklets: (a) the Army-Wide Performance Rating Booklet, (b) the MOS Performance Rating Booklet, and (c) the Combat Performance Prediction Scales. Those supervisors who had been deployed to Southwest Asia as part of Operation Desert Shield/Storm along with the soldier they were rating were also asked to complete a second set of combat performance ratings, the Combat Performance Questionnaire. This latter

rating booklet was a criterion measure designed by ARI to measure performance in combat. Analyses of these combat performance ratings data are not presented in the current volume, but will be provided in a subsequent report.

Personnel File Form

The Personnel File Form (PFF) was used to ask soldiers to report information that could be obtained from archival sources, but which can be more efficiently and as accurately gathered through self-report (Campbell, 1987). Administrative indices of performance gathered using this form related to awards and commendations received, education, promotion history, disciplinary actions received, and operational test results (e.g., Individual Weapons Qualification).

Situational Judgment Test (SJT)

The SJT consisted of 49 written, multiple-choice test items that covered supervisory-related job content. Each item depicted a scenario involving a realistic problem situation that might face a first-line supervisor. From the three to five response alternatives provided for each question, soldiers indicated which response they believed would be most effective for handling the situation, and which response would be least effective.

Supervisory Simulation Exercises

Critical supervisory tasks were simulated by having civilian test administrators play the subordinate's role in each of three scenarios. The scenarios presented problems which required (a) personal performance counseling, (b) disciplinary counseling, and (c) one-on-one training.

Supplemental Instruments

Background Information Form

There were three versions of the Background Information Form: one for examinees, one for supervisor raters, and one for the NCOs who administered the hands-on tests. The form asked for identifying and background information such as social security number, test date, test site, and primary and duty MOS.

MOS-Specific Job History Questionnaire

For each technical task tested via the job knowledge and/or hands-on tests, soldiers were asked to indicate how frequently they had performed the task within the previous 6 months and how long ago they had last performed the task.

Supervisory Experience Questionnaire

On this form, soldiers indicated their experience with various supervisory tasks that were tested via the SJT and the supervisory simulation exercises. They reported how often they had performed each task within the previous 6 months and the first time they performed the task. In addition, soldiers indicated how often they were given responsibility to supervise others.

Army Job Satisfaction Questionnaire (AJSQ)

The AJSQ measures satisfaction with six aspects of Army life: Work, Pay, Promotions, Co-Workers, Supervision, and the Army as an Organization. The AJSQ that was administered during this data collection was a slightly modified form of the version administered to soldiers in the LVI/CVII data collection.

Assessment of Background and Life Experiences (ABLE)

To collect test-retest data on one of the more promising Project A predictor tests, the 114-item ABLE was administered to LVII soldiers as time permitted. That is, soldiers were asked to complete the ABLE if doing so did not interfere with their ability to complete the other instruments.

Leader and Unit Attitudes Questionnaire

This short questionnaire was developed by ARI to support research interests related to the broader ARI research program. The 24 questions asked soldiers about their attitudes towards their supervisors, their unit, and the Army as a whole.

OBTAINING AND SCHEDULING THE REQUIRED TROOP SUPPORT

The original project plan called for the LVII data collection to take place July-December 1991. Second-tour criterion data were to be collected from at least 150 soldiers in each of nine MOS for an overall sample size of 1,350. The MOS are the Batch A MOS (excluding 19E) that were listed in Figure 1.2.

Even before the deployment of troops to Southwest Asia created havoc with the data collection plans, project personnel anticipated difficulty obtaining required LVII sample sizes. Several obstacles that were encountered during the LVI data collection were expected to be factors in the LVII data collection as well. These problems included difficulty projecting future location of soldiers targeted for testing because of frequent reassignments, and difficulty getting individual soldiers to testing (e.g., because of limited access due to training or alert status, leave, and so forth). The difficulty of projecting troop location was compounded by a tasking system which requires that Troop Support Requests (TSRs) be submitted by ARI one year prior to data collection. Moreover, before data collection planning activities began, the Army was starting to respond to directives to downsize and to reduce the proportion of troops stationed in Germany. It was anticipated that this would lower reenlistment rates and compound the problems associated with tracking individual soldiers and scheduling them for testing.

These concerns led to the development of a data collection strategy which would be flexible enough to accommodate the problems that were anticipated. The strategy determined (a) which soldiers would be eligible for testing, (b) whether hands-on tests would be administered, (c) where testing would take place, (d) how soldiers would be tasked for testing, and (e) when the testing would take place. Each of these planning elements will be described briefly.

Eligibility for Testing

The amount of data available for Project A soldiers varies depending upon the data collections in which they have been included. Specifically, soldiers may have predictor data (collected in 1986-1987), first-tour criterion data (collected in 1988-1989), or both. In May 1990, the World Wide Locator (WWL) system data base was queried to determine the number of Project A soldiers who were still in the Army and their locations. At this time it became clear that sample size requirements would not be met if only soldiers having predictor and first-tour criterion data were tested. Accordingly, the decision was made to test soldiers for whom predictor and/or first-tour criterion data were available. Soldiers with no Project A data were not eligible for testing.

Hands-On Tests

Shortly after the LVI data collection ended, MOS 31C began declining in strength due to the phasing out of certain radio equipment. The collection of hands-on data is inordinately resource-intensive for small numbers of examinees. On the basis of these considerations, hands-on tests were dropped from the 31C soldiers' performance measures.

The WWL information from May 1990 also indicated that soldiers in the 71L and 88M MOS were relatively few in number and were spread out in many locations that data collectors would be unable to reach. While contingency plans were made to drop hands-on tests from these MOS if necessary, these plans were never implemented.

Testing Locations

The initial query to the WWL data base also indicated that appreciable concentrations of Project A soldiers were stationed in locations other than those identified in the original research plan. Accordingly, requests for troop support were written to include some of these new sites (e.g., Fort Drum, Eighth Army, US Army Pacific). Data were subsequently collected at some, but not all, of the new sites.

Soldier Taskings

To minimize problems associated with forecasting the exact location of soldiers, the Troop Support Request package (originally submitted in May 1990) did not identify the specific soldiers to be tested at each test site. Moreover, it requested a large number of soldiers for testing at each test site even though it was anticipated that fewer would actually be tested. Each test site was then provided with a computer diskette containing the social security numbers of all soldiers eligible for testing. By matching these social security numbers with each installation's own personnel files, the most accurate identification of soldiers available for testing at each location was obtained.

Data Collection Schedule

The original research plan called for LVII data to be collected July-December 1991. To accommodate the interests of supporting commands, it was

agreed that test sites could be scheduled to conduct testing as early as May 1991 and as late as February 1992.

Once the formal Troop Support Request was submitted and approved in principle by supporting commands, details regarding the data collection (including specific test dates) were coordinated with individual test sites. Coordination procedures were somewhat different for two commands (US Army Europe [USAREUR] and the Eighth Army) because the test sites were outside the Continental United States.

The data collection strategy described in the preceding section was established before hostilities involving U.S. troops in Southwest Asia arose. After initial negotiations with individual test sites were underway, the U.S. Forces Command (FORSCOM), which had been tasked to provide the majority of LVII soldiers, invoked a moratorium on research support. This moratorium was imposed in September 1990 and lifted in April 1991. The flexibility of our original troop support request strategy helped ensure that the required data could be collected despite this unforeseen obstacle. However, an unexpectedly large proportion of the LVII data was collected overseas in Germany and the Republic of Korea and the data collection window was extended further into 1992.

The complete LVII data collection schedule is shown in Table 2.2. Note that four data collection teams were sent to Germany, whereas one team of data collectors was sent to each of the other test sites. The first LVII data collection occurred in June 1991 and the last in July 1992. Composition of the teams, in terms of project staff, varied from location to location.

SITE COORDINATION

Once the test dates and daily schedule were negotiated for each test site, the required personnel, facilities, and equipment were located and obligated. Required personnel included name-requested examinees, their supervisors, senior NCOs to administer the hands-on tests, and support NCOs. Indoor facilities were required to accommodate written testing, some hands-on and simulation (i.e., role-play) administration, supervisor rating sessions, and general office and storage needs. Large outdoor areas were required for most hands-on testing. The hands-on administration also required varied pieces of equipment and other materials.

The Army provided a point of contact (POC) for each test site to negotiate a testing schedule and manage on-site data collection preparation activities. The POC was usually an officer assisted by a senior NCO. Initial contact and coordination with test site POCs was usually made by the Task 1 leader. Once a test site manager (TSM) was designated for the test site, coordination efforts shifted to that individual.

Many lessons regarding advance site coordination were learned during the Project A criterion-related data collections. To make the most of this prior experience, a manual was prepared and provided to the POC at each test site. This manual was designed to orient the POC to the purpose and nature of the LVII data collection. It provided detailed instructions for locating and

Table 2.2

LVII Data Collection Schedule

<u>Command</u>	<u>Location</u>	<u>Test Dates</u>
<u>1991</u>		
USAREUR	Germany	7 June - 27 June
USAREUR	Germany	5 July - 2 August
USAREUR	Germany	5 July - 3 August
Eighth Army	Republic of South Korea	5 July - 9 August
USAREUR	Germany	September - October
HSC	Fort Sam Houston, TX	October
FORSCOM	Fort Lewis, WA	9 December - 19 December
<u>1992</u>		
FORSCOM	Fort Drum, NY	13 January - 24 January
TRADOC	Fort Bliss, TX	20 January - 31 January
MDW & AMC	Fort Belvoir, VA	February
TRADOC	Fort Knox, KY	2 March - 6 March
FORSCOM	Fort Bragg, GA	16 March - 3 April
TRADOC	Fort Benning, GA	31 March - 3 April
FORSCOM	Fort Riley, KS	6 April - 10 April
FORSCOM	Fort Hood, TX	4 May - 15 May
FORSCOM	Fort Campbell, KY	11 May - 15 May
FORSCOM	Fort Carson, CO	1 June - 5 June
FORSCOM	Fort Stewart, GA	15 June - 23 June
TRADOC	Fort Polk, LA	13 July - 16 July

USAREUR	U.S. Army Europe
HSC	Health Services Command
FORSCOM	Forces Command
TRADOC	Training and Doctrine Command
MDW	Military District of Washington
AMC	Army Materiel Command

tasking the required personnel, facilities, and equipment, and answered questions regarding the requirements which had frequently arisen in earlier data collections.

DATA COLLECTION PROCEDURES

Generally, each test site was staffed with a team comprised of the following personnel:

- 1 Test Site Manager (TSM)
- 1-2 Hands-on Managers (HOMs)
- 3 Test Administrators (TAs)

All of these positions were filled by permanent employees of the contractor consortium. The Army installations also provided personnel to help support the data collection activities. In addition to the test site POC, each test site provided eight senior NCOs for each MOS (except 31C) to administer and score the hands-on tests and two to four NCOs to fill general supporting roles (e.g., to track down soldiers who fail to report for testing and handle problems with defective equipment).

TSMs were responsible for all aspects of the data collection activities on-site. HOMs were responsible for training NCOs to administer and score the hands-on tests and for supervising all aspects of the hands-on testing activities. Because all individuals selected to be TSMs and HOMs had considerable experience with earlier Project A criterion-related data collections, their training focused on the specific requirements of the LVII data collection.

TAs were responsible for (a) administering the written measures, (b) playing the role of the subordinate in one or more of the simulation exercises, and (c) collecting performance ratings from supervisors. Many of the TAs had prior experience with Project A data collections. Those who did not have Project A experience had had experience collecting data in other military and/or civilian projects.

Data Collection Team Training

One day of classroom training and considerable follow-up on-the-job training was provided to TAs for the written test and supervisor rating procedures. One to two days of additional training was provided to each TA for each subordinate role a TA was responsible for playing. TAs were trained to play only one role at a time and most TAs played only one role during the course of the data collection. The training for the simulation exercise emphasized the need for standardization in role-playing and scoring, and provided for considerable practice.

In addition to covering test administration and role-playing requirements, TA training reviewed (a) background of the Project A/Career Force research program, (b) things to know on an Army post (e.g., rank insignia), and (c) procedures for the secure maintenance of test materials and data. Two documents were developed to support TA training: the Test Administrator's Manual and the Supervisory Role-Play Exercises Administration Manual.

NCO hands-on scorers were trained the day before the administration of the hands-on tests to soldiers in a given MOS. The training followed the same basic procedures as those that had been used in the CV and LVI/CVII data collections (R. Campbell, 1985). It focused on the need to administer and score the tests in a standardized fashion, and provided for several practice dry-runs.

Instrument Administration Procedures

Ordinarily, only one MOS was tested each day. If another MOS was scheduled for testing the following day, NCO hands-on scorer training for that MOS was conducted concurrently with test administration activities for the preceding MOS. The schedule generally followed for administering the measures

is shown in Table 2-3. There were exceptions to this general schedule, usually to accommodate late arrivals, inclement weather conditions, and various other contingencies. Testing was typically restricted to 20 or fewer soldiers per day to allow for timely completion of the hands-on tests and simulation exercises.

Table 2.3

LVII Daily Testing Schedule^a

0730	In-process soldiers
0800	Hands-on tests and supervisory simulation exercises
.	" "
:	" "
.	" "
1200	Lunch
1300	Job knowledge test
1400	Situational Judgment Test
1500	Personnel File Form, Job History Questionnaire, Army Job Satisfaction Questionnaire
1600	Supervisory Experience Questionnaire, Leader and Unit Attitudes Questionnaire, Assessment of Background and Life Experiences

^a Supervisor rating sessions were generally conducted during the afternoons, concurrent with the written testing sessions.

Each day of test administration began with soldier in-processing. After it was determined whether any soldiers scheduled for testing were missing, soldiers were given a briefing which explained the purpose of the project and described the day's activities. The Privacy Act was read aloud at this time.

Half a day was devoted to hands-on and simulation administration. The tests were set up so that soldiers rotated through nine test stations. One test station comprised the three supervisory simulation exercises and the remaining test stations each comprised one or more technical task tests. Before testing began, the HOM oriented the soldiers to the testing rotation arrangement. Soldiers were not required to complete the tests in any particular order; sign-off cards were used to keep track of which tests they had or had not taken.

The second half of the day was devoted to the written tests. Although the order of test administration was fairly structured, the administration times shown in Table 2.3 are approximations only. Examinees were given all the time they needed to complete the criterion measures. The Test Administrator's Manual provided standard instructions for administering each written measure.

Project staff attempted to collect performance ratings from at least two supervisors per soldier. Although test site POCs were responsible for identifying supervisor raters prior to the arrival of the data collection team, the lists were often incomplete and/or inaccurate. Once on-site, project staff identified additional raters based on input from examinees and other supervisors. This information, as well as the names of supervisors who had not reported as scheduled, was relayed to the test site POC. The POC, with the assistance of his or her support staff, was then responsible for contacting and scheduling or rescheduling raters.

Post Data Collection Activities

Various procedures and documents were used to handle completed data collection instruments before shipping them to the facility where they would be processed and keypunched. Test site personnel checked measures for completeness and legibility, and documented explanations for data which were incomplete or otherwise anomalous. Transmittal documents were used to help ensure that data could be tracked once it left the test site.

After testing at a given location was completed, the TSM prepared and submitted a report to ARI. This report summarized the support provided by the installation (e.g., number of examinees and supervisor raters provided) and described any significant problems encountered during testing.

Chapter 3 ANALYSES OF LVII PERFORMANCE MEASURES

Deirdre Knapp, Charlotte Campbell, Mary Ann Hanson, Ken Bruskiwicz,
Cheryl Paullin, Carolyn Hill-Fotouhi, Chris Sager, and Leissa Nelson

This chapter will describe how basic scores for the LVII performance criterion measures were developed. The measures were introduced in the preceding chapter and have been described in detail elsewhere (Campbell, 1988; Campbell & Zook, 1990). They were originally administered to second-tour soldiers in the CVII sample and were subsequently revised in preparation for administration to the LVII sample.

Analyses of the data from the LVII sample had three major objectives: (a) to examine and evaluate the psychometric properties of the LVII measures, (b) to compare the psychometric properties of the LVII scores with the CVII scores, and (c) to develop basic scores to be used in modeling second-tour performance. Description of the measures and the derivation of basic scores will emphasize the similarities and the differences between the LVII and CVII research.

JOB KNOWLEDGE AND HANDS-ON TESTS

A set of 28-30 tasks had been selected for performance measurement in each MOS. The procedures used to select tasks and to develop task tests for each of the nine Batch A MOS are described in previous reports (Campbell, 1989; Campbell & Zook, 1990). All tasks were assessed using a written job knowledge test format. Performance on a subset (14-17) of the tasks was assessed using a hands-on performance test format. The knowledge test items were multiple choice, with one correct answer per item. Performance steps for each task tested hands-on were scored GO or NO-GO by a trained NCO scorer. A list of the tasks comprising the hands-on and job knowledge test components for each MOS is presented in Appendix A.

Soldiers are responsible for tasks at their own and lower skill levels. The set of tasks selected for performance measurement in each MOS included (a) common tasks which were drawn from the Soldier's Manual of Common Tasks, Skill Level 1 (STP 21-1-SMCT, October 1985) and the Soldier's Manual of Common Tasks, Skill Level 2/3/4 (STP 21-24-SMCT, Draft, January 1987), and (b) MOS-specific tasks which were drawn from the relevant MOS-specific Soldier's Manuals. Common tasks are basic soldiering tasks that all soldiers are expected to know how to perform (e.g., first aid, personal weapons, map reading); MOS-specific tasks are central to the jobs of the soldiers working in a given MOS and are typically unique to that MOS. Tasks that were seldom performed at Skill Level 2 were not selected for testing (see Campbell, 1989).

Some tasks are performed differently depending upon the type of equipment a soldier uses (e.g., an M16A1 rifle versus an M16A2 rifle). To deal appropriately with such situations, tracked (i.e., parallel) tests were prepared for tasks where equipment might vary. In some cases, equipment variations required only minor changes in the task steps. In other cases, the omission of only a few steps resulted in the tasks being judged as having similar behavioral requirements.

Before the measures that had been developed for CVII could be used again, technical currency reviews were also conducted. Each job knowledge and hands-on test was reviewed against Army doctrinal training materials by project staff. Revisions were made to test items and to supporting graphics and handouts as necessary. All revisions were evaluated by the MOS proponent agencies. This evaluation led to the decision to drop some steps, items, or task tests because they were no longer doctrinally appropriate.

At the time of the CVII data collection, the Army's transition from the M60-series tank (used by MOS 19E) to the M1-series tank (used by MOS 19K) was in progress. Second-tour performance measures for 19K had not been developed at that time. Consequently, second-tour job knowledge and hands-on tests for 19K had to be developed for LVII. Test development for the 19K tasks followed essentially the same steps as were followed for CVII (see Campbell & Zook, 1990). However, as mentioned in Chapter 2, for MOS 31C the equipment transition led to the decision to administer job knowledge tests, but not hands-on tests, to 31C soldiers.

Finally, many of the hands-on task tests result in a product generated by the test taker (e.g., a completed maintenance form, a typed memorandum, a set of grid coordinates, a firing data record). In previous data collections, NCO scorers were trained to score these products. To reduce the burden on NCO scorers and increase the accuracy of the scoring process, LVII products resulting from the hands-on tests were scored by the Hands-On Manager.

Scoring Adjustments

Specifications for the basic scores for the LVII job knowledge and hands-on measures depended largely on previous work in CVI, CVII, and LVI. As with the previous data collections, five potential sources of systematic error were addressed: variation in the number of steps/items per task test, multiple tracks, missing data, site differences, and marginal items. The procedures used to minimize the effects of these sources of variance were, for the most part, the same as for previous analyses.

Number of Test Items. Because the number of items in a task test was not necessarily related to the importance of the task, job knowledge and hands-on task scores were calculated as percent-correct (or percent-GO) scores at all score levels.

Tracked Tests. The data for tracked tests were examined for evidence of level and dispersion differences between tracks in the test scores that would reflect differences in test difficulty rather than individual differences among soldiers. No anomalous differences were found. The percent-correct/GO scoring scheme was considered adequate for correcting for variation in number of items or steps performed between tracked versions of the task tests.

Missing Data. On hands-on tests, data could be missing for one of three principal reasons: (a) the scorer failed to observe a step or failed to record the observation, (b) the scorer marked both GO and NO-GO, or (c) equipment was unavailable for testing part or all of a task. Whatever the reason, the fact that the observation was missing was irrelevant to the soldier's performance. In the job knowledge tests, there were two likely

reasons for missing data: Either the soldier skipped an item or the soldier did not get to one or more items at the end of the test booklet. Methods used to adjust for the missing data are discussed in Chapter 4.

Site Differences. Because it was not always possible to faithfully replicate testing conditions at the various test sites, hands-on test scores could potentially reflect site differences. Type of testing facility, condition of equipment, local operating procedures, and weather and terrain conditions all interfered with standardization of test administration. Analysis of variance was used to examine site differences within tasks, and statistically significant differences were found for almost all tasks. Therefore, as with the previous data collections, hands-on test scores were standardized by site at the task level to control for site differences.

Marginal Items. An adjustment which affected only the job knowledge tests concerned marginal items. Because of changes in equipment and changes in the prescribed steps in performance between the CVII testing and the LVII testing, not all test items were keyed correctly when the tests were administered--this despite rigorous currency review and careful proponent agency examination. In some cases, no correct answer was included in the list of responses, and those items were dropped. Between one and four items per MOS were dropped because of such doctrinal changes.

Table 3.1 shows the overall number of items in the job knowledge component for each MOS and the range of items per task test. Table 3.2 shows the overall number of steps in the hands-on component for each MOS and the range of steps per task test.

Table 3.1

Number of LVII Job Knowledge Tests and Items by MOS

MOS	No. of Tasks	Items Dropped	Total Items	Items Per Task	Average Items Per Task
11B Infantryman ^a	29	2	128	2-12	4.4
13B Cannon Crewmember ^a	30	3	119-120	2-8	4.0
19K M1 Armor Crewman	28	4	142	3-12	5.1
31C Single Channel Radio Operator ^a	30	1	111-112	3-5	3.7
63B Light Wheel Vehicle Mechanic	27	2	102	2-6	3.8
71L Administrative Specialist ^a	30	2	125	2-12	4.2
88M Motor Transport Operator	30	1	119	3-12	4.0
91A/B Medical Specialist	30	3	113	2-6	3.6
95B Military Police	29	4	109	2-7	3.8

^a One or more task tests were tracked; tracked tests do not necessarily have the same number of items.

Table 3.2

Number of LVII Hands-On Tests and Steps by MOS

MOS	No. of Tasks	Total Steps	Steps Per Task	Average Steps Per Task
11B Infantryman	9	121	5-31	13.4
13B Cannon Crewmember ^a	12	258-259	7-67	21.5-21.6
19K M1 Armor Crewman	10	167	8-37	16.7
63B Light Wheel Vehicle Mechanic ^a	8	142	7-44	17.8
71L Administrative Specialist ^b	14	140-146	2-44	10.0-10.4
88M Motor Transport Operator ^a	10	193-195	4-44	19.3-19.5
91A/B Medical Specialist ^a	13	216	6-44	16.6
95B Military Police ^a	10	223-227	7-37	22.3-22.7

^a One or more task tests were tracked; tracked tests do not necessarily have the same number of steps.

^b One task was scored on a continuous scale; it is not included in calculating total steps, steps per task, or average steps per task.

Score Construction

After data editing, four levels of scores were constructed. The four levels (Tasks, Functional Categories, Task Factors, and Task Constructs) are the same as those described in Chapter 1 and depicted in Figure 1.10. The four-level scoring scheme evolved from earlier research. The Functional Categories were constructed for the CVI and CVII tests by asking expert judges to sort tasks into homogenous categories. Using CVI data, Functional Category scores were, in turn, reduced by a series of exploratory and confirmatory analyses to a smaller set of Task Factors.

Task Factor scores were then subjected to another round of empirical factor analysis along with other criterion scores (from various rating scales and administrative records). The scores split between two higher-order factors, labeled General Soldiering Proficiency and Core Technical Proficiency. This resulted in two Construct scores: a Basic (non-MOS-specific) score comprised of tasks that loaded on General Soldiering Proficiency and a Technical (MOS-specific) score comprised of tasks that loaded on Core Technical Proficiency.

As the first LVII step in replicating the CVII procedures for constructing the basic scores, tasks were clustered into Functional Categories. The Functional Category rules developed for CVII define 10 across-MOS categories, plus one to five MOS-specific Technical Categories. At the next stage, tasks were sorted into six Task Factors (Safety/Survival, Basic Techniques, Communication, Identify Targets, Vehicles, and Technical). Finally, tasks were combined to form two Task Construct scores: General (formerly termed Basic) and MOS-Specific (formerly termed Technical).

The assignment of tasks to Functional Categories, Task Factors, and Task Constructs is shown in Appendix A. At each level of aggregation, hierarchical scores were computed using task-level data. That is, each category, factor, and construct score was computed by calculating the mean percentage of items correct (or percentage of steps passed) across all constituent tasks.

Final Basic Scores for Job Knowledge and Hands-On Measures

Descriptive statistics calculated across MOS for both the Task Construct and Task Factor scores are provided below.

With regard to the Task Construct scores for the job knowledge tests, the mean General score across all MOS except 11B was 64.94 (SD = 10.38, N = 1,238) and the mean MOS-Specific score was 61.84 (SD = 11.05, N = 1,238). The correlation between these two sets of scores was .506 and their split-half reliability estimates were .658 for the General score and .517 for the MOS-Specific score. For the hands-on tests, the mean General score across MOS was 70.85 (SD = 11.60, N = 1,152) and the mean MOS-Specific score was 70.65 (SD = 12.54, N = 1,145). The correlation between the two sets of hands-on scores was .205.

Tables 3.3 and 3.4 show the means, standard deviations, and inter-correlations among the 11 sets of Task Factor scores (six job knowledge and five hands-on), across MOS. Means and standard deviations for all four levels of scores (i.e., Task, Functional Category, Task Factor, Task Construct), computed by MOS, are shown in Appendix B.

Table 3.3

Intercorrelations Among LVII Job Knowledge Task Factor Scores Across MOS

Task Factor	Safety/ Survival	Basic Soldiering	Commo.	Identify	Vehicles	Technical (MOS)
Safety/ Survival	1.00					
Basic Soldiering	.46	1.00				
Communications	.25	.36	1.00			
Identify	.24	.30	.23	1.00		
Vehicles	.20	.31	.18	.2	1.00	
Technical (MOS)	.35	.42	.24	.30	.29	1.00
Mean	68.43	60.13	71.00	78.96	55.41	61.13
Standard Deviation	13.24	11.67	20.68	17.08	21.06	11.94
N	1,583	1,583	1,583	1,583	915	1,238

Overall, the Task Factor results for the LVII testing do not differ much from the results for the CVII soldiers tested (which may be found in Campbell & Zook, 1990). For the hands-on tests, the task factor scores (across MOS) for the two sample groups are within 4 percentage points for four of the factors (Safety/Survival, Basic Soldiering, Communications, and Technical-MOS). On both of the other two factors (Identify and Vehicles), CVII soldiers scored higher than did LVII soldiers, by an average of about 13 and 7

Table 3.4

Intercorrelations Among LVII Hands-On Task Factor Scores Across MOS

Task Factor	Safety/ Survival	Basic Soldiering	Commo.	Identify	Vehicles	Technical (MOS)
Safety/ Survival	1.00					
Basic Soldiering	.19 (1,380)	1.00				
Communications	.23 (919)	.24 (919)	1.00			
Identify	.13 (363)	.16 (363)	.07 (363)	1.00		
Vehicles	.17 (593)	.19 (593)	.39 (286)	.17 (203)	1.00	
Technical (MOS)	.21 (1,056)	.26 (1,056)	.17 (595)	.21 (363)	.12 (593)	1.00
Mean	77.44	77.15	57.05	63.37	65.87	72.34
Standard Deviation	15.74	11.83	23.82	23.95	16.33	13.37
N	1,483	1,483	961	378	652	1,143

Note. Sample sizes are shown in parentheses.

percentage points (less than one standard deviation), respectively. For both groups, the Communications score was lower than other scores (59% for CVII and 57% for LVII): scores on the other factors ranged from about 71 percent to about 79 percent for CVII soldiers, while the range for LVII soldiers was between about 63 percent and 77 percent.

Similarly, on the job knowledge tests, the differences between CVII and LVII soldiers' scores were less than 7 percentage points. The lowest scores for both groups were for the Vehicles factor (56% for CVII and 55% for LVII); the remaining scores ranged from 61 percent to 74 percent for CVII soldiers, and from 60 percent to 79 percent for LVII soldiers.

Task Factor (otherwise known as CVBITS) scores had been used in the performance modeling exercises conducted for CVI and LVI; however, Task Construct scores (i.e., MOS-Specific and General) were used for this purpose in CVII. Although Task Factors preserve somewhat more information than the more highly aggregated Task Construct scores, they have the disadvantage of differing across MOS as to the availability of each of the six scores (e.g., no Vehicles (V) score can be computed for several MOS). This problem is compounded by the considerably smaller sample sizes available for the two second-tour data collections relative to the two first-tour data collections. Moreover, in both CVI and LVI, the Technical (T) Task Factor score invariably loaded on the Core Technical Proficiency performance construct while the other five Task Factor scores invariably loaded on the General Soldiering Proficiency performance construct. Therefore, the two Task Construct scores were selected for use in the LVII performance modeling exercise described in Chapter 5. Because all "General" tasks are central to MOS 11B, only one Task Construct score was constructed for this MOS.

As mentioned above, means and standard deviations for the job knowledge and hands-on Task Construct scores are provided in Appendix B. Calculated across MOS, split-half reliability estimates (corrected to the number of items) were .79 for the General job knowledge score and .68 for the MOS-

Specific job knowledge score. Only a total score reliability estimate was calculated for the hands-on tests as it was not possible to derive equivalent halves for the two subscores in each MOS. This split-half reliability estimate (corrected to the number of tasks) was .59. Given the variability in test content across tasks, these estimates are reasonable.

PERFORMANCE RATING SCALES

As reported previously (Campbell, 1989), the dimensions covered by the second-tour rating scales (with the exception of the Combat Performance Questionnaire) were grounded in an analysis of second-tour jobs. The scale anchors were developed by revising and adapting rating scales developed for first-tour soldiers. Based on the CVII data analyses, additional minor modifications were made to these three sets of scales: the Army-Wide ratings, the MOS-Specific ratings, and the Combat Performance Prediction scales.

Army-Wide Rating Booklet. The Army-Wide rating booklet included 12 behavior-based dimensions, seven task-based leadership dimensions, a rating of overall effectiveness, and a rating of senior NCO potential. To construct this booklet, the first-tour Army-wide behavior-based dimensions were first modified for the CVII sample on the basis of additional samples of critical incidents (Campbell, 1988) to reflect the somewhat different job performance requirements and increased supervisory responsibilities of second-tour soldiers. Seven task-based leadership dimensions were also added on the basis of extensive job analyses of second-tour MOS conducted prior to CVII. These seven task-based dimensions, in addition to three of the behavior-based dimensions, were intended to assess important aspects of leadership or supervision.

Raters in the CVII sample tended to make frequent use of the highest rating scale values when evaluating the performance of second-tour soldiers. This suggested that the rating scale behavioral anchors may have been too lenient for more experienced soldiers (e.g., the behaviors depicted in the moderate range of the rating scale actually reflected relatively low-level performance). To offset this tendency in the LVII sample, the behavioral anchors for most rating dimensions were revised somewhat to make the scale values reflect a slightly higher level of performance than was the case in the CVII research.

MOS-Specific Rating Booklets. The MOS-Specific rating booklets included from 7 to 14 technically oriented behavior-based dimensions and a rating of overall MOS effectiveness. They were developed with the same procedure used for the Army-wide ratings. A set of scales suitable for second-tour MOS 19K soldiers were developed by adapting the second-tour MOS 19E scales that had been used in CVII. For all scales, the behavior-based dimensions were the same as those used in the CVII research which, in turn, were similar in nature to the dimensions used for first-tour soldiers. In five of the nine MOS, one or two of the MOS-specific dimensions measured some aspect of leadership (e.g., Leading the Team for MOS 11B). As with the Army-wide rating dimensions, the CVII behavioral anchors for most MOS-specific rating dimensions were revised to reflect slightly higher levels of performance. The names of all of the second-tour Army-wide and MOS-specific rating dimensions are presented in Appendix C.

Combat Performance Prediction Scales. The Combat Performance Prediction Scales consisted of 14 items which depict examples of soldier behaviors under

combat conditions. The rater's task was to estimate the likelihood that the ratee would behave as described in the behavioral example. Ratings were made on a 7-point scale ranging from very likely to very unlikely. The items were a subset of the 40 items that appeared on the original CVI version of the Combat Performance Prediction Scales. Unlike the LVI/CVII data collections, LVII Combat Performance Prediction Scale ratings were collected for both male and female soldiers.

Rater Training

An extremely important aspect of each rating session was a rater orientation and training program developed to reduce various rating errors (e.g., halo) and to persuade raters to provide evaluations that were as accurate as possible. The orientation/training program used in LVI, CVII, and LVII was an adaptation of the program developed for raters participating in the CVI data collection (Pulakos & Borman, 1986).

Summary of Ratings Data

Table 3.5 shows, by MOS, the number of supervisors who provided ratings for each member of the LVII sample. Across all nine MOS, two or more ratings were obtained for 75 percent of the soldiers (1,194 of 1,595) and at least one rating was obtained for 94 percent of the sample (1,494 of 1,595). The soldiers who received ratings averaged 1.82 raters per ratee. These figures pertain to the Army-Wide rating booklet; for one reason or another, raters were not always able to complete the MOS-specific and Combat Performance Prediction booklets.

Rater Familiarity With Ratees

Supervisors who made ratings were asked to report how familiar they were with the ratees' job performance. Frequencies were computed based on their answers to these questions.

Table 3.6 shows the self-reported familiarity of the raters with ratees' job performance. Most of the supervisors (89%) reported observing the ratees' performance at least several times each week for one month or more. These data suggest that the raters were sufficiently familiar with ratees' job performance to provide accurate ratings. Note also that supervisors were not required to rate soldiers on aspects of performance which they believed they had had insufficient opportunity to observe.

Analysis Procedure

Substantive analyses for the Army-wide and Combat Performance Prediction Scale ratings were carried out on the total sample; MOS-specific ratings were, of course, analyzed separately by MOS. The first set of analyses for the Army-wide and MOS-specific rating scales focused on the distributions of the individual ratings (e.g., means and standard deviations) and reliability estimates. This was followed by principal factor analyses with varimax rotation to determine the composition of basic scores.

Analysis of the Combat Performance Prediction Scales began with principal factor analyses with varimax rotation to determine the composition of the basic score(s). This was followed by the computation of descriptive statistics and reliability estimates for the recommended composite score.

Table 3.5

Number of Raters Per LVII Ratee by MOS

	MOS										Total Sample	
	11B	13B	19K	31C	63B	71L	88M	91A/B	95B			
<u>Number of Ratings</u>												
0	21	10	16	1	4	7	1	12	3	75		
1	57	34	62	9	17	31	20	45	25	300		
2	262	131	89	59	171	117	67	162	136	1,194		
3	7	4	1	1	1	1	1	3	4	23		
4	0	0	0	0	1	1	0	0	0	2		
5	0	1	0	0	0	0	0	0	0	1		
LVII Sample Total N	347	180	168	70	194	157	89	222	168	1,595		
<u>Mean Number of Ratings Per Ratee</u>												
Total LVII Sample	1.73	1.74	1.45	1.86	1.89	1.73	1.76	1.70	1.84	1.74		
Rated Soldiers Only	1.85	1.84	1.60	1.88	1.93	1.81	1.78	1.80	1.87	1.82		

Table 3.6

Self-Reported Familiarity of LVII Raters With Ratees (Percent)

Length of Time Worked With Ratee	Opportunity to Observe Job Performance (on Average)				Total Sample
	Daily	Several Times Each Week	About Once a Week	Less Than Once a Week	
Less than one month	2.6	.8	.2	.9	4.5
1-3 months	11.0	4.3	.6	.8	16.7
4-6 months	16.6	5.2	1.2	.4	23.4
7-12 months	17.8	6.9	1.3	.4	26.4
More than 12 months	20.9	6.2	1.1	.8	29.0
Total Sample	68.9	23.4	4.4	3.3	100.0

Note. Sample size is 2,779.

Army-Wide Rating Scale Results

Descriptive Statistics. Table 3.7 displays the Army-wide rating distributions and demonstrates that raters tended to make less use of the highest values in the LVII sample, as compared to the CVII sample. This is probably a direct result of revising the behavioral anchors to reflect slightly more stringent performance standards. Supervisors tended to provide lower ratings on the leadership dimensions compared to the nonleadership dimensions, in both LVII and CVII.

These conclusions are supported by the data in Table 3.8 as well. The overall mean across the leadership-oriented dimensions is 4.38, compared to 4.89 for the nonleadership dimensions. Table 3.8 also indicates that ratings of the LVII soldiers on the nonleadership dimensions are somewhat lower than the corresponding ratings of CVII soldiers. Again, this is probably a direct result of revising the behavioral anchors. Ratings on these non-leadership dimensions are also higher in the LVII research than they were in the LVI research. This is a reasonable outcome, because second-tour soldiers should perform at a higher level on the technical part of the job compared with their first-tour counterparts. This outcome is particularly interesting in view of the fact that the anchors for the second-tour rating scales already reflect higher levels of performance than the corresponding first-tour anchors.

Table 3.7

LVII Army-Wide Rating Distributions: Use of Scale Points (Percent)

Dimension	7	6	5	4	3	2	1
<u>Behavior Scales</u>							
1. Tech Knowledge/Skill*	6	26	33	23	9	3	0
2. Effort*	11	25	25	19	13	7	1
3. Supervising	3	12	25	26	21	11	2
4. Following Regs/Orders*	12	26	29	16	10	6	2
5. Integrity*	18	28	25	13	8	5	2
6. Training/Development	5	17	26	24	19	8	2
7. Maintaining Equipment*	12	26	29	18	9	5	2
8. Physical Fitness*	20	20	26	16	11	5	2
9. Self-Development*	8	16	27	24	15	8	2
10. Consideration for Subord	9	24	31	21	12	3	1
11. Military Bearing*	15	22	30	16	9	6	1
12. Self Control*	18	26	23	15	11	5	2
<u>Task-Based Leadership Scales</u>							
13. Role Model	4	14	29	23	17	9	2
14. Communication	5	18	31	24	15	6	1
15. Personal Counseling	4	12	27	28	19	8	2
16. Monitoring	4	15	27	25	18	8	2
17. Organizing Missions/Opers	5	17	30	25	16	6	1
18. Personnel Administration	5	15	25	27	18	7	2
19. Performance Counseling	4	15	27	27	18	7	2
20. Overall Effectiveness	5	23	35	21	10	5	1
21. Senior NCO Potential	8	24	26	18	14	8	3
LVII Mean Non-Supervisory*	13.33	23.89	27.44	17.78	10.56	5.56	1.56
LVII Mean Supervisory	4.80	15.90	27.80	25.00	17.30	7.30	1.70
CVII Mean Non-Supervisory*	16.89	25.22	27.00	15.44	9.11	5.00	1.33
CVII Mean Supervisory	6.90	18.00	28.20	22.00	15.10	8.10	1.70

Note. LVII sample sizes range from 2,592 to 2,798 for the behavior scales and from 2,432 to 2,744 for the task-based leadership scales; CVII sample sizes range from 1,602 to 1,732 for the behavior scales and from 1,502 to 1,654 for the task-based leadership scales.

* Indicates non-supervisory scales.

Table 3.8

LVII Army-Wide Ratings: Dimension-Level Means and Standard Deviations

<u>Dimension</u>	<u>Mean^a</u>	<u>SD</u>
<u>Behavior Scales</u>		
1. Technical Knowledge/Skill*	4.85	1.02
2. Effort*	4.79	1.27
3. Supervising	4.09	1.21
4. Following Regs/Orders*	4.90	1.25
5. Integrity*	5.11	1.28
6. Training/Development	4.35	1.19
7. Maintaining Equipment*	4.95	1.17
8. Physical Fitness*	4.93	1.40
9. Self-Development*	4.48	1.26
10. Consideration for Subord	4.85	1.10
11. Military Bearing*	4.95	1.29
12. Self-Control*	5.05	1.28
<u>Task-Based Leadership Scales</u>		
13. Role Model	4.28	1.23
14. Communication	4.55	1.11
15. Personal Counseling	4.24	1.15
16. Monitoring	4.32	1.17
17. Organizing Missions/Operation	4.47	1.24
18. Personnel Administration	4.34	1.20
19. Performance Counseling	4.30	1.16
20. Overall Effectiveness	4.75	1.09
21. Senior NCO Potential	4.62	1.35
LVII Mean Non-Supervisory*	4.89	1.25
LVII Mean Supervisory	4.38	1.18
CVII Mean Non-Supervisory*	5.08	1.25
CVII Mean Supervisory	4.50	1.23
LVI Overall Mean	4.42	1.51

Note. Sample size ranges from 1,437 to 1,538 for LVII, from 857 to 927 for CVII, and from 9,907 to 9,928 for LVI. CVII and LVI means and SDs based on supervisor ratings only.

* Indicates non-supervisory scales.

^a On a scale in which 7 = Highest and 1 = Lowest.

The differentiation across rates is indicated by the standard deviations in Table 3.8. On average, the ratings showed somewhat less differentiation for these second-tour soldier ratings than was the case for the first-tour soldier ratings (LVI sample). However, the ratings of LVII soldier performance have about the same degree of variability as did the ratings of CVII soldiers.

Overall, the LVII rating distributions seem appropriate. The means are about where expected, and the variability of the ratings is sufficient to reveal relationships between these ratings and other variables.

Reliability Estimates. Army-wide dimension-level interrater reliability results are presented in Table 3.9. This table contains intraclass correlations that reflect the reliability of a single rater and the reliability of the mean rating across all raters. The latter intraclass correlations depend in part on the average number of raters per ratee.

First, Table 3.9 shows that the degree of interrater reliability for the LVII ratings is almost exactly the same as was found in the LVI and CVII research. Second, the task-based leadership dimensions are slightly less reliable than the behavior-based rating dimensions, but they are still quite reliable. Third, the mean ratings in the LVII sample have about the same level of reliability as the mean ratings in the LVI and CVII samples.

Factor Analysis Results. Several factor analyses were conducted on the LVII sample. Army-wide ratings on the nine second-tour nonleadership dimensions were intercorrelated and factor analyzed so that the LVI and LVII factor structures could be compared. Then, the ratings on the 10 leadership dimensions for the LVII sample were intercorrelated and factor analyzed to assess the possibility of multiple underlying leadership/supervision factors. Finally, the same procedure was followed for all 19 of the Army-wide dimensions. For this analysis, the factor structure obtained in the LVII sample was compared to the factor structure obtained in the CVII sample.

The striking similarity of the rotated factor structures for the nine nonleadership/supervision dimensions that are common to the first-tour and second-tour rating scales is shown in Table 3.10. The three factors obtained in the LVI sample were closely replicated with the LVII data.

Factor analysis of the 10 supervisory dimensions resulted in a single leadership/supervision factor. Consequently, these results are not presented.

The four-factor rotated solutions obtained in the LVII and CVII samples are shown in Table 3.11. The two solutions are very similar. Both include three factors that are quite similar to the three LVI factors, plus a separate leadership/supervision factor.

Table 3.9

LVII Army-Wide Ratings: Dimension-Level Interrater Reliability Results

Dimension	Single Rater (r_{11})	N-Rater (r_{kk}) ^a
<u>Behavior Scales</u>		
1. Technical Knowledge/Skill	.36	.51
2. Effort	.45	.60
3. Supervising	.40	.53
4. Following Regs/Orders	.37	.52
5. Integrity	.35	.49
6. Training/Development	.31	.44
7. Maintain Equipment	.26	.39
8. Physical Fitness	.50	.64
9. Self-Development	.41	.56
10. Consideration for Subord	.30	.43
11. Military Bearing	.46	.61
12. Self-Control	.29	.43
<u>Task-Based Leadership Scales</u>		
13. Role Model	.44	.58
14. Communication	.32	.46
15. Personal Counseling	.28	.39
16. Monitoring	.32	.45
17. Organizing Missions/Operations	.29	.41
18. Personnel Administration	.31	.43
19. Performance Counseling	.29	.41
20. Overall Effectiveness	.44	.59
21. Senior NCO Potential	.46	.61
LVII Median for Behavior Scales	.37	.52
LVII Median for Task Leadership Scales	.31	.43
CVII Median for Behavior Scales	.36	.51
CVII Median for Task Leadership Scales	.33	.47
LVI Overall Median	.38	.52

Note. The total number of ratings used to compute interrater reliabilities ranges from 2,432 to 2,798 for LVII, from 1,495 to 1,735 for CVII, and from 9,907 to 9,928 for LVI. The average number of ratings per ratee is 1.78 for LVII, 1.52 for CVII, and 1.79 for LVI. CVII and LVI figures are based on supervisor ratings only.

^a k is the mean number of ratings per ratee.

Table 3.10

Comparison of LVI and LVII Factor Analysis^a Results: Non-Supervisory Dimensions

Dimension	Factor Loadings (LVI/LVII)		
	1	2	3
Leadership	<u>.67</u> /---	.31/---	.36/---
Technical Knowledge/Skill	<u>.67</u> /.70	.27/.23	.32/.28
Effort	<u>.66</u> /.73	.44/.34	.28/.27
Self-Development	<u>.53</u> /.52	.34/.31	.44/.44
Maintain Equipment	<u>.50</u> /.49	.37/.29	.40/.33
Following Regs/Orders	.41/.41	<u>.68</u> /.65	.29/.29
Self-Control	.19/.16	<u>.61</u> /.55	.22/.19
Integrity	.44/.49	<u>.62</u> /.57	.28/.26
Military Bearing	.33/.27	.35/.33	<u>.54</u> /.57
Physical Fitness	.24/.22	.18/.16	<u>.47</u> /.53
Percent Common Variance	41.6/42.7	34.5/31.6	24.0/25.7

Note. Sample size is 9,728 for LVI and 1,521 for LVII. LVI analyses based on supervisor ratings only.

^a Principal factor analysis, varimax rotation.

Table 3.11

Comparison of LVII and CVII Army-Wide Factor Analysis^a Results:
All Dimensions

Dimension	Factor Loadings (LVII/CVII)			
	1	2	3	4
1. Technical Knowledge/Skill	.47/.41	.23/.24	.26/.22	<u>.56/.65</u>
2. Effort	.45/.39	.34/.31	.26/.27	<u>.58/.68</u>
3. Supervising	<u>.63/.57</u>	.22/.21	.24/.28	.42/.53
4. Following Regs/Orders	.32/.29	<u>.63/.63</u>	.29/.30	.31/.36
5. Integrity	.38/.32	<u>.58/.66</u>	.24/.22	.34/.37
6. Training/Development	<u>.60/.52</u>	.20/.24	.27/.27	.38/.52
7. Maintain Equipment	.36/.32	.27/.33	.32/.25	.38/.50
8. Physical Fitness	.17/.20	.14/.18	<u>.53/.60</u>	.16/.19
9. Self-Development	.48/.41	.29/.27	.41/.44	.32/.48
10. Consideration for Subord	<u>.61/.47</u>	.40/.44	.16/.26	.28/.40
11. Military Bearing	.26/.30	.32/.34	<u>.62/.63</u>	.12/.22
12. Self-Control	.16/.17	<u>.57/.56</u>	.20/.18	.07/.09
13. Role Model	.51/.53	.37/.40	.56/.51	.25/.31
14. Communication	<u>.61/.62</u>	.39/.34	.22/.23	.26/.35
15. Personal Counseling	<u>.74/.72</u>	.24/.31	.27/.26	.11/.19
16. Monitoring	<u>.68/.63</u>	.18/.31	.30/.22	.30/.41
17. Organizing Missions/Operations	<u>.66/.70</u>	.22/.26	.27/.20	.30/.36
18. Personnel Administration	<u>.65/.63</u>	.28/.20	.22/.24	.17/.29
19. Performance Counseling	<u>.72/.72</u>	.22/.20	.23/.29	.24/.32
Percent Common Variance	45.3/37.6	25.4/20.3	18.2/16.9	16.9/25.3

Note. Sample size is 1,388 for LVII and 823 for CVII. CVII analyses based on supervisor ratings only.

^a Principal factor analysis, varimax rotation.

Basic Scores. Factor analyses of the Army-wide ratings suggest that the four-factor CVII solution can be replicated in the present data. Accordingly, the four composites shown in Table 3.12 and the overall effectiveness rating were used to summarize the LVII Army-wide rating data. The composite scores are identical to the CVII Army-wide rating composites. As in the CVII research, each dimension in a composite was unit weighted. Definitions for each of the composites are shown in Table 3.13.

Table 3.12

Composition of LVII Army-Wide Rating Composites

Percent Common Variance Accounted for by Relevant Factor	Composite Name	Dimensions Included
45.3	1. Leading/Supervising	Supervising Training/Development Consideration for Subord Communication Personal Counseling Monitoring Organizing Missions/Opers Personnel Administration Performance Counseling
25.4	2. Personal Discipline	Following Regs/Orders Integrity Self-Control
16.9	3. Technical Skill/Effort	Technical Knowledge/Skill Effort Maintain Equipment
18.2	4. Physical Fitness/ Military Bearing	Military Bearing Physical Fitness

Note. Two dimensions were not included in any composites: Acting as a Role Model and Self-Development.

Table 3.13

Definitions of LVII Army-Wide Rating Composites

Leading/Supervising:

Effectively organizing, monitoring, and, when necessary, correcting subordinates; providing proper training experiences; communicating effectively to keep subordinates and superiors informed and providing support and help to subordinates when needed.

Technical Skill/Effort:

Displaying technical knowledge and skill in accomplishing job tasks and completing assignments; showing conscientiousness and initiative on the job and exerting considerable effort to get jobs and tasks done effectively.

Personal Discipline:

Adhering to Army rules and regulations; exercising self-control; demonstrating integrity in day-to-day behavior; not causing disciplinary problems.

Physical Fitness/Military Bearing:

Maintaining an appropriate military appearance and bearing and staying in good physical condition.

The interrater reliabilities of the four Army-wide composites are shown in Table 3.14. The reliabilities tend to be slightly lower than the reliabilities for the same composites in the CVII sample. This is due in part to the slightly smaller average number of ratings per ratee in the LVII sample. Even though the reliabilities are slightly lower in the LVII sample, they are high enough for the rating factors to be quite useful as criterion measures.

Correlations among the four Army-wide composites are presented in Table 3.15. LVII correlations are very similar to those obtained in CVII. Although some of these correlations are quite high, prior results from CVII indicate that differentiation between these LVII composites should be sufficient to provide multidimensional performance information.

Table 3.14

Interrater Reliability Results for CVII and LVII Army-Wide Rating Composites

	Leading/ Supervising	Technical Skill/ Effort	Personal Discipline	Fitness/ Bearing
<u>LVII Ratings</u>				
r_{11}	.45	.46	.44	.51
r_{kk}^a	.58	.60	.58	.66
Average Ratings Per Ratee	1.68	1.79	1.81	1.82
Mean Rating ^b	4.39	4.86	5.02	4.97
Standard Deviation	.95	.99	1.07	1.16
Sample Size	1,427	1,521	1,537	1,537
<u>CVII Supervisor Ratings</u>				
r_{11}	.50	.48	.45	.56
r_{kk}^a	.64	.63	.60	.70
Average Ratings Per Ratee	1.75	1.86	1.86	1.86
Mean Rating ^b	4.51	5.04	5.16	5.18
Standard Deviation	1.01	1.03	1.09	1.17
Sample Size	857	918	920	925

Note. The total number of ratings used to compute reliabilities ranges from 2,385 to 2,792 for LVII and from 1,485 to 1,725 for CVII. CVII analyses based on supervisor ratings only.

^a k is the average number of ratings per ratee.

^b On a scale in which 7 = Highest and 1 = Lowest.

Table 3.15

Intercorrelations Among LVII and CVII Army-Wide Rating Composites

	Leading/ Supervising	Technical Skill/ Effort	Personal Discipline	Fitness/ Bearing
<u>Based on LVII Ratings</u>				
Leading/Supervising	1.00			
Technical Skill/Effort	.80	1.00		
Personal Discipline	.68	.66	1.00	
Fitness/Bearing	.54	.49	.52	1.00
<u>Based on CVII Supervisor Ratings</u>				
Leading/Supervising	1.00			
Technical Skill/Effort	.81	1.00		
Personal Discipline	.68	.67	1.00	
Fitness/Bearing	.60	.56	.55	1.00

Note. Sample sizes used to compute the intercorrelations range from 1,427 to 1,538 for LVII and from 852 to 919 for CVII.

MOS-Specific Rating Scale Results

Descriptive Statistics. Table 3.16 presents the means and standard deviations of the MOS-specific ratings for each MOS. Results are shown separately for the leadership- and nonleadership-oriented dimensions. In general, the means and standard deviations of these ratings are quite similar for the LVII and CVII samples and the means are somewhat higher than those for the Army-wide dimensions. The unweighted mean across MOS for the MOS-specific ratings is 5.07, whereas the mean across the Army-wide dimensions is 4.66.

Reliability Estimates. Interrater reliabilities for the MOS-specific scales are presented in Table 3.17. The single-rater MOS dimension reliabilities are generally lower than the single-rater Army-wide dimension reliabilities. Moreover, the single-rater reliabilities in the LVII sample tend to be somewhat lower than the single-rater reliabilities in the CVII sample. Reliabilities of the mean ratings across raters are of course higher than the single-rater estimates.

Table 3.16

MOS-Specific Ratings: LVII and CVII Means (Across Rating Dimensions) of Dimension Means* and Standard Deviations*

	11B	136	19K	31C	636	71L	88M	91A/B	95B
<u>LVII Ratings</u>									
Non-Sup. Dimensions									
Mean	5.16 (11)	5.27 (10)	5.25 (8)	5.03 (7)	4.69 (12)	5.13 (7)	5.12 (11)	5.17 (9)	4.96 (13)
SD	1.07	.98	1.10	1.15	1.08	1.11	1.08	1.13	1.03
Supervisory Dimensions									
Mean	4.83 (2)	5.01 (1)	5.10 (1)	4.91 (1)	--	--	--	--	4.54 (1)
SD	1.15	1.15	1.27	1.42	--	--	--	--	1.12
Overall Rating									
Mean	5.22	5.23	5.25	5.10	4.94	5.28	5.28	5.21	4.96
SD	1.09	.98	1.15	1.25	1.11	1.02	1.05	1.11	.91
Number of Ratees	236-323	151-165	144-149	59-66	198-194	135-150	70-87	149-209	134-165
<u>CVII Supervisor Ratings</u>									
Non-Sup. Dimensions									
Mean	5.18 (11)	5.28 (10)	5.57 (8)	5.08 (7)	4.97 (12)	5.17 (7)	5.40 (11)	5.37 (9)	5.17 (13)
SD	1.10	1.08	.94	1.19	1.15	1.19	1.14	1.07	1.06
Supervisory Dimensions									
Mean	4.61 (2)	4.96 (1)	5.52 (1)	4.93 (1)	--	--	--	--	4.87 (1)
SD	1.23	1.38	1.65	1.35	--	--	--	--	1.15
Number of Ratees	98-103	120-135	39-41	73-78	86-88	70-84	115-121	73-85	105-123

Note. Values in parentheses represent number of dimensions used in computing mean ratings.

* Rated on a scale in which 7 = Highest and 1 = Lowest.

Table 3.17

LVII MOS-Specific Ratings: Dimension Interrater Reliability Results

	11B	13B	19K	31C	63B	71L	88M	91A/B	95B
<u>Number of Rating Dimensions</u>	13	11	9	8	12	7	11	9	14
<u>Dimension Reliabilities (r_{kk})^b</u>									
Median Dimension Reliability	.31	.15	.30	.38	.24	.30	.25	.37	.21
Range Across Dimensions	.20-.48	.00-.37	.19-.55	.30-.61	.11-.29	.25-.38	.00-.48	.14-.42	.00-.42
Reliability of Overall Rating	.52	.22	.52	.57	.44	.33	.18	.42	.34
Average Ratings Per Ratee ^a	1.77	1.78	1.53	1.82	1.89	1.78	1.75	1.65	1.67
<u>Dimension Reliabilities (R_{kk})^b</u>									
Median Dimension Reliability	.44	.23	.39	.53	.37	.43	.37	.48	.30
Range Across Dimensions	.31-.62	.00-.51	.26-.64	.44-.73	.19-.43	.38-.52	.00-.62	.00-.62	.22-.55
Reliability of Overall Rating	.65	.34	.63	.71	.60	.47	.28	.55	.48
Average Ratings Per Ratee ^a	1.77	1.78	1.53	1.82	1.89	1.78	1.75	1.65	1.67

Note. The total number of ratings used to compute reliabilities for each MOS ranges from 103 to 586.

^a Mean across all dimensions.

^b k is the average number of ratings per ratee.

Factor Analysis Results. As in the first-tour analyses, factor analyses of MOS-specific rating data within MOS revealed that a single factor can account for the vast majority of the variance in the MOS-specific ratings. Rotation of additional factors yielded solutions that were difficult to interpret. Thus, none of these solutions are presented here.

Basic Scores. Because the factor analysis results did not indicate multiple factors for any of the MOS-specific rating analyses, a unit-weighted composite of all dimension ratings for each MOS was constructed. This is identical to the scoring system used in CVII, and yields comparable reliability estimates (see Table 3.18). Note also that the single-rater reliabilities of the MOS rating composites are comparable to the single-rater reliabilities of the Army-wide dimensions.

Combat Performance Prediction Scale Results

Factor Analysis Results. Results of the principal components analysis on the combined LVII sample confirmed the findings that were obtained in LVI and CVII. Specifically, two factors were identified; however, the second factor was simply a reflection of the first (i.e., it was comprised of the negatively worded items). Therefore, the factor loadings are not presented here.

Basic Score. Because the factor analysis again indicated only one substantive factor, the 14 items were summed to form a single composite score. This scoring system was used in the LVI and CVII research as well.

Descriptive Statistics. The mean LVII Combat Performance Prediction scale composite score is 70.67 with a standard deviation of 12.44 (N = 1,483), indicating a reasonable degree of variability in these ratings. This is virtually identical to the mean of the supervisor ratings of soldiers in the CVII sample (mean = 70.82, SD = 12.57, N = 814). This is true even though the CVII sample did not include female soldiers, and LVII female soldiers tended to receive lower scores than males (mean of 64.73 compared to 71.60). Moreover, second-tour soldiers scored higher than first-tour soldiers on these scales, with the LVI sample having a mean score of 63.30 (SD = 13.65, N = 8,713).

Reliability Estimates. Interrater reliability estimates for the LVII and CVII ratings are provided in Table 3.19. The LVII estimates compare favorably with the CVII estimates. Furthermore, the estimates are comparable to those obtained for the Army-wide composite scores.

Coefficient alpha, an index of internal consistency, was also computed for the composite score. Again, the findings are comparable with CVII. Coefficient alpha was .929 for the LVII sample, compared to .926 for the CVII supervisor ratings. Thus, both the interrater reliability and internal reliability estimates associated with the Combat Performance Prediction Scales are reasonably high.

Table 3.18

MOS-Specific Ratings: Composite Interrater Reliability Results for LVII and CVII

	11B	13B	11	9	31C	638	71L	88H	91A/B	95B
<u>Number of Rating Dimensions</u>	13	11	11	9	8	12	7	11	9	14
<u>LVII Composite</u>										
r_{11}	.49	.18	.18	.49	.57	.31	.39	.27	.44	.28
r_{kk}^a	.63	.28	.28	.59	.71	.46	.53	.39	.55	.36
Average Ratings Per Ratee ^b	1.77	1.78	1.78	1.53	1.82	1.89	1.78	1.75	1.65	1.67
<u>CVII Composite</u>										
r_{11}	.45	.37	.37	.40	.74	.43	.53	.57	.34	.40
r_{kk}^a	.61	.52	.52	.54	.83	.54	.65	.57	.47	.57
Average Ratings Per Ratee ^b	1.93	1.90	1.90	1.78	1.73	1.58	1.59	1.90	1.68	1.96

Note. The total number of ratings used to compute reliabilities for each MOS ranges from 103 to 586. CVII analyses based on supervisor ratings only.

^a k is the average number of ratings per ratee.

^b Mean across all dimensions.

Table 3.19

Interrater Reliability Results for Combat Performance Prediction Scales Score
for LVII and CVII

	LVII Ratings	CVII Supervisor Ratings
r_{11}	.463	.423
r_{kk}^a	.610	.575
Average Ratings Per Ratee	1.82	1.84
Mean Rating ^b	70.67	70.82
Standard Deviation	12.44	12.57
Sample Size	1483	8471

Note. The total number of ratings used to compute reliabilities is 2,698 for LVII and 1,501 for CVII.

^a k is the average number of ratings per ratee.

^b Maximum possible score is 98.

ADMINISTRATIVE MEASURES: THE PERSONNEL FILE FORM

The LVII Personnel File Form was used to gather self-reports of archival/administrative information dealing with personnel actions reflective of individual performance. The first-tour versions (CVI and LVI) of the PFF requested information regarding (a) evidence of exemplary performance, including awards and memoranda/certificates of appreciation, commendation, and achievement; (b) receipt of disciplinary actions (i.e., Articles 15 and flag actions); and (c) test results, including Physical Readiness test scores, individual weapon qualification scores, and Skill Qualification Test scores.

The original second-tour version of the PFF developed for CVII included these same types of variables and added others. The additional items were related to education (military training and civilian college courses) and promotions (e.g., how often recommended for accelerated promotion, number of promotion board points received). Another modification was to distinguish between awards, memoranda, and disciplinary actions received while in grades E-1 through E-3 and those received while in grades E-4 and above.

Before being administered to the LVII sample, the second-tour PFF was revised in several minor ways. Most of these revisions were intended to increase the interpretability/accuracy of responses and to reduce the amount of missing data. For example, the PFF response format was changed so that soldiers could indicate if they had earned more than one Army Achievement Medal.

Item Scoring and Analysis Procedure

The first set of analyses examined the extent of missing and invalid data for individual variables included on the PFF and the amount of variance associated with them. Next, tentative basic scores modeled on the content of the CVII basic scores were constructed. Descriptive statistics and score intercorrelations were then computed to evaluate the psychometric properties of these basic scores.

Because of the diverse nature of the items on the PFF and the reliance on the CVII scoring system as a starting point, analyses leading to the development of tentative basic scores will be discussed for each portion of the PFF in turn.

Positive Recognition Items

Awards that soldiers might have received were listed in a checklist format on the PFF. Additionally, PFF respondents indicated how many memoranda and certificates of appreciation, commendation, or achievement they had received while in grade E-4 and above (i.e., while in second tour). The distribution of responses to these items was similar to that found with the CVII sample, so a composite score was constructed in the same manner as in CVII. The scoring algorithm makes use of the NCO promotion board process which differentially weights awards. For example, an Army Achievement Medal receives three times as much as an Air Assault Badge.

The use of this weighting scheme in CVII increased the variability associated with the resulting composite score and appeared to reflect the relative job performance of the soldiers more accurately than did the unit-weighted approach used in CVI. Thus, an "Awards and Certificates" composite score was constructed by summing the weighted awards with the number of memoranda and certificates received while in grades E-4 and above.

Disciplinary Action Items

Both an Article 15 (a disciplinary action) and a Flag Action (suspension of a favorable personnel action) are considered to be indices of poor soldier performance. As with the memoranda and certificates, soldiers were asked to indicate how many Articles 15 and Flag Actions they had received while at different paygrades. Examination of the distribution of responses indicated that the scoring scheme previously used for these items would be appropriate. Thus, a disciplinary action composite score was constructed by summing the number of Articles 15 and Flag Actions received while in grades E-4 and above.

Test Scores

Soldiers are periodically administered physical fitness and marksmanship tests. As with previous data collections, the Physical Readiness test score and the individual weapon (usually M16) qualification score exhibited reasonable degrees of variability, and appeared to cover important aspects of soldier performance. Thus, they were tentatively identified as basic scores. The Skill Qualification Test score was not used as a basic score in CVII or LVII because of problems with incomplete data.

Education

The second-tour PFF included a checklist of military training courses; respondents were also asked to indicate how many hours of college courses they had successfully completed. The military training items were not used in LVII because (a) several of the courses were not comparable and there was insufficient information to weight them appropriately, and (b) the training course composite developed in CVII did not correlate with scores from any of the other criterion measures.

Although the college course response format was changed from the CVII version to improve the accuracy of responses, examination of the response distributions suggested that the data were still questionable. Since there was no way to definitively assess the accuracy of the information, it was once more not used in any basic scores. Thus, no education-related scores were generated for LVII soldiers.

Promotions

A promotion rate variable had been constructed for first-tour soldiers based on information in the Army's computerized Enlisted Master File. This was a grade deviation score in which each soldier's paygrade was adjusted to the mean of those who shared his or her time in service. The second-tour PFF also asked for other information related to promotions which could potentially be used to supplement the grade deviation score. This information was related to (a) the number of administrative and board points assigned at each promotion board appearance, and (b) whether the soldier had ever been recommended for an accelerated promotion.

Analysis of the CVII data indicated that information regarding promotion points was of limited usefulness because soldiers confused administrative points and board points. The relevant items were revised in an effort to make the LVII data more interpretable, but an inordinate percentage of invalid responses were still evident in the LVII data. Therefore, the information was not used.

Greater success was achieved with the accelerated promotion data. In CVII, the promotion rate score was a composite of the grade deviation score and a dichotomously scored accelerated promotion rate item. On the LVII form, soldiers indicated how many times they had been recommended for an accelerated promotion. Thus, the promotion rate composite for LVII was based on the number of accelerated promotion recommendations plus the grade deviation score.

Descriptive Statistics and Intercorrelations

Means and standard deviations for the administrative indices of performance are presented in Table 3.20. The corresponding descriptive statistics for CVII are not comparable for the Awards and Certificates score because of response format differences between the CVII and LVII instruments. Otherwise, the means and standard deviations for the LVII and CVII scores are very similar.

Table 3.20

Administrative Indices Descriptive Statistics for LVII and CVII

Measure		N	Mean	SD	Range
Awards and Certificates ^a	CVII	928	10.53	5.63	0-44
	LVII	1,577	14.69	6.79	0-40
Disciplinary Actions	CVII	930	.42	.87	0-8
	LVII	1,577	.37	.76	0-6
Physical Readiness Score	CVII	998	250.11	30.68	121-300
	LVII	1,522	248.81	31.27	23-288
Weapon Qualification	CVII	1,036	2.52	.67	1-3
	LVII	1,565	2.58	.67	1-3
Promotion Rate	LVII	1,513	100.00	7.79	61-128
Promotion Rate (CVII Scoring)	CVII	901	100.14	8.09	67-121
	LVII	1,513	99.98	7.48	57-121

^a Differences in LVII and CVII results reflect differences in PFF response format.

Correlations among the CVII and LVII basic scores are shown in Table 3.21. Again, the LVII results are generally similar to the CVII results. The correlations with promotion rate (not presented) tended to be a bit smaller when promotion rate was computed in a manner identical to CVII (i.e., using dichotomous scoring for accelerated promotion recommendations).

Table 3.21

Intercorrelations Among LVII and CVII Administrative Indices of Second-Tour Performance

Measure	Awards	Discipline	Physical	Weapon	Promotion
Awards and Certificates	1.00				
Disciplinary Actions	-.11/-.08	1.00			
Physical Readiness	.16/.13	-.14/-.11	1.00		
Weapon Qualification	.19/.14	-.04/-.03	.19/.11	1.00	
Promotion Rate	.26/.31	-.21/-.19	.16/.14	.16/.14	1.00

Note. LVII correlations are on the left, sample sizes range from 1,461 to 1,577; CVII correlations are on the right, sample sizes range from 817 to 1,035.

Basic Scores

The analyses reported herein suggest that the basic scores tentatively derived for the PFF satisfactorily capture the useful information on that form. Therefore, they were made available for use in the second-tour performance modeling exercise.

SITUATIONAL JUDGMENT TEST (SJT)

The SJT was designed to evaluate the effectiveness of NCO judgments concerning what the NCO should do in difficult supervisory situations. Thus, the SJT can be viewed as a job knowledge test pertaining to the leadership/supervision components of second-tour positions. For each SJT item, soldiers were asked to read a description of a difficult supervisory situation, examine three to five possible responses to the situation, then select the most and the least effective response alternatives. Figure 3.1 presents an item which is representative of the type of items that are included in the SJT.

You are a squad leader on a field exercise, and your squad is ready to bed down for the night. The tent has not been put up yet, and nobody in the squad wants to put up the tent. They all know that it would be the best place to sleep because it may rain, but they are tired and just want to go to bed. What should you do?

- a. Tell them the first four men to volunteer to put up the tent will get light duty tomorrow.
 - b. Make the squad sleep without tents.
 - c. Tell them that they will all work together and put up the tent.
 - d. Explain that you are sympathetic with their fatigue, but the tent must be put up before they bed down.
-

Figure 3.1. Example of a Situational Judgment Test type of item.

As reported previously (Campbell, 1989), development of the SJT involved asking groups of soldiers similar to the target NCOs (i.e., at the E-4 and E-5 level) to describe a large number of difficult but realistic situations that Army first-line supervisors face on their jobs. After a large number of these situations had been generated, a wide variety of possible actions (i.e., response alternatives) for each situation were gathered, and ratings of the effectiveness of each of these actions were collected from both experts (senior NCOs) and the target group (E-4 and E-5 NCOs). These effectiveness ratings were used to select situations and response alternatives to be included on the SJT.

The sample of subject matter experts (SMEs) was a group of 90 senior NCOs who were students and instructors at the Sergeants Major Academy. These NCOs were among the highest ranking enlisted soldiers in the Army (rank of E-8 or E-9), and they all had extensive experience as Army supervisors. For each situation, these NCOs rated the effectiveness of each response alternative on a 7-point scale (1 = least and 7 = most effective). Each NCO rated the response alternatives for a subset of the items that were included on the SJT; thus, about 25 expert judgments were available for each of the SJT items. The effectiveness ratings from this sample of experts were used to develop SJT scoring procedures.

The initial version of the SJT, which was administered to the CVII sample, consisted of 35 items. Results of the CVII data analyses were very encouraging. SJT scores showed an adequate amount of variability, and internal consistency reliabilities were moderately high. The SJT's highest zero-order correlations were with the job knowledge test scores, but its secondary correlations were with measures that compose the effort/leadership factor. Because the CVII data analysis results indicated that the SJT was a promising measure of supervisory performance, this test was lengthened for the LVII data collection to increase the internal consistency reliability and facilitate the identification of SJT subscores.

Because the 35-item SJT proved to be rather difficult for the CVII sample, an effort was made to select relatively less difficult additional items to include in the lengthened test. Difficulty was estimated using the g -ratios obtained in the original pilot testing activities described in Campbell (1989). Also, the content of the new items was intended to be similar to the content of the existing SJT items. To aid in this judgment, the original 35 items were item analyzed against both the SJT total score and the criterion factor scores, using CVII data. The new items were intended to be similar to the original items that both had similar correlations with other measures and had comparable item-total correlations with the SJT itself. A total of 14 new items were selected, and the resulting 49-item SJT was administered to the LVII sample.

Analysis Procedure

The data were first screened for invalid responses and incomplete data. The results of data screening are provided in Chapter 4. Next, frequency counts were conducted to determine whether there was variability across alternative responses for an item. If the correct answer was obvious, it would be impossible for SJT scores to discriminate among the LVII soldiers.

Development of Scoring Procedures

Procedures for scoring the SJT were identical to those used in CVII. Five different scores were computed. The most straightforward was a simple number correct score. For each item, the response alternative that was given the highest mean effectiveness rating by the experts (the senior NCOs at the Sergeants Major Academy) was designated the "correct" answer. Respondents were scored based on the number of items for which they indicated that the "correct" response alternative was the most effective.

The second scoring procedure involved weighting each response alternative by the mean effectiveness rating given to that response alternative by

the expert group. This gave respondents more credit for choosing "wrong" answers that are still relatively effective than for choosing wrong answers that are very ineffective. These item-level effectiveness scores for the chosen alternative were then averaged to obtain an overall effectiveness score for each soldier. Averaging item-level scores instead of summing them placed respondents' scores on the same 1 to 7 effectiveness scale as the experts' ratings and ensured that respondents were not penalized for missing data.

Scoring procedures based on respondents' choices for the least effective response to each situation were also used. Being able to identify the least effective response alternatives might be seen as an indication of the respondent's knowledge and skill for avoiding these very ineffective responses, or in effect, to avoid "screwing up." As with the choices for the most effective response, a simple number correct score was computed: the number of times each respondent correctly identified the response alternative that the experts rated the least effective. To differentiate it from the number correct score based on choices for the most effective response, this score will be referred to as the L-Correct score, and the score based on choices for the most effective response (described previously) will be referred to as the M-Correct score.

Another score was computed by weighting respondents' choices for the least effective response alternative by the mean effectiveness rating for that response, and then averaging these item-level scores to obtain an overall effectiveness score based on choices for the least effective response alternative. This score will be referred to as L-Effectiveness, and the parallel score based on choices for the most effective responses (described previously) will be referred to as M-Effectiveness.

Finally, a scoring procedure that involved combining the choices for the most and the least effective response alternative into one overall score was also examined. For each item, the mean effectiveness of the response alternative each soldier chose as the least effective was subtracted from the mean effectiveness of the response alternative they chose as the most effective. Because it is actually better to indicate that less effective response alternatives are the least effective, this score can be seen as a composite of the two effectiveness scores described previously (i.e., subtracting a negative number from a positive number is the same as adding the absolute values of the two numbers). Consequently, this is not a "difference" score but a simple sum. These item-level scores were then averaged together for each soldier to generate the fifth total score. This score will be referred to as M-L Effectiveness.

Each of these scores was computed twice for the LVII soldiers, once using all 49 SJT items and once including only the 35 SJT items that had been administered to the CVII sample as well. The 35-item SJT scores were computed for two reasons. First, these scores can be more directly compared with the SJT scores for the CVII sample because they are based on the same set of items. Second, these scores can be used to determine whether adding 14 items did, in fact, increase the internal consistency reliability of the SJT and decrease test difficulty.

Descriptive statistics and internal consistency reliabilities were computed for each of the five scoring procedures for both the 49-item and the 35-item versions of the SJT. Intercorrelations were computed among the five

scores generated by the five different scoring procedures for the 49-item SJT only. Finally, item analyses were conducted for each of the scoring procedures, again for the 49-item SJT only. These item analyses included the item-total correlations for all of the scoring procedures and also the proportion of the sample answering each item correctly for the M-Correct and L-Correct scoring procedures.

Development of Factor Score Composites

It is conceivable that what is measured by the SJT is not a single, unidimensional construct but rather several relatively distinct aspects of supervisory knowledge that underlie distinct components of supervisory performance. Distinct subconstructs, if they could be identified, would provide a better understanding of what is measured by the SJT. Such factors could provide the basis for developing SJT subscores. Since several scores are available for each of the supervisory simulation exercises (see next section for a description), it may be possible to more clearly delineate the supervisory aspects of the second-tour soldier job if several different scores could be identified for the SJT as well. It may even be possible to identify more than one component of supervisory/leadership performance in the overall latent structure of second-tour performance.

Efforts to identify distinct SJT subscores in analyses for the CVII sample were not particularly successful. Results of the CVII item-level factor analyses of the SJT failed to reveal any clearly defined factors and were for the most part uninterpretable. Some partially identifiable factors emerged in a few of these analyses that involved (a) disciplining when appropriate, (b) avoiding disciplining when inappropriate, and (c) assigning work tasks effectively, but the content of these factors was not very distinct.

The LVII version of the SJT contained almost 40 percent more items, so it was conceivable that a more interpretable solution would emerge for the LVII data. In addition, a content analysis of the SJT items conducted by Hanson and Borman (in press), as part of a research program aimed at explicating the nature of the construct measured by the SJT, revealed some promising new SJT subscales. Thus, the dimensionality of the SJT for the LVII sample was investigated both rationally and empirically, with the primary goal to develop a set of more homogeneous SJT subscores.

The content analysis of the SJT by Hanson and Borman was aimed at identifying what each SJT item was measuring. In other words, the goal was to determine how the more effective response alternatives differ from the less effective response alternatives. This content analysis was based on the content of the item stems, the content of the response alternatives, and the effectiveness of the various response alternatives. For example, an SJT item stem might describe a subordinate who is performing poorly, the more effective response alternatives might involve giving that subordinate a second chance, and the less effective response alternatives might involve disciplining harshly. This item could be seen as tapping the ability to identify situations in which it is most effective to "avoid inappropriately harsh discipline." Eleven such content-based "item types" were identified that appeared to have potential for identifying relatively homogeneous subsets of SJT items.

Ratings were then obtained from five psychologists concerning which, if any, of these 11 item types captured the essence of each SJT item. Based on these ratings, most of the CVII SJT items were categorized according to the item type involved. Categories that contained no items were dropped and categories with just a few items were combined with other conceptually similar categories. This resulted in categorizing 32 of the 35 CVII SJT items into five item-type categories. Hanson and Borman (in press) provide more details concerning the development of these item-type categories. They also developed SJT subscales based on this categorization and reported the psychometric characteristics of these subscales in the CVII sample, as well as their correlations with other criterion measures.

The 14 new items in the LVII version of the SJT were not included in the Hanson and Borman research because it was based on the CVII data. Consequently, the current analysis involved categorizing the additional 14 SJT items into the 11 content categories, using their procedures. This resulted in a total of 43 of the 49 LVII SJT items categorized into seven item-type categories.

The item-level M-L Effectiveness scores for the LVII sample were then intercorrelated and factor analyzed using principal factor analysis, and between 2 and 12 factors were retained. Both orthogonal and oblique rotations of these factors were examined. The orthogonal rotation was varimax, and the oblique rotations were Promax (Hendrickson & White, 1964) and Harris-Kaiser case II (Harris & Kaiser, 1964). The item-type categories were used to interpret the results of these factor analyses.

Subgroup Analyses

Descriptive statistics were computed separately for soldiers from combat and noncombat MOS and for soldiers from each of the nine MOS included in the research. These analyses will provide information concerning whether the SJT is an equally appropriate measure of supervision for all nine MOS. Some of the participants in the SJT development workshops reported that supervision in combat MOS is somewhat different than supervision in noncombat MOS. For example, some of them reported that supervisors in combat MOS are expected to take a stricter approach to subordinate misconduct. If the "correct" answer to SJT items varies by MOS, this may be reflected in differences in the mean scores of soldiers from different MOS.

Results

Item-Level Frequencies

The item-level responses from the LVII sample were well distributed across the response alternatives for each item. For example, the percentage of respondents choosing the most popular response alternative for each item as the most effective ranged from 32 to 83, with a median of 53 percent. This suggests that the correct responses to SJT items were not at all obvious to the soldiers in this sample.

Descriptive Statistics for the Five Scoring Procedures

Table 3.22 presents descriptive statistics for the 35-item SJT for both the LVII and the CVII samples. This table includes the mean score for each of the five scoring procedures. The maximum possible for the M-Correct scoring procedure is 35 (i.e., all 35 items answered correctly). In the LVII sample, the mean M-Correct score for the 35-item SJT was only 17.51. The mean number of least effective response alternatives correctly identified by this group was only 15.64. The mean M-Correct score for the CVII sample was 16.52 and the mean L-Correct score was 14.86. Clearly the SJT was difficult for both the CVII and the LVII soldiers.

Table 3.22

Comparison of LVII and CVII Situational Judgment Test Data: Means, Standard Deviations, and Internal Reliabilities

<u>Scoring Method</u>	<u>N</u>	<u>Mean</u>	<u>SD</u>	<u>Coefficient Alpha</u>
<u>LVII 35-Item SJT</u>				
M-Correct ^a	1,580	17.51	4.11	.56
M-Effectiveness	1,580	4.99	.31	.64
L-Correct ^a	1,581	15.64	3.81	.48
L-Effectiveness ^b	1,581	3.47	.29	.65
M-L Effectiveness	1,580	1.53	.54	.72
<u>CVII SJT (35 items)</u>				
M-Correct ^a	1,025	16.52	4.29	.58
M-Effectiveness	1,025	4.91	.34	.68
L-Correct ^a	1,007	14.86	3.86	.49
L-Effectiveness ^b	1,007	3.54	.31	.68
M-L Effectiveness	1,007	1.36	.61	.75

^a Maximum possible score is 35.

^b Low scores are "better"; mean effectiveness scale values for L responses should be low.

In addition, two-tailed t-tests revealed that the LVII sample had significantly higher M-Correct ($t = 5.93$, $p < .001$) and L-Correct ($t = 5.01$, $p < .001$) scores than did the CVII sample. Likewise, the LVII sample also scored significantly higher than the CVII sample on the M-L Effectiveness score ($t = 6.75$, $p < .001$).

These differences between the LVII and CVII samples may be, in part, a function of the level of supervisory training the soldiers in each sample had received. Sixty-two percent of the LVII sample reported having received at least basic supervisory training, whereas only 53 percent of the CVII sample had received such training. It may be that, because the LVII soldiers had more supervisory training than the CVII soldiers, they also had more supervisory job knowledge.

Table 3.22 also presents the standard deviation for each of the five scoring procedures. All of the scoring procedures resulted in a reasonable amount of variability in both the LVII and CVII samples. The internal consistency reliabilities for all of these scoring procedures are also acceptably high. The internal consistency reliabilities are very similar for the two samples. The most reliable score for both samples is M-L Effectiveness, probably because this score contains more information than the other scores (i.e., choices for both the most and the least effective responses).

Table 3.23 presents descriptive statistics for both the 35- and the 49-item versions of the SJT in the LVII sample. For the 49-item SJT, the maximum possible M-Correct score is 49. The mean in the LVII sample is only 25.84, indicating that this longer version of the SJT was also relatively difficult. However, there is some evidence to suggest that the additional 14 items did make the SJT easier. Two-tailed t-tests revealed that the 49-item SJT had a higher mean L-Effectiveness score ($t = 11.29$, $p < .001$) and a higher mean M-L Effectiveness score ($t = 4.87$, $p < .001$) than did the 35-item SJT. However, the difference between the 35-item and the 49-item M-Effectiveness scores was not significant.

Table 3.23 also presents the internal consistency reliabilities for both the 35- and the 49-item versions of the SJT for each of the five scoring procedures in the LVII sample. All of the scoring methods for both versions of the SJT have moderate to high internal consistency reliabilities. The most reliable score for both versions is M-L Effectiveness. In addition, the longer 49-item SJT (with the additional 14 items) did result in considerably higher reliabilities for all of the scoring methods.

In fact, the 49-item SJT is slightly more reliable than would be expected based on the number of items that were added. For example, based on the reliability of the 35-item SJT and using the Spearman-Brown prophecy formula (Cureton, 1965), a reliability of about .78 would be expected for the 49-item M-L Effectiveness score, but the obtained reliability for this score in the LVII sample was .81.

Table 3.23

Comparison of LVII 35-Item and 49-Item Situational Judgment Test Scores:
Means, Standard Deviations, and Internal Reliabilities

Scoring Method	N	Mean	SD	Coefficient Alpha
<u>LVII 35-Item SJT</u>				
M-Correct ^a	1,580	17.51	4.11	.56
M-Effectiveness	1,580	4.99	.31	.64
L-Correct ^a	1,581	15.64	3.81	.48
L-Effectiveness ^b	1,581	3.47	.29	.65
M-L Effectiveness	1,580	1.53	.54	.72
<u>LVII 49-Item SJT</u>				
M-Correct ^c	1,577	25.84	5.83	.69
M-Effectiveness	1,577	4.97	.32	.74
L-Correct ^c	1,577	22.35	5.14	.60
L-Effectiveness ^b	1,577	3.35	.29	.76
M-L Effectiveness	1,576	1.62	.57	.81

^a Maximum possible score is 35.

^b Low scores are "better"; mean effectiveness scale values for L responses should be low.

^c Maximum possible score is 49.

The intercorrelations among the scores obtained using the five different scoring procedures for the 49-item version of the SJT are shown in Table 3.24. These intercorrelations range from moderate to very high. Correlations between scores that are based on the same set of responses (e.g., M-Correct with M-Effectiveness) are higher than correlations between scores that are based on different sets of responses (e.g., M-Correct with L-Correct). The correlation between L-Effectiveness scores and the other scores is negative, because lower L-Effectiveness scores represent better performance. The high (negative) correlation between M-Effectiveness and L-Effectiveness seems to indicate that these two scores measure similar or related constructs.

Table 3.24

LVII 49-Item Situational Judgment Test: Score Intercorrelations for Various Scoring Methods

	M-Correct	M-Eff.	L-Correct	L-Eff.	M-L Eff.
M-Correct	1.00				
M-Effectiveness	.96	1.00			
L-Correct	.57	.61	1.00		
L-Effectiveness	-.68	-.73	-.88	1.00	
M-L Effectiveness	.89	.94	.79	-.92	1.00

Note. Sample sizes range from 1,576 to 1,577.

The median and range of the item-total correlations obtained using each of the scoring procedures for the 49-item SJT are shown in Table 3.25. These correlations are reasonably high, although there is quite a bit of variability across items. As would be expected, the scoring procedures that yield more internally consistent scores also have, on average, higher item-total correlations.

The proportion of the sample answering each item correctly was appropriate only for the M-Correct and L-Correct scoring procedures, and there was a great deal of variability in this measure of item difficulty across the SJT items. For the LVII sample, some items were answered correctly by as few as 14 percent of the sample and others by up to 84 percent. This large range of item difficulties is likely to be useful in discriminating among respondents across the entire range of SJT scores. The median proportion of the sample choosing the correct M and L responses was near .50 (.52 and .44, respectively).

Based on the descriptive statistics presented here, the M-Correct and L-Correct scores appear to have less desirable psychometric characteristics than the scores obtained using the other three scoring procedures. Further, the M-L Effectiveness score is the most reliable and, based on its high correlations with both the M-Effectiveness and the L-Effectiveness scores, appears to provide an adequate summary of the information contained in the SJT responses. Thus, the remaining analyses focus on the M-L Effectiveness score, which is hereafter referred to as the SJT Total Score.

Table 3.25

LVII 49-Item Situational Judgment Test: Summary of Item Analysis Results

Scoring Procedure	Corrected Item- Total Correlations ^a		Proportion Answering Items Correctly	
	Range	Median	Range	Median
M-Correct	-.08-.37	.17	.22-.84	.52
M-Effectiveness	.02-.38	.20	----	----
L-Correct	-.03-.32	.12	.14-.77	.44
L-Effectiveness	-.04-.37	.23	----	----
M-L Effectiveness	.01-.44	.27	----	----

^a This is the correlation between scores on a single item and scale scores computed using the other items in the set.

Development of Factor-Based Subscales

The factor pattern matrices for all three rotated factor solutions that were examined were remarkably similar. Where these solutions differed, the Harris-Kaiser solutions tended to be the most interpretable and also yielded factors that contained more nearly equal numbers of items. The eight-factor Harris-Kaiser solution was selected as the most interpretable. This solution also tended to converge with the item-type categories previously identified by the SMEs.

A set of "factor-based" SJT subscales were developed by rationally combining the item-type categorization with this factor analysis solution. Some of the factors had only a few items with high loadings and these factors were either dropped or collapsed with other factors. Items that did not load clearly on one particular factor were, where possible, assigned to scales based on their item-type categories. Those few items for which the item-type category and the factor pattern matrix clearly led to different conclusions were categorized based on their content and their correlations with the other items in the relevant factor-based subscales.

This process resulted in six factor-based subscales that contained between six and nine items each, and six remaining items that were not included in any subscale. Definitions of these factor-based subscales and the number of items included on each scale are presented in Table 3.26. Scores on these subscales were computed for soldiers in the LVII sample by averaging their item-level M-L Effectiveness scores for the items assigned to the subscales. Scores were not computed for soldiers who were missing more than 40 percent of the item-level scores for a particular subscale.

Table 3.26

Situational Judgment Test: Definitions of Factor-Based Subscales

1. Discipline soldiers when necessary (Discipline). This subscale is made up of items on which the most effective responses involve disciplining soldiers, sometimes severely, and the less effective responses involve either less severe discipline or no discipline at all. (Six items.)
2. Focus on the positive (Positive). This subscale is made up of items on which the more effective responses involve focusing on the positive aspects of a problem situation (e.g., a soldier's past good performance, appreciation for a soldier's extra effort, the benefits the Army has to offer). (Six items.)
3. Search for underlying reasons (Search). This subscale is made up of items on which the more effective responses involve searching for the underlying causes of soldiers' performance or personal problems rather than reacting to the problems themselves. (Eight items.)
4. Work within the chain of command and with supervisor appropriately (Chain/Command). For a few items on this subscale the less effective responses involve promising soldiers rewards that are beyond a direct supervisor's control (e.g., "comp" time). The remaining items involve working through the chain of command appropriately. (Six items.)
5. Show support/concern for subordinates and avoid inappropriate discipline (Support). This subscale is made up of items where the more effective response alternatives involve helping the soldiers with work-related or personal problems and the less effective responses involve not providing needed support or using inappropriately harsh discipline. (Eight items.)
6. Take immediate/direct action (Action). This subscale is composed of items where the more effective response alternatives involve taking immediate and direct action to solve problems and the less effective response alternatives involve not taking action (e.g., taking a "wait and see" approach) or taking actions that are not directly targeted at the problem at hand. (Nine items.)

The coefficient alpha internal consistency estimates for each subscale and their intercorrelations are presented on Table 3.27. The factor-based subscales demonstrate moderately high internal consistency reliabilities. This is especially encouraging considering that the subscales are comprised of relatively small numbers of items. The Spearman-Brown prophecy formula (Cureton, 1965) was used to estimate the reliability that would be expected if a 49-item test as reliable as the SJT was shortened to the number of items that are included in each of the subscales. The actual subscale reliabilities are considerably higher than these predicted reliabilities. For example, the reliability of the *Search for underlying reasons* subscale is .61, whereas the predicted reliability is only .44. This is evidence that the subscales are relatively homogeneous in content and that, minimally, they are more homogeneous than the total SJT.

Table 3.27

Situational Judgment Test: Score Intercorrelations for the Factor-Based Subscales and SJT Total Score

	Action	Chain/ Command	Discipline	Positive	Search	Support	SJT Total Score
Action	(.52)						
Chain/Command	.38	(.44)					
Discipline	.25	.13	(.44)				
Positive	.39	.35	.16	(.47)			
Search	.39	.37	.04	.40	(.61)		
Support	.46	.42	.13	.42	.48	(.61)	
SJT Total Score	.73	.61	.42	.64	.71	.76	(.81)

Note. Sample sizes range from 1505 to 1506; a correlation of about .10 is significant at the .01 level. Internal consistency reliabilities are presented on the diagonal in parentheses.

Correlations between the factor-based subscales and SJT Total Scores (M-L Effectiveness) can also be found in Table 3.27. The *Take immediate/direct action*, *Search for underlying reasons*, and *Show support/concern for subordinates* subscales correlate most highly with SJT Total Score (all correlations exceeding .70). *Discipline soldiers when necessary* has the lowest correlation with SJT total score ($r = .42$).

The intercorrelations among the subscales range from insignificant to moderately high. *Show support/concern for subordinates* correlates most highly with all of the other subscales except *Discipline soldiers when necessary*. It is interesting to note that the *Discipline soldiers when necessary* subscale has very low correlations with all of the other subscales. Its highest correlation is with *Take immediate/direct action* ($r = .25$). This is understandable, at least after the fact, because some supervisory situations require immediate disciplinary action and to take a "wait and see" attitude would be inappropriate.

Subgroup Analyses

The mean SJT Total Scores for soldiers in combat and noncombat MOS are shown in Table 3.28. Soldiers in combat MOS (11B, 13B, and 19K) have mean SJT Total Scores that are about a quarter of a standard deviation lower than the means for soldiers in the other five MOS. Table 3.28 also shows the mean SJT Total Scores for each of the nine different MOS. The MOS with the highest mean scores are 95B and 71L, and the MOS with the lowest mean scores are

19K and 88M. Analysis of variance showed that MOS differences accounted for more variance in SJT scores than did combat/noncombat differences (4% versus 1%). These results are very similar to those obtained for the CVII sample.

Table 3.28

Situational Judgment Test Scores by Combat/NonCombat and by MOS

	N	SJT Total Score		
		Mean	SD	d ^a
Combat MOS	689	1.54	.61	
Noncombat MOS	887-888	1.68	.52	-.24
<u>MOS^b</u>				
11B	345	1.58	.57	
13B	178	1.52	.70	
19K	166	1.48	.67	
31C	70	1.61	.58	
63B	191	1.53	.52	
71L	153	1.78	.44	
88M	88-89	1.49	.49	
91A/B	217	1.76	.52	
95B	168	1.79	.52	

^a This is the standardized mean difference between two subgroups' scores. A negative value indicates that soldiers in noncombat MOS scored higher than those in combat MOS.

^b Effect sizes were not computed for separate MOS.

Final Basic Scores

The results of the SJT data analyses indicate that the measure has appropriate distributional characteristics in the LVII sample. The five scoring procedures all resulted in scores with reasonable variance and internal consistency reliabilities, and item-total correlations were quite high. Results also indicate that the lengthening of the SJT for the LVII achieved the desired results, both higher reliabilities and a somewhat easier test.

Based on these psychometric characteristics, the most promising score appears to be M-L Effectiveness (i.e., SJT Total Score), which has an internal consistency reliability of .81. This score also appears to be a good summary of the information contained in the SJT. The SJT Total Score was used in the modeling of second-tour performance for the CVII sample as well, but during the CVII it was based on 35 items.

It was also possible to identify six relatively homogeneous subscales in this lengthened version of the SJT. These subscales have potential for more clearly delineating the leadership/supervision aspects of the second-tour soldier job and will be included in one of the major alternative models of second-tour performance to be evaluated in subsequent confirmatory analyses.

SUPERVISORY SIMULATION EXERCISES

The supervisory simulation measures were designed to assess areas of second-tour job performance that deal with specific components of supervisor/subordinate interaction. These areas included personal counseling, disciplinary counseling, and one-on-one training. A trained evaluator (role-player) acted out the role of a subordinate to be counseled or trained and the examinee assumed the role of a first-line supervisor who was to conduct the counseling or training. In each exercise, evaluators scored the examinees on a number of rating scales.

The subordinate and supervisor roles were essentially the same as those used in the CVII data collection. The role-players who assumed the role of the subordinate in each of these exercises were trained to play the roles in a standardized fashion. Before each role-play began, examinees were given a one-half page description of the problem and several minutes to consider their approach to handling the subordinate. The respective roles of the subordinates (role-players) and supervisors (examinees) are briefly summarized below.

Personal Counseling Simulation

- Supervisory problem: A private first class (PFC) is exhibiting declining job performance and personal appearance. Recently the PFC's wall locker was left unsecured. The supervisor has decided to counsel the PFC about these matters.
- Subordinate role: The soldier is having difficulty adjusting to life in Korea and is experiencing financial problems. The role-player is trained to initially react defensively to the counseling but to calm down if the supervisor handles the situation in a non-threatening manner. The subordinate will not discuss personal problems unless prodded.

Disciplinary Counseling Simulation

- Supervisory problem: There is convincing evidence that a PFC lied to get out of coming to work today. The PFC has arrived late to work on several occasions and has been counseled for lying in the past. The PFC has been instructed to report to the supervisor's office immediately.
- Subordinate role: The soldier's work is generally up to standards which leads the soldier to believe that he or she is justified in occasionally "slacking off." The subordinate has slept in to nurse a hangover and then lied to cover it up. The role-player is trained to initially react to the counseling in a very polite

manner but to deny that he or she is lying. If the supervisor conducts the counseling effectively, the subordinate eventually admits guilt and begs for leniency.

Training Simulation

- Supervisory problem: The commander will be observing the unit practice formation in 30 minutes. This private, although highly motivated, is experiencing problems with the hand salute and about face.
- Subordinate role: The role-player is trained to demonstrate feelings of embarrassment that contribute to the soldier's clumsiness. Role-player training also includes making very specific mistakes when performing the hand salute and about face.

For the CVII sample, examinees were rated on their performance on each exercise independently. Using a 3-point scale, ratings were made on from 11 to 20 behaviors tapped by each exercise. The three rating points were anchored with a description of performance on the particular behavior being rated. Examinees were also rated on a 5-point overall effectiveness scale following each of the three exercises. Additionally, examinees were rated on a 5-point overall affect scale following the personal counseling exercise and on a 5-point overall fairness scale following the disciplinary counseling exercise.

The rating system used to evaluate LVII examinees was modified in several ways from CVII. First, the CVII analyses identified the scales which appeared to be (a) difficult to rate reliably, (b) conceptually redundant with other rated behaviors, and/or (c) not correlated with other rated behaviors in meaningful ways. These behavior ratings were dropped to allow raters to concentrate more fully on the remaining behaviors. Some of the behavioral anchors were also changed to improve rating reliability, and the rating scale was expanded from 3 to 5 points. The overall effectiveness rating was retained, but the overall affect and fairness rating scales were eliminated. Thus, examinees were rated on each exercise on from 7 to 11 behavioral scales and on one overall effectiveness scale. Examples of two behavior rating scales from the Personal Counseling exercise are shown in Figure 3.2.

Another important difference between the CVII and LVII measures was the background of the evaluators. The smaller size of the LVII data collection allowed for the selection and training of role-players/evaluators who were formally educated as personnel researchers and who were employed full-time by organizations in the project consortium. In contrast, the scope of the LVI/CVII data collection required the hiring of a number of temporary employees to serve as role-players. Most of these individuals had no formal research training or related research experience. Informal observations of the simulation training and testing across the two data collections suggest that, in comparison to the CVII exercises, the LVII exercises were played in a more standardized fashion and examinees were rated more consistently both within and across evaluators.

States the purpose of the counseling session clearly and concisely.

- 5 = Outlines specific topics to be covered (e.g., the purpose is to discuss the wall locker that was left open last night, any problems the subordinate may be having and what might be done to resolve them, etc.).
- 3 = States at least one general topic to be discussed (e.g., says the purpose is to talk about the subordinate's recent poor performance).
- 1 = Fails to state a purpose for the session; instead, jumps directly into the problems.

Gives the subordinate positive feedback for his/her overall good past performance.

- 5 = Clearly/strongly acknowledges the subordinate's past effective performance; does so prior to the subordinate bringing up his/her own effective performance.
 - 3 = Acknowledges the subordinate's past effective performance but does not do so clearly/strongly or waits until the subordinate brings up his/her performance before recognizing it.
 - 1 = Fails to acknowledge the subordinate's past effective performance.
-

Figure 3.2. Sample scales from LVII Personal Counseling Simulation Exercise.

Data Analysis Procedure

Descriptive analyses were conducted, followed by a series of factor analyses. The purpose of the factor analyses was to identify the content of basic criterion scores for each of the simulation exercises. Maximum likelihood factor analyses with oblique rotations were performed within each exercise. The factor analyses were within exercise because analyses of the CVII data indicated that when the factor analyses included scales from multiple exercises, method factors associated with each exercise dominated the factor structure.

Raw scale ratings and scale ratings standardized by MOS, evaluator, and test site were factor analyzed because there was some concern that non-performance-related variables associated with MOS, evaluator, and/or test site might affect the factor structure of the raw scale ratings. No orthogonal rotations were used because, based on the CVII analyses, the factors were expected to be at least moderately correlated.

The overall effectiveness ratings were not considered for inclusion in the basic scores because they are conceptually distinct from the behavior ratings. Interrater reliability estimates could not be computed because there were insufficient "shadow score" data to conduct the required analyses.

Descriptive Statistics

Descriptive statistics which summarize the ratings of the specific scales in each of the three simulation exercises are contained in Table 3.29. Overall, the means and standard deviations are within expected ranges. The median and the range of the scale means and the median and the range of the scale standard deviations, for each exercise, indicate that (a) there is a reasonable amount of variation in the scale ratings, (b) none of the scale ratings show a floor effect, and (c) a reasonable number of the ratings do not show a ceiling effect.

Table 3.29

Descriptive Statistics for LVII Simulation Exercises

Scale Statistic	Personal Counseling	Disciplinary Counseling	Training
Number of Items	11	7	9
Number of Rates	1,482	1,480	1,457
Median Mean Rating ^a	3.70	3.32	3.84
Range of Rating Means	2.55-4.57	1.68-4.59	2.62-4.23
Median Standard Deviation	1.20	.86	1.16
Range of Standard Deviations	.80-1.62	.66-1.52	.98-1.59
Mean Correlation Among Ratings	.275	.128	.337
Mean Overall Efficiency Rating ^a	3.10	3.27	3.29
SD Overall Efficiency Rating	1.07	1.07	1.15

^a The ratings are on a 5-point scale; 1 indicates poor performance and 5 indicates excellent performance.

Factor Analysis Results

Summary statistics for factor analyses performed on the raw scale ratings in all three exercises are presented in Table 3.30. The summary statistics for the factor analyses of the standardized scale ratings are not shown. In terms of relative magnitude, they are similar to the results presented in Table 3.30.

Personal Counseling Exercise

Table 3.31 presents the pattern matrices resulting from the factor analyses of the standardized and raw score Personal Counseling exercise ratings that specified two factors. The two-factor structure was preferred over the one- or three- (or more) factor structures based on the superior simple structure and interpretability of the rotated two-factor pattern matrix. Factor 1 was labeled "Communication/Interpersonal Skills," and Factor 2 was labeled "Diagnosis/Prescription."

Table 3.30

Factor Analysis Summary Statistics for LVII Simulation Exercises^{ab}

Exercise	Factors	df	χ^2	p ^c	RMSEA ^d
Personal Counseling	1	44	1210.45	.0001	.134
	2	34	586.18	.0001	.105
Disciplinary Counseling	1	14	356.49	.0001	.129
	2	8	85.42	.0001	.081
	3	3	7.37	.0610	.039
Training	1	27	533.11	.0001	.113
	2	19	138.33	.0001	.066

^a Maximum likelihood factor analysis with an oblique rotation.

^b These are the results from analyses of the raw scale ratings.

^c The probability associated with the chi-square.

^d Root mean square error of approximation.

As indicated by the notations in Table 3.31, the factor analysis results for LVII did not exhibit the same pattern as that obtained in CVII. This is at least in part because the CVII exercise included nine scales that were not included in the LVII measure. The superscript 1 in Table 3.31 indicates that the same (or a similarly worded) scale was assigned to the CVII basic score titled "Personal Counseling - Content." The superscript 2 indicates that the same (or a similarly worded) scale was assigned to the CVII basic score titled "Personal Counseling - Process." Finally, the superscript OMIT indicates that a similarly worded scale was part of the equivalent CVII measure, but was not assigned to a basic score in CVII.

Disciplinary Counseling Exercise

Table 3.32 presents the pattern matrices resulting from the factor analyses of the standardized and raw scale Disciplinary Counseling exercise ratings that specified three factors. The three-factor structure was preferred over the one-, two-, or four- (or more) factor structures based on the superior simple structure and interpretability of the rotated three-factor pattern matrix. Factor 1 was labeled "Structure," Factor 2 was labeled "Interpersonal Skill," and Factor 3 was labeled "Communication."

Again, the scales listed in Table 3.32 are annotated to allow comparison with CVII results. Note that the equivalent CVII measure included four scales that were not included in the LVII measure and the factor analysis resulted in two rather than three factors.

Table 3.31

LVII Personal Counseling Exercise Scales and Factor Analysis Results^a

Scale	Factor 1		Factor 2		h ²	
	S	R	S	R	S	R
<u>Communication/Interpersonal Skill</u>						
1. States the purpose of the counseling session clearly and concisely. ¹	<u>.45</u>	<u>.24</u>	-.04	.08	.18	.08
2. Gives the subordinate positive feedback for his/her overall good past performance. ¹	<u>.74</u>	<u>.52</u>	-.10	.02	.48	.28
3. Explains what the soldier did wrong and why it was or can be a problem. ¹	<u>.38</u>	<u>.33</u>	-.06	-.02	.12	.10
7. Maintains eye contact during the interview. ²	<u>.30</u>	<u>.61</u>	.14	.05	.16	.28
8. Behaves in a manner that demonstrates support and concern for subordinate. ^{OMIT}	<u>.52</u>	<u>.73</u>	.30	.17	.54	.66
9. Conducts the counseling session in a professional manner. ²	<u>.47</u>	<u>.61</u>	.12	.05	.29	.40
10. Maintains open communication. ²	<u>.13</u>	<u>.49</u>	.45	.21	.27	.38
<u>Diagnosis/Prescription</u>						
4. Asks open-ended, fact-finding questions that uncover important and relevant information. ¹	.01	.24	<u>.78</u>	<u>.61</u>	.61	.56
5. Provides advice to the subordinate concerning actions that should be taken to solve problems. ¹	-.04	.04	<u>.87</u>	<u>.93</u>	.73	.89
6. Sets a time or date to follow-up with the subordinate. ¹	.01	.11	<u>.52</u>	<u>.50</u>	.27	.31
<u>Omitted Item</u>						
11. Does not interrupt the subordinate when he/she is talking. ²	.08	.43	.17	.02	.05	.19
Eigenvalue ^b	6.73	12.1	1.39	2.41		

Note. The underline indicates which composite the scale was assigned to for the construction of simulation exercise basic scores; h² = Communality; S = From analysis of standardized scale ratings; R = From analysis of raw scale ratings.

^a Maximum likelihood factor analysis with an oblique rotation.

^b Eigenvalues of the first two unrotated factors.

¹ A similar (or the same) scale was assigned to the Personal Counseling - Content composite score in CVII.

² A similar (or the same) scale was assigned to the Personal Counseling - Process composite score in CVII.

^{OMIT} A similar scale was not assigned to a composite score in the CVII analyses.

Table 3.32

LVII Disciplinary Counseling Exercise Scales and Factor Analysis Results^a

Scale	Factor 1		Factor 2		Factor 3		h ²	
	S	R	S	R	S	R	S	R
<u>Structure</u>								
1. Remains focused on the immediate problems (i.e., the subordinate's absences and/or lying). ¹	<u>.38</u>	<u>.30</u>	.12	.24	-.08	-.09	.17	.15
2. Determines an appropriate corrective action. ¹	<u>.57</u>	<u>.44</u>	.04	.13	-.02	-.07	.33	.21
3. States the exact provisions of the punishment. ¹	<u>.57</u>	<u>.70</u>	-.01	-.06	.07	.03	.34	.50
<u>Interpersonal Skill</u>								
6. Conducts the counseling session in a professional manner. ²	.07	.02	<u>.72</u>	<u>.71</u>	-.02	.04	.53	.51
7. Defuses rather than escalates potential arguments. ²	-.04	-.03	<u>.67</u>	<u>.76</u>	-.02	.00	.44	.57
<u>Communication</u>								
4. Explains the ramifications of the soldier's actions. ^{omit}	.01	-.01	-.03	.01	<u>.82</u>	<u>.66</u>	.66	.44
5. Allows the subordinate to present his/her view of the situation. ²	.14	.17	.08	.04	<u>.29</u>	<u>.33</u>	.14	.17
Eigenvalue ^b	2.62	2.53	1.52	1.45	1.02	0.78		

Note. The underline indicates which composite the scale was assigned to for the construction of simulation exercise basic scores; h² = Communality; S = From analysis of standardized scale ratings; R = From analysis of raw scale ratings.

^a Maximum likelihood factor analysis with an oblique rotation.

^b Eigenvalues of the first three unrotated factors.

¹ A similar (or the same) scale was assigned to the Disciplinary Counseling - Content score in CVII.

² A similar (or the same) scale was assigned to the Disciplinary Counseling - Interpersonal Skills score in CVII.

^{omit} A similar item was not assigned to a composite score in the CVII analyses.

Training Exercise

Table 3.33 presents the pattern matrices resulting from the factor analyses of the standardized and raw scale Training exercise ratings that specified two factors. The two-factor structure was preferred over the one- or three- (or more) factor structures based on the superior simple structure and interpretability of the rotated two-factor pattern matrix. Factor 1 was labeled "Structure" and Factor 2 was labeled "Motivation Maintenance." Each factor label listed above was designed to be descriptive of the scales that loaded highest on the particular factor.

The CVII training exercise included three scales that were not included in the LVII measure and only a single factor was identified by the factor analysis of those data. In rather striking contrast, a pronounced two-factor structure was evident in the LVII data.

Table 3.33

LVII Training Exercise Scales and Factor Analysis Results^a

Scale	Factor 1		Factor 2		h ²	
	S	R	S	R	S	R
<u>Structure</u>						
2. Organizes and presents the training steps in a logical sequence.	<u>.64</u>	<u>.67</u>	-.03	-.01	.39	.44
3. Demonstrates the task steps for the trainee.	<u>.58</u>	<u>.57</u>	.07	.12	.39	.44
4. Identifies and corrects the trainee's errors.	<u>.74</u>	<u>.71</u>	-.16	-.20	.41	.35
5. Makes the trainee practice each movement required to perform the task.	<u>.66</u>	<u>.60</u>	-.03	.04	.41	.40
6. Provides specific feedback to the trainee following good performance.	<u>.70</u>	<u>.77</u>	.04	.01	.53	.60
<u>Motivation Maintenance</u>						
7. Provides positive feedback to the trainee following good performance.	-.01	-.01	<u>.81</u>	<u>.87</u>	.65	.74
8. Encourages the trainee when mistakes are made.	-.07	-.05	<u>.80</u>	<u>.76</u>	.57	.53
<u>Omitted Items</u>						
1. Presents an overview of what will be learned.	.18	.15	.21	.24	.13	.13
9. Speaks in a clear, distinct, and understandable manner.	.28	.30	.26	.18	.25	.19
Eigenvalues ^b	6.12	7.21	1.32	1.42		

Note. The underline indicates which composite the scale was assigned to for the construction of simulation exercise basic scores. In the CVII analyses scales similar (or identical) to those above were assigned to a single Training Exercise composite score. h² = Communality; S = From analysis of standardized scale ratings; R = From analysis of raw scale ratings.

^a Maximum likelihood factor analysis with an oblique rotation.

^b Eigenvalues of the first two unrotated factors.

Basic Scores

Scales were assigned to composite scores based primarily on the patterns of their relative factor loadings in the factor structure for each exercise. This procedure resulted in empirically derived basic scores for each exercise that seemed to have considerable substantive meaning.

Two basic scores were created to represent performance on the Personal Counseling exercise (see Table 3.31). Scales 1 through 3 and 7 through 8 were assigned to the *Personal Counseling - Communication/Interpersonal Skills* composite because those scales loaded highest on Factor 1 in the analyses of the raw and the standardized scale ratings. Scale 10 loaded highest on Factor 2 in the analyses of the standardized scale ratings and on Factor 1 in the analyses of the raw scale ratings. Because Scale 10 appears to be conceptually more related to Factor 1 than to Factor 2, it was also assigned to the *Personal Counseling - Communication/Interpersonal Skills* composite. Scales 4 through 6 were assigned to the *Personal Counseling - Diagnosis/ Prescription* basic composite because they loaded highest on Factor 2 in the analyses of the raw and the standardized scale ratings. Scale 11 was not assigned to either composite score because the analyses of raw and standardized scale ratings disagreed about the factor on which the scale loaded highest and the scale's communality was relatively low (.19). Two basic scores were generated for the Personal Counseling exercise in CVII as well; however, they were structured significantly differently than those described here.

Three basic scores were created to represent performance on the Disciplinary Counseling exercise (see Table 3.32). Scales 1 through 3 were assigned to the *Disciplinary Counseling - Structure* composite because they loaded highest on Factor 1 in the analyses of the raw and the standardized scale ratings. Scales 6 and 7 were assigned to the *Disciplinary Counseling - Interpersonal Skill* composite because the scales loaded highest on Factor 2 in the analyses of the raw and the standardized scale ratings. Scales 4 and 5 were assigned to the *Disciplinary Counseling - Communication* composite because they loaded highest on Factor 3 in the analyses of the raw and the standardized scale ratings. Only two basic scores had been derived from the CVII Disciplinary Counseling exercise data.

Two basic scores were created to represent performance on the Training exercise (see Table 3.33). Scales 2 through 6 were assigned to the *Training - Structure* composite because those scales loaded highest on Factor 1 in the analyses of the raw and the standardized scale ratings. Scales 7 and 8 were assigned to the *Training - Motivation Maintenance* composite because they loaded highest on Factor 2 in the analyses of the raw and the standardized scale ratings. Scales 1 and 9 were not assigned to either composite score because the analyses of raw and standardized scale ratings show that these scales have relatively small loadings on both factors and relatively small communalities. Only one basic score was derived from the CVII Training exercises data.

Across all exercises, each basic composite score was generated by (a) standardizing the ratings on each scale within each evaluator, (b) scaling each standardized rating by its raw score mean and standard deviation, and (c) calculating the mean of the transformed (i.e., standardized and scaled) ratings that were assigned to that particular basic criterion composite. The ratings were standardized within evaluator because (a) each evaluator rated

examinees in only some MOS and (b) there was more variance in mean ratings across evaluators than there was in mean ratings across MOS. The standardized ratings were scaled with their original overall means and standard deviations so that each scale would retain its relative central tendency and variability. The correlations among the supervisory simulation basic scores are presented in Table 3.34.

Table 3.34

Correlations Among LVII Simulation Exercise Basic Scores

Basic Score	PCI	PDP	DST	DIS	DCO	TST	TMN
Personal Counseling - Communication/ Interpersonal Skill	1.00						
Personal Counseling - Diagnosis/Prescription	.51	1.00					
Disciplinary Counseling - Structure	.07	.09	1.00				
Disciplinary Counseling - Interpersonal Skill	.15	.19	.17	1.00			
Disciplinary Counseling - Communication	.15	.06	.12	.16	1.00		
Training - Structure	.25	.21	.09	.18	.09	1.00	
Training - Motivation Maintenance	.28	.18	.05	.20	.16	.49	1.00

SUMMARY OF BASIC CRITERION SCORES

The analyses described in this chapter resulted in an array of basic criterion scores which were available for the performance modeling activities described in Chapter 5. These scores are summarized in Figure 3.3.

Hands-On Performance Test

1. MOS-specific task performance score
2. General (common) task performance score

Job Knowledge Test

3. MOS-specific task knowledge score
4. General (common) task knowledge score

Army-Wide Rating Scales

5. Overall effectiveness rating
6. Leadership/supervision composite
7. Technical skill and effort composite
8. Personal discipline composite
9. Physical fitness/military bearing composite

MOS-Specific Rating Scales

10. Overall MOS composite

Combat Performance Prediction Scales

11. Overall Combat Prediction scale composite

Personnel File Form

12. Awards and certificates
13. Disciplinary actions (Articles 15 and Flag actions)
14. Physical readiness
15. Promotion rate

Situational Judgment Test

16. Total composite *or, alternatively,*
17. Discipline soldiers when necessary
18. Focus on the positive
19. Search for underlying causes
20. Work within chain of command
21. Show support/concern for subordinates
22. Take immediate/direct action

Supervisory Simulation Exercises

23. Personal counseling - Communication/Interpersonal skill
24. Personal counseling - Diagnosis/Prescription
25. Disciplinary counseling - Structure
26. Disciplinary counseling - Interpersonal skill
27. Disciplinary counseling - Communication
28. Training - Structure
29. Training - Motivation maintenance

Figure 3.3. Summary list of LVII basic criterion scores.

Chapter 4
THE LVII DATA FILE

Geofrey Wilson, Charles T. Keil, Jr., Scott H. Oppler, and Deirdre Knapp

This chapter describes the data file generated by the Longitudinal Validation Second-tour (LVII) data collection. The initial sample sizes and the LVII performance instruments will be specified in the opening sections. Subsequent sections will summarize the extent of missing data, the treatment of missing data for each of the individual instruments, and the final sample sizes.

INITIAL SAMPLE SIZES

The LVII data were collected from 1,595 soldiers in nine Military Occupational Specialties, designated as Batch A MOS in previous data collections. The sample, by MOS, is shown in Table 4.1. Table 4.2 and Table 4.3 show the distribution of the sample by gender and race, respectively.

Table 4.1

LVII Sample by MOS

MOS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
11B	347	21.8	347	21.8
13B	180	11.3	527	33.0
19K	168	10.5	695	43.6
31C	70	4.4	765	48.0
63B	194	12.2	959	60.1
71L	157	9.8	1,116	70.0
88M	89	5.6	1,205	75.5
91A/B	222	13.9	1,427	89.5
95B	168	10.5	1,595	100.0

Table 4.2

LVII Sample by Gender

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	206	12.9	206	12.9
Male	1,389	87.1	1,595	100.0

Table 4.3

LVII Sample by Race

Race	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Black	516	32.4	516	32.4
Native American	26	1.6	542	34.0
Hispanic	120	7.5	662	41.5
White	894	56.1	1,556	97.6
Other	39	2.4	1,595	100.0

LVII PERFORMANCE INSTRUMENTS

As noted in previous chapters, the Longitudinal Validation second-tour (LVII) sample was assessed on a number of measures over a one-day administration. Summary descriptions of the instruments can be found in Chapter 2 and more detailed descriptions of the instruments and the scores derived from them are provided in Chapter 3. The construction of these instruments has been described in listed detail in previous reports (Campbell, 1987; Campbell & Zook, 1990).

Performance Criterion Instruments

Approximately 75 percent of the assessment time was devoted to the measurement of second-tour performance. The individual instruments that were used are listed below.

- Job knowledge tests
- Hands-on performance tests
- Performance ratings scales
 - Army-Wide Ratings
 - MOS-Specific Ratings
 - Combat Performance Prediction Ratings
- Personnel File Form
- Situational Judgment Test
- Three Supervisory Simulation (role-play) Exercises
 - Personal Counseling
 - Disciplinary Counseling
 - Training

Supplemental Instruments

A number of supplemental instruments were also administered to the sample for purposes of sample stratification, to account for the effects of individual differences in experience, or to support other Army research interests:

- Background Information Form
- MOS-Specific Job History Questionnaire
- Supervisory Experience Questionnaire
- Army Job Satisfaction Questionnaire (AJSQ)
- Assessment of Background and Life Experiences (ABLE)
- Leader/Unit Attitudes
- Combat Performance Questionnaire (Operation Desert Shield/Storm)

Recall that the Combat Performance Questionnaire was administered only to those rater-ratee pairs who had been deployed to Operation Desert Shield/Storm. Although it was intended for use as a performance measure, the small sample sizes dictate that this instrument be excluded from the category of primary criterion measures.

The initial sample sizes for each principal criterion instrument administered in LVII are given in Table 4.4. The column headed N gives the number of soldiers, by MOS, from whom any data were collected on any instrument. The columns for each specific instrument show the number of soldiers from whom at least some data were collected for that instrument. The sample sizes for the supplemental instruments are shown in Table 4.5.

EXTENT OF MISSING DATA

Every effort was made to collect complete information from each soldier for all instruments. However, as described in Chapter 2, that was not always possible. For any instrument, information could be partially or completely missing. For example, for the hands-on measures, the necessary pieces of equipment might have been unavailable for use, making it impossible to score some or all of the steps of a particular task test. In the written tests, soldiers may have skipped a question they could not answer or they may not have been able to finish the test in the time provided. For supervisor ratings, the supervisors may have felt that they were not able to use a particular rating scale because of too few opportunities to observe that aspect of performance. For the Personnel File Form, soldiers may have left questions unanswered if they did not know or chose not to provide the requested information.

The number of soldiers that are missing all data on a particular instrument can be determined from Table 4.4. For example, only 341 of the 347 MOS 11B soldiers participated in hands-on testing while all 347 soldiers in the 11B sample participated in the job knowledge test administration.

Table 4.4

Number of LVII Soldiers With Complete or Partial Data by Criterion Instrument and MOS

MOS	N	Job Knowledge	Hands-On	Army-Wide Ratings	MOS Ratings	Combat Prediction	PFF	SJT	Simulation Exercises
11B	347	347	341	333	321	313	347	346	341
138	180	179	174	167	164	160	179	178	174
19K	168	168	160	156	149	152	161	166	156
31C*	70	70	--	65	66	64	68	70	--
638	194	192	187	182	191	176	193	193	188
71L	157	155	156	153	150	150	155	157	156
88M	89	89	88	86	87	85	88	89	88
91A/B	222	220	215	212	208	205	218	220	214
958	168	168	168	167	162	163	168	168	167
Total	1,595	1,589	1,489	1,521	1,498	1,468	1,577	1,587	1,485

Note. PFF = Personnel File Form; SJT = Situational Judgment Test.

* Hands-on and Supervisory Simulation Exercises data were not collected for MOS 31C.

Table 4.5

Number of LVII Soldiers With Data by Supplemental Instrument and MOS

MOS	Background Information	Job History	Supervisory Experience	Army Job Satisfaction	Leader/Unit Attitudes	Combat Performance Questionnaire ^a	ABLE
11B	347	344	343	345	338	44	308
13B	180	175	174	178	173	41	136
19K	168	164	164	164	162	51	110
31C	70	66	65	67	65	8	46
63B	194	189	188	190	185	30	135
71L	157	156	156	156	155	7	106
88M	89	87	86	89	86	19	54
91A/B	222	216	215	218	215	48	182
95B	168	167	167	167	167	8	35
Total	1,595	1,564	1,558	1,574	1,546	256	1,112

^a The Combat Performance Questionnaire was administered only to those rater-ratee pairs who had been deployed to Operation Desert Shield/Storm.

TREATMENT OF MISSING DATA

Various methods were used for each criterion instrument to deal with partially missing data. For some instruments, missing data were simply left as missing; these were the Personnel File Form and Simulation Exercises. For the other measures, various strategies were used to treat missing data. The following sections provide summaries of the amount of missing data for each performance measure and describe how it was handled.

Generally speaking, the minimum amount of data required for computing a basic criterion score was consistent with decision rules adopted in earlier data collections. These rules vary by measure, depending upon factors such as test length, item type, and extent of missing data. For example, 90 percent complete data was required to compute a job knowledge test score whereas 80-85 percent complete data (depending upon the task) was required to compute a hands-on score. Because of the relatively small sample sizes, no data imputation procedures were applied to the LVII criterion data.

Job Knowledge Tests

There were two main reasons for partially missing data for the Job Knowledge tests. Soldiers may have either skipped over a question within the test or been unable to complete the test within the time allotted. First, to be included in the job knowledge data set, soldiers could miss no more than 10 percent of the item responses. If individuals were missing more than 10 percent, their data were deleted from the Job Knowledge data set. Missing

item responses for individuals with 10 percent or less missing were treated as incorrect.

As Table 4.6 shows, only one soldier's Job Knowledge data were deleted because of excessive missing data.

Table 4.6

Number of LVII Soldiers With Incomplete Job Knowledge Data^a

MOS	None Missing	10% or Less Missing	More Than 10% Missing
11B	298	49	0
13B	151	28	0
19K	150	18	0
31C	62	8	0
63B	175	17	0
71L	137	18	0
88M	74	15	0
91A/B	191	29	1
95B	152	16	0
Total	1,390	198	1

^a Calculated for those who have at least some JK data.

For the Job Knowledge tests, as described in Chapter 3, two sets of scores were calculated. The first set was Task Factor scores: Communications, Vehicles, Basic Soldiering, Identify Targets, Technical, and Safety/Survival (CVBITS). The second set was Task Construct scores: MOS-Specific and General. Each item was assigned to a particular score category, and the composite scores were calculated by summing the number of correct responses made to the items within each category. For some MOS, only a subset of scores were computed; this occurred when no items were assigned to a particular category for a given MOS. The percentage of soldiers in the LVII sample for whom Job Knowledge scores were not computed is reported in Table 4.7. Note that the maximum amount of missing Percent Correct scores was 1.3 percent for MOS 71L. No attempt was made to calculate General Task Construct scores for MOS 11B because all common soldiering tasks can be considered technical tasks for this MOS.

Table 4.7

Percentage of LVII Soldiers With Missing Job Knowledge Scores by MOS

MOS	Percent Missing
11B	.00
13B	.56
19K	.00
31C	.00
63B	1.03
71L	1.27
88M	.00
91A/B	.90
95B	.00

Hands-On Tests

The hands-on measure consisted of observing and scoring the performance of each soldier on 14-17 independent job tasks. Tasks consisted of a varying number of discrete steps that were scored GO or NO GO. Within each task, data were missing generally because (a) the scorer failed to observe a step or failed to record the observation, (b) the scorer marked both GO and NO-GO, or (c) equipment was not available for testing all or part of a task.

For the most part, few data were missing at the step level. A Percent GO score was calculated for each task, using the step-level data. To receive a Percent GO score for a task, each soldier had to have scores for at least 85 percent of the steps (except as noted in the next paragraph). In other words, each soldier could have only 15 percent, or less, of the step data missing for each task for a Percent GO score to be calculated for that task. The Percent GO score was calculated on the basis of the scored steps.

Within certain MOS, some tasks were scored differently. For the MOS 11B task, Engage Targets with LAW, there were no step-level data. Soldiers received a Percent GO based on the number of targets hit. For the MOS 63B task, Perform Annual Preventive Maintenance Checks and Services (PMCS), soldiers could have up to 20 percent of the data missing for a Percent GO score to be calculated. For the MOS 71L task, Prevent Shock, soldiers could be missing up to 20 percent of the step-level data. The more liberal rules for these tasks were established because of the particularly severe missing data problems associated with them.

Task scores for a soldier were missing if the soldier was unable to be tested on the task at all. The task scores for these individuals were assigned values as follows. The Mean Percent GO score within an MOS for all soldiers who had completed that task was substituted for soldiers with a missing score for that task. Within each MOS, soldiers could have no more than two assigned task scores. If a soldier was missing more than two task scores, that soldier's data were deleted from the hands-on data base.

Each task was assigned to particular Task Factor (CVBITS) and Task Construct categories, just as items were assigned to score categories for the scoring of the Job Knowledge tests. For the Hands-On tests, composite scores were calculated as the mean of the Percent GO scores for the tasks assigned to each category, respectively. Note that the Percent GO scores were first standardized by Post. This was done to allow for differences in testing conditions (e.g., equipment, amount of space) across data collection sites. Also note that only a subset of CVBITS scores were completed for each MOS (except for 91A). This occurred when no tasks were assigned to a particular CVBITS category for a given MOS. The percentage of soldiers in the LVII sample for whom hands-on CVBITS scores have not been computed is shown in Table 4.8. Because of the nature of the MOS, no General Task Construct scores were calculated for MOS 11B.

Table 4.8

Percentage of LVII Soldiers With Missing Hands-On Scores by MOS

MOS	Percent Missing
11B	2.31
13B	4.44
19K	4.76
31C	-- ^a
63B	9.28
71L	1.27
88M	1.12
91A/B	5.41
95B	2.38

^a Hands-on data were not collected for MOS 31C.

Performance Rating Scales

Missing data on the rating scales were sometimes the result of the unavailability of suitable raters. Raters also left rating dimensions blank if they had had insufficient opportunity to observe performance on the dimension in question. This tended to be a particular problem for supervisory-related dimensions and MOS-specific dimensions which were not relevant for some of the rated soldiers (e.g., they did not supervise). Other data were lost due to administrative errors (i.e., Combat Performance Questionnaire administered in place of Combat Performance Prediction scales; page missing from MOS-specific rating booklet).

Army-Wide Performance Ratings

All raters who made ratings for individuals in the LVII sample were considered to be supervisors. No attempts were made to collect ratings from peers, and virtually all raters identified themselves as supervisors. Those who did not do so were in fact serving in a supervisory capacity but for some reason still considered themselves peers and so identified themselves. For each soldier, the ratings for each individual scale were averaged across all raters.

Four Army-Wide rating scale composites were calculated by taking the mean of the designated scales assigned to that composite. A soldier needed to have at least 60 percent of the scales used in calculating each rating composite. If not, the rating composite was set to missing. The four rating composites were labeled Leading and Supervising, Technical Skill and Effort, Personal Discipline, and Physical Fitness and Military Bearing. The single overall effectiveness rating was also used as a basic score. The percentage of soldiers in each MOS in the LVII sample with missing data for each of the Army-Wide rating composites and the overall effectiveness rating is shown in Table 4.9.

Table 4.9

Percentage of LVII Soldiers With Missing Data for Performance Rating Composite Scores by MOS

Composite Score	11B	13B	19K	31C	63B	71L	88M	91A/B	95B
Army-Wide Ratings									
Overall Effectiveness	6.34	5.56	9.52	1.43	2.06	4.46	1.12	5.41	1.79
Leading and Supervising	8.36	8.89	11.90	7.14	4.12	9.55	10.11	13.06	6.55
Technical Skill and Effort	6.05	5.56	9.52	1.43	2.06	4.46	1.12	5.41	1.79
Personal Discipline	6.05	5.56	9.52	1.43	2.06	4.46	1.12	5.41	1.79
Physical Fit/Mil Bearing	6.05	6.11	9.52	1.43	2.06	4.46	1.12	5.41	1.79
MOS-Specific Ratings									
Overall MOS Composite	9.22	11.67	12.50	10.00	2.06	6.37	4.49	17.57	10.71
Combat Performance Prediction									
Overall Combat Rating	6.34	13.89	20.83	1.43	4.12	4.46	10.11	7.21	4.76

MOS-Specific Ratings

As was the case for the Army-Wide ratings, the LVII MOS ratings for each soldier were averaged across all raters for each individual scale. The overall MOS rating composite was calculated as the mean of all the behavior-based scales for each MOS. Again, the soldier needed to have data for at least 60 percent of the individual scales if an overall mean was to be calculated; otherwise, the composite was coded as missing. The percentage of soldiers in each MOS in the LVII sample with missing data for the MOS overall composite is also shown in Table 4.9.

Combat Performance Prediction Scales

Missing data rules were used at two different points in the processing of the Combat Performance Prediction data. First, if an individual rater was missing more than 6 of the 14 individual rating scores, the ratings for that rater were dropped. After these ratings were dropped, the remaining ratings for each scale were averaged across all remaining raters. The overall rating composite was calculated by taking the sum over all items. If soldiers were missing any individual item (i.e., no rater rated it), their overall rating composite was set to missing. The percentage of soldiers in the LVII sample with missing data for the overall Combat Prediction composite is shown in Table 4.9 by MOS.

Personnel File Form

For the Personnel File Form, items were missing if the soldier (a) did not recall the information requested, (b) did not wish to provide the information requested, or (c) misunderstood the directions to complete the form. Five basic scores were calculated from the PFF. If one or more items used to calculate each basic score were missing, then the basic score was coded as missing. The percentage of soldiers in the LVII sample with missing data for each of the five Personnel File Form basic scores is shown in Table 4.10. Note that several MOS 19K soldiers did not complete the self-report measure at all, making missing data on these scores more of a problem for this MOS.

Table 4.10

Percentage of LVII Soldiers With Missing Data for Personnel File Form Basic Scores by MOS

Personnel File Form Basic Score	11B	13B	19K	31C	63B	71L	88M	91A/B	95B
Awards and Certificates	.00	.56	4.17	2.86	.52	1.27	1.12	1.80	.00
Disciplinary Actions	.00	.56	4.17	2.86	.52	1.27	1.12	1.80	.00
Promotion Rate	2.31	2.78	11.90	2.86	2.06	4.46	5.26	3.15	3.57
Physical Readiness	3.46	3.89	8.03	7.14	4.12	2.55	3.37	7.21	1.79
Weapon Qualification	.29	.56	7.74	1.47	.52	2.55	1.12	2.25	.60

Situational Judgment Test

Data could be missing for the Situational Judgment Test (SJT) for various reasons. For example, the soldier may have skipped a question or questions, or may not have followed directions properly. Moreover, the soldier could have been exceptionally slow and thus unable to complete the test in the allotted time.

To calculate the "Most-least" effectiveness total score, soldiers could be missing up to four "Most" and/or "Least" responses for the 49 questions. If the soldier was missing more than four responses, the "Most-Least" effectiveness basic score was coded as missing. Table 4.11 summarizes the percentage of missing data by MOS for the SJT "Most-Least" effectiveness basic score.

Supervisory Simulation Exercises

Data for the supervisory simulation exercises were missing if the soldier could not be tested (e.g., because of insufficient time) or if the scorer left items on the score sheet blank. As described in Chapter 3, a series of factor analyses were performed to identify the scorer rating scales that should make up the basic scores for each Simulation Exercise. The Disciplinary Counseling Simulation had three basic scores: Structure, Communication, and Interpersonal Skill. The Personal Counseling Simulation had two basic scores: Communication/Interpersonal and Diagnosis/Prescription. The Training Simulation had two basic scores: Structure and Motivation Maintenance.

Table 4.11

Percentage of Soldiers With Missing Data for the Situational Judgment Test Total Score by MOS

MOS	Percent Missing
11B	.52
13B	1.11
19K	1.19
31C	1.41
63B	1.55
71L	2.55
88M	1.12
91A/B	2.25
95B	.00

Basic scores were calculated as the mean across all rating scales included in that score. If any component scale was missing, the basic score was coded as missing. Based on these rules, the percentage of soldiers in the LVII sample with missing data for each of the Simulation Exercise basic scores is shown in Table 4.12.

Table 4.12

Percentage of LVII Soldiers With Missing Data for Simulation Exercises Basic Scores by MOS^a

Simulation Exercise Basic Score	11B	13B	19K	63B	71L	88M	91A/B	95B
Disciplinary Counseling								
Structure	1.44	3.89	7.74	4.64	.64	1.12	3.60	.60
Communication	1.44	3.89	7.74	4.64	.64	1.12	3.60	.60
Interpersonal	1.44	3.89	7.74	4.64	.64	1.12	3.60	.60
Personal Counseling								
Communication/Interpersonal	1.73	3.89	7.14	4.12	.64	1.12	3.60	.60
Diagnosis/Prescription	1.73	3.89	7.14	3.61	.64	1.12	3.60	.60
Training								
Structure	2.02	14.44	8.33	4.12	.64	1.12	4.05	1.19
Motivation Maintenance	2.02	14.44	8.33	4.12	.64	2.25	4.05	1.19

^a Simulation Exercises data were not collected for MOS 31C.

SUMMARY OF MISSING DATA TREATMENT

The percentage of assigned values for missing data for each performance instrument is shown in Table 4.13. That is, these are the individuals in the sample who had some missing data but not enough to be dropped from the data set for a particular instrument. Instead, their scores were computed using the rules described previously. Note that these percentages are generally very low; almost all are less than one percent except for the MOS Ratings Scales.

Table 4.13

Percentage of LVII Assigned Values by Type of Instrument and MOS

MOS	Job Knowledge	Hands-On	Army-Wide Rating Scales	MOS Rating Scales	Combat Ratings	Personnel File Form	Situational Judgment Test	Supervisory Simulation Exercises
11B	.00	.88	.19	2.88	.00	.00	.14	.00
13B	.00	1.55	.55	2.00	.00	.00	.17	.00
19K	.00	.00	.03	.68	.00	.00	.06	.00
31C	.00	-- ^a	.46	1.79	.00	.00	.15	-- ^a
63B	.00	1.92	.54	.66	.00	.00	.12	.00
71L	.00	.92	.85	1.75	.00	.00	.11	.00
88M	.00	.91	.44	3.95	.00	.00	.23	.00
91A/B	.00	.92	.78	8.08	.00	.00	.09	.00
95B	.00	.85	.61	5.67	.00	.00	.09	.00
Total Sample	.00	.98	.47	3.33	.00	.00	.12	.00

^a Hands-on and Supervisory Simulation Exercises data were not collected for MOS 31C.

Table 4.14 is a summary of the percentage of missing data at the basic score level. That is, this is the percentage of individuals for whom a particular score was missing altogether, or set to missing because of insufficient data. The ratings show the largest percentage of missing data, up to 20 percent, for the Combat Performance Prediction ratings. For the other instruments, the missing data percentages are generally low, approximately 1 to 2 percent. A summary of the amount of complete data for each performance instrument by MOS after deleting records because of missing data rules, and after applying scoring rules, is shown in Table 4.15.

Table 4.14

LVII Combined Criteria Data: Percentage of Soldiers With Missing Data for Composite or Basic Scores by MOS

Criteria	11B	13B	19K	31C	63B	71L	88M	91A/B	95B
Job Knowledge Scores (All)	.00	.56	.00	.00	1.03	1.27	.00	.90	.00
Hands-On Scores (All)	2.31	4.44	4.76	--	9.28	--	1.12	5.41	2.38
Army-Wide Ratings									
Overall Effectiveness	6.34	5.56	9.52	1.43	2.06	4.46	1.12	5.41	1.79
Leading and Supervising	8.36	8.89	11.90	7.14	4.12	9.55	1.11	13.06	6.55
Technical Skill and Effort	6.05	5.56	9.52	1.43	2.06	4.46	1.12	5.41	1.79
Personal Discipline	6.05	5.56	9.52	1.43	2.06	4.46	1.12	5.41	1.79
Physical Fitness and Military Bearing	6.05	6.11	9.52	1.43	2.06	4.46	1.12	5.41	1.79
MOS-Specific Ratings									
Overall MOS Composite	9.22	11.67	12.50	1.00	2.06	6.37	4.49	17.57	1.71
Combat Performance Prediction									
Overall Composite	6.34	13.89	2.83	1.43	4.12	4.46	1.11	7.21	4.76
Personnel File Form									
Awards and Certificates	.00	.56	4.17	2.86	.52	1.27	1.12	1.80	.00
Flag Actions and Articles 15	.00	.56	4.17	2.86	.52	1.27	1.12	1.80	.00
Promotion Rate	2.31	2.78	11.90	2.86	2.06	4.46	5.26	3.15	3.47
Physical Readiness Test Score	3.46	3.89	8.93	7.14	4.12	2.55	3.37	7.21	1.79
Weapon Qualification	.29	.56	7.74	1.47	.52	2.55	1.12	2.25	.60
Situational Judgment Test Total Score	.58	1.11	1.19	1.41	1.55	2.55	1.12	2.25	.00
SE - Disciplinary Counseling									
Structure	1.44	3.89	7.74	--	4.64	.64	1.12	3.60	.60
Communication	1.44	3.89	7.74	--	4.64	.64	1.12	3.60	.60
Interpersonal Skill	1.44	3.89	7.74	--	4.64	.64	1.12	3.60	.60
SE - Personal Counseling									
Communication/Interpersonal	1.73	3.89	7.14	--	4.12	.64	1.12	3.60	.60
Diagnosis/Prescription	1.73	3.89	7.14	--	3.61	.64	1.12	3.60	.60
SE - Training									
Structure	2.02	14.44	8.33	--	4.12	.64	1.12	4.05	1.19
Motivation Maintenance	2.02	14.44	8.33	--	4.12	.64	2.25	4.05	1.19

Note. -- Indicates that the particular score was not calculated for that MOS. SE = Supervisory Simulation Exercises.

Table 4.15

Numbers of Soldiers With Complete Data (After Applying Scoring Rules) Across All Instruments and by Type of Instrument and MOS

MOS	#	All Criteria	All Excluding Combat ^a	Job Knowledge	Hands-On	Army-Wide Rating	MOS Rating	Combat Prediction	Personnel File	Situational Judgment Test	Supervisory Simulation Exercises
11B	347	280	281	347	339	318	315	325	328	345	340
13B	180	105	117	179	172	164	159	155	169	178	152
19K	168	89	105	168	160	148	147	133	141	166	153
31C	70	--	--	70	^b	64	63	63	65	70	-- ^a
63B	194	155	157	192	176	186	190	186	183	191	184
71L	157	129	129	155	155	142	147	150	148	153	156
88H	89	64	69	89	88	80	85	80	82	88	87
91A/B	222	153	156	220	210	193	183	206	203	217	213
95B	168	126	130	168	164	157	150	160	159	168	166
Total	1,595	1,101	1,144	1,588	1,464	1,452	1,439	1,458	1,478	1,576	1,451

^a All criterion measures excluding Combat Prediction Scales. The information is provided because some modeling analyses were conducted on this subset of measures.

^b Hands-on and Simulation Exercises data were not collected for MOS 31C.

Chapter 5
DEVELOPMENT OF THE SECOND-TOUR PERFORMANCE MODEL
FROM THE LONGITUDINAL VALIDATION SAMPLE

Mary Ann Hanson, John P. Campbell, Amy Schwartz McKee, and Rodney A. McCloy

INTRODUCTION

This chapter describes analyses of the Longitudinal Validation sample second-tour (LVII) criterion scores to determine how the total covariation in these scores can best be represented by a smaller number of basic performance factors. That is, a major objective was to evaluate alternative factor models of the latent structure of second-tour NCO performance. A second objective was to determine the extent to which a hierarchical set of even more parsimonious models (i.e., that postulate fewer and fewer underlying factors) can account for the observed covariation in the LVII basic criterion scores.

Analyses were guided by the same general framework that was used in modeling the covariation among performance measures for first-tour performance (Campbell, McHenry, & Wise, 1990). Total performance is assumed to be composed of a small number of relatively distinct components such that aggregating them into one score covers up too much information about relative proficiency on the separate factors. The meaning of each separate component is independent (conceptually at least) of measurement method. The major components that are hypothesized to exist comprise the so-called latent structure of performance.

The Problem

The analysis task was to determine which model (i.e., a particular specification of the number of components and their substantive content) of the latent structure best fits the observed data. A good fit implies that the composite scores used to measure each major component are both a parsimonious and a valid representation of the basic nature of performance.

A preliminary model of second-tour performance had been developed based on data from the Project A Concurrent Validation second-tour (CVII) sample. This model, referred to as the Training and Counseling model, is described in detail in Campbell and Oppler (1990). Briefly, the development of the model involved the following steps: (a) identifying a set of basic performance criterion scores; (b) examining the correlations among the scores, using exploratory factor analyses; (c) suggesting several alternative models for "confirmation"; and (d) comparing the "fit" of the model across jobs, using the CVII data.

The LVII data provide an opportunity to confirm the fit of the CVII Training and Counseling model in an independent sample. An additional objective was to evaluate the fit of alternative a priori models. In general, the LVII data should provide a better understanding of second-tour performance because the LVII sample is somewhat larger than the CVII sample and because several of the individual performance measures had been revised and refined on the basis of the results of the CVII analyses.

The Measures

The data were collected from the LVII sample using the measures of second-tour performance that were developed as part of Project A (Campbell, 1989) and later modified based on the results of the CVII data analyses (Campbell & Zook, 1990). Chapter 2 described how the CVII measures were modified for the LVII data collection and Chapter 3 described how each of the major sets of performance measures was reduced from a large number of item, task, or individual scale scores to a smaller set of basic performance scores.

The LVII criterion scores are similar to the scores that served as input for the CVII modeling analyses. One notable difference is in the scores from the two measures of supervisory performance: the Situational Judgment Test and the Supervisory Simulation Exercises. A larger number of scores were derived from these two measures in LVII than in CVII, and there are also several substantive differences.

The results of this first level of aggregation have been referred to as the "basic" array of LVII criterion scores. Following is a brief review of the LVII criterion measures and the differences between the CVII and LVII scores.

Hands-On Performance Tests. As in the CVII data, analyses of the Percent GO scores for the various hands-on task tests for all MOS except 11B suggested two overall clusters of tasks: MOS (i.e., job) specific tasks and general, or common, tasks. For the 11B MOS, all the tasks formed a single cluster. Because a subset of these common tasks form the technical component of the infantry MOS, this score was treated as the job-specific hands-on score for 11Bs. Hands-on performance data were not collected for soldiers in MOS 31C during the LVII data collection because of ongoing equipment changes.

Job Knowledge Tests. The job knowledge tests also were organized around specific samples of tasks. Parallel to the hands-on performance scores, a two-factor model with separate general soldiering and MOS-specific scores was indicated for eight of the nine MOS. All of the MOS 11B job knowledge tasks formed a single cluster, and this was treated as the MOS-specific job knowledge score for 11Bs.

Army-Wide Performance Ratings. Both the LVII and the CVII analyses utilize supervisory ratings. Some peer ratings had been collected for the CVII sample, but these data were considerably less complete than for supervisors. The same four factors identified in analyses of the CVII ratings emerged in the LVII factor analyses. Consequently, the basic criterion composite scores derived from these ratings are identical to those used in CVII: Leading/Supervising, Technical Skill/Effort, Personal Discipline, and Physical Fitness/Military Bearing. The Army-Wide overall effectiveness rating was included in the LVII analyses but had not been included in the CVII modeling.

MOS-Specific Performance Ratings. As in CVII, no consistent factor structure was found within the MOS-specific ratings and a single composite score (the mean overall behavior-based scales) was used to provide a summary of the information contained in these ratings.

Combat Performance Prediction Ratings. During the CVII data collection, only males were rated on the Combat Performance Prediction scales. So as not to exclude females, scores from these scales were not included in the CVII modeling analyses. During the LVII data collection, females were also rated on the Combat scales, and these scales were included in the present analyses. A single score was obtained by summing across all 14 items. The results of exploratory factor analyses did not support the use of subscales.

Personnel File Form Measures. Analyses of the items on the administrative records questionnaire and the supplemental data from the Enlisted Master File suggested five scores: awards, disciplinary actions, promotion rate, physical readiness, and weapons qualification. These same variables were included in the CVII analysis as well. The weapons qualification score did not fit well in any of the models tested in CVII, however, and was not included in the final CVII model. Consequently, this score was excluded from all of the LVII analyses. One additional variable that was included in the CVII analysis--number of military training courses completed--was not included in the present analyses because of problems with the interpretation and distribution of responses.

Situational Judgment Test (SJT). The SJT was lengthened for the LVII data collection, and factor analyses of this longer version of the SJT yielded six relatively homogeneous subscores. These six factor-based subscores were initially included in the present analyses in place of the SJT Total Score that was used in the CVII.

Supervisory Simulation Exercises. The revised rating scales that were used to score the three Supervisory Simulation Exercises during the LVII data collection yielded a somewhat different factor solution than was obtained in the CVII analyses; this in turn led to a somewhat different set of basic criterion scores for the LVII Supervisory Simulation Exercises. Seven Supervisory Simulation scores were identified in the LVII analyses whereas the CVII included only five.

The criterion scores used to model LVII performance are listed in Table 5.1.

The Sample

The sample used in the LVII modeling analyses included soldiers from eight of the nine Batch A MOS for which a full set of criterion measures had been developed (C.H. Campbell et al., 1990). Because complete data on the entire array of basic criterion scores were required and because soldiers from the MOS 31C did not have hands-on performance scores, these soldiers were excluded from all of the present analyses. In addition, 43 of the soldiers in the LVII sample who had otherwise complete basic score data had not been rated on the Combat Performance Prediction scales during the LVII data collection. To include these soldiers, the Combat scales were omitted from the initial analyses. No score imputations or other treatments of missing data were carried out at the factor score level. If any one of the remaining basic scores was missing, the individual was eliminated from the sample.

As a result of these considerations, a total sample of 1,144 soldiers with complete data was available for the initial modeling analyses. The MOS breakdown is shown in Table 5.2. Fourteen percent of these soldiers were

Table 5.1

List of Basic Criterion Scores Used in LVII Performance Modeling Exercise

Hands-On Performance Test

1. MOS-specific task performance score
2. General (common) task performance score

Job Knowledge Test

3. MOS-specific task knowledge score
4. General (common) task knowledge score

Army-Wide Rating Scales

5. Overall effectiveness rating
6. Leadership/supervision composite
7. Technical skill and effort composite
8. Personal discipline composite
9. Physical fitness/military bearing composite

MOS-Specific Rating Scales

10. Overall MOS composite

Combat Performance Prediction Scales

11. Overall Combat Prediction scale composite

Personnel File Form

12. Awards and certificates
13. Disciplinary actions (Articles 15 and Flag actions)
14. Physical readiness
15. Promotion rate

Situational Judgment Test

16. Total composite or, *alternatively,*
17. Discipline soldiers when necessary
18. Focus on the positive
19. Search for underlying causes
20. Work within chain of command
21. Show support/concern for subordinates
22. Take immediate/direct action

Supervisory Simulation Exercises

23. Personal counseling - Communication/interpersonal skill
24. Personal counseling - Diagnosis/Prescription
25. Disciplinary counseling - Structure
26. Disciplinary counseling - Interpersonal skill
27. Disciplinary counseling - Communication
28. Training - Structure
29. Training - Motivation maintenance

female, and the racial breakdown was as follows: 56 percent white, 33 percent black, 8 percent Hispanic, and 2 percent Native American (the remainder reported "other").

Table 5.2

Number of LVII Soldiers With Complete Array of Basic Criterion Scores
(Excluding Combat Performance Prediction Scales) by MOS

MOS		Number With Complete Data
11B	Infantryman	281 ^a
13B	Cannon Crewmember	117
19K	M1 Armor Crewman	105
31C	Single Channel Radio Operator	0
63B	Light-Wheel Vehicle Mechanic	157
71L	Administrative Specialist	129
88M	Motor Transport Operator	69
91A/B	Medical Specialist	156
95B	Military Police	130
Total Sample		1,144

^a These soldiers do not have general soldiering scores for the hands-on or job knowledge tests.

THE MODELING ANALYSIS PROCEDURE

The basic steps in the modeling analysis were as follows. First, several alternative models of second-tour soldier performance were hypothesized. The fit of these alternative models was then assessed using the LVII data and compared with the fit of the CVII Training and Counseling model. Second, because the Combat Performance Prediction Scales were not included in this initial modeling, key analyses were rerun with these scales included to confirm that the Combat scales fit the models as expected and to determine whether including them would affect the degree of fit. Once a best fitting model was identified, subsequent analyses were conducted to determine whether the model fit equally well across MOS and across demographic subgroups. Finally, based on the results of these analyses, a set of criterion construct scores to be used in the LVII validation analyses were specified.

The CVII Model as One Alternative

The Training and Counseling model of second-tour performance developed on the basis of the CVII data is shown in Table 5.3. This model is similar to the model of first-tour soldier performance that was identified by Campbell, McHenry, and Wise (1990) using the CVI sample and was later confirmed in the LVI sample by Oppler, Childs, and Peterson (1994). The first-tour model contained five substantive factors -- (1) Core Technical Proficiency, (2) General Soldiering Proficiency, (3) Effort and Leadership, (4) Personal Discipline, (5) Physical Fitness/Military Bearing -- and two method factors.

Table 5.3

CVII Training and Counseling Model^a

Latent Variable	Scores Loading on Latent Variables
Core Technical Proficiency (CT)	MOS-Specific Hands-On MOS-Specific Job Knowledge
General Soldiering Proficiency (G)	General Hands-On General Job Knowledge
Effort and Leadership (EL)	Awards and Certificates Promotion Rate Army-Wide Ratings: Leading/Supervising Composite Army-Wide Ratings: Technical Skill/Effort Composite Overall Effectiveness Rating MOS Ratings: Overall Composite Combat Prediction: Overall Composite SJT: Total Score
Personal Discipline (PD)	Disciplinary Actions (reversed) Army-Wide Ratings: Personal Discipline Composite
Physical Fitness/Military Bearing (PF)	Physical Readiness Score Army-Wide Ratings: Physical Fitness/ Bearing Composite
Training and Counseling Subordinates (TC)	SE - Counseling Diagnosis/Prescription SE - Counseling Communication/Interpersonal Skills SE - Disciplinary Structure SE - Disciplinary Communication SE - Disciplinary Interpersonal Skill SE - Training Structure SE - Training Motivation Maintenance
Written Methods	MOS-Specific Knowledge General Job Knowledge SJT: Total Score
Ratings Methods	Four Army-Wide Ratings Composites Overall Effectiveness Rating MOS Ratings: Overall Composite Combat Prediction: Overall Composite

Note. SJT = Situational Judgment Test; SE = Simulation Exercise.

^a Scores shown on this table are those used in the LVII modeling analyses.

The primary difference between the model of first-tour soldier performance and the Training and Counseling model of second-tour performance is that the second-tour model was expanded to incorporate the supervisory aspects of the second-tour NCO position. Those elements were represented by a sixth factor, called Training and Counseling Subordinates, and included all scores from the Supervisory Simulation Exercises. Campbell and Oppler (1990) note that the Supervisory Simulation Exercise scores defined a new factor in large part because they show a good deal of internal consistency, but have very low correlations with any of the other performance measures.

Two other supervisory measures, the Situational Judgment Test and the Leading/Supervising rating composite, were constrained to load on the factor called Effort and Leadership. Finally, whereas promotion rate was part of the Personal Discipline factor in the model of first-tour performance, the revised promotion rate variable fit more clearly with the Effort and Leadership factor in the second-tour model. Apparently for soldiers in their second tour a relatively high promotion rate is due to positive achievement rather than simply the avoidance of disciplinary problems.

The CVII Training and Counseling model has one undesirable characteristic: the Training and Counseling factor itself confounds method variance with substantive variance. One of the objectives in generating alternative hypotheses of the underlying structure of second-tour soldier performance was to avoid this problem. The larger LVII sample and the improved methods used to collect these data provide a better opportunity for exploring the nature of second-tour performance than did the CVII sample.

Expert-Generated Alternatives

Definitions of the LVII basic criterion scores used in the modeling exercise were circulated to the project staff, and a variety of hypotheses concerning the nature of the underlying structure of second-tour soldier performance were obtained. These hypotheses were consolidated into one principal central alternative model, several variations on this model, and a series of more parsimonious models that involved collapsing two or more of the substantive factors.

The central alternative, the Consideration/Initiating Structure model, is presented in Table 5.4. It differs from the CVII Training and Counseling model primarily in that it includes two leadership factors. The composition of these two factors -- given their traditional labels of Consideration and Initiating Structure -- is based on the general findings of the Ohio State Leadership Studies and virtually all subsequent leadership research (Fleishman, 1973; Fleishman, Zaccaro, & Mumford, 1991). Based on staff judgment, each of the SJT and Supervisory Simulation scores was assigned to one of these two factors. Because the majority of the scales contained in the Army-wide Leading/Supervising composite appear to involve initiating structure, this score was assigned to the Initiating Structure factor.

However, some of the rating scales included in the Army-wide Leading/Supervising rating basic score are clearly more related to consideration than to structure. Thus, one variation of this model that was tested involved rationally assigning the scales from this basic rating score to the appropriate Leadership factor. Another variation on this model was to assign both of the scores from the Personal Counseling exercise to the Consideration factor, because this entire exercise could be seen as more related to consideration than to initiating structure.

The analysis plan was to first compare the fit of the Consideration/Initiating Structure model and the variations of this model with each other and with the fit of the Training and Counseling model, and to identify the

Table 5.4

Consideration/Initiating Structure Model

Latent Variable	Scores Loading on Latent Variables
Core Technical Proficiency (CT)	MOS-Specific Hands-On MOS-Specific Job Knowledge
General Soldiering Proficiency (GP)	General Hands-On General Job Knowledge
Achievement and Effort (AE)	Awards and Certificates Promotion Rate Army-Wide Ratings: Technical Skill/Effort Composite Overall Effectiveness Rating MOS Ratings: Overall Composite Combat Prediction: Overall Composite
Personal Discipline (PD)	Disciplinary Actions (reversed) Army-Wide Ratings: Personal Discipline Composite
Physical Fitness/Military Bearing (PF)	Physical Readiness Score Army-Wide Ratings: Physical Fitness/ Bearing Composite
Leadership: Initiating Structure (IS)	Army-Wide Ratings: Leading/Supervising Composite SE - Disciplinary Structure SE - Counseling Diagnosis/Prescription SE - Training Structure SJT - Disciplining SJT - Immediate/Direct Action SJT - Chain of Command
Leadership: Consideration (LC)	SE - Disciplinary Communication SE - Disciplinary Interpersonal Skill SE - Counseling Communication/Interpersonal Skills SE - Training Motivation Maintenance SJT - Support SJT - Search for Reasons SJT - Focus on the Positive
Written Methods	Technical Knowledge Basic Job Knowledge All Six SJT Scores
Ratings Methods	All Four Army-Wide Ratings Composites Overall Effectiveness Rating MOS Ratings: Overall Composite Combat Prediction: Overall Composite
Disciplinary Simulation Exercise Methods	All Three SE - Disciplinary Counseling Scores
Counseling Simulation Exercise Methods	Both SE - Personal Counseling Scores
Training Simulation Exercise Methods	Both SE - Training Scores

alternatives that best fit the LVII covariance structure. The next set of analyses involved comparing a series of nested models to determine the extent to which the observed correlations could be accounted for by fewer underlying factors.

Confirmatory Analysis Steps

Because the within-MOS sample sizes in the LVII sample were relatively small (ranging from 69 to 281), initial tests of the models were conducted using the entire LVII sample. For MOS 11B, as discussed previously, all hands-on task scores are summed to form a technical or MOS-specific basic score and all job knowledge test items are summed to form a technical or MOS-specific knowledge basic score; there are no general soldiering hands-on or job knowledge basic scores.

This MOS represents approximately one quarter of the LVII sample, so it was not appropriate to exclude these soldiers from the modeling analyses. However, the modeling analyses required complete data on the entire array of basic criterion scores. It could be argued that the MOS-specific components of the infantry job overlap almost completely with its general soldiering components; consequently, there is some conceptual rationale for using their MOS-specific hands-on and job knowledge test scores in place of general soldiering scores (or vice versa). In fact, this was done in the present analyses by adding error (a random normal deviate with a variance equal to the estimated standard error of measurement for the MOS-specific score) to the job-specific scores for these soldiers and using these new scores as their general soldiering scores.

To check whether this "imputing" of data for the infantryman MOS biased the modeling results, all of the analyses were run twice, once for the total sample and once including only those soldiers from the seven MOS for which actual general soldiering scores were available.

Procedure

Criterion scores were first standardized within each MOS, then the intercorrelations among these standardized basic scores were computed across all MOS. The total sample matrix was used as input for the analyses. Table 5.5 shows the resulting correlation matrix that was used for the total sample, and Table 5.6 shows this correlation matrix with MOS 11B excluded. Due to space limitations, the matrices presented on these tables do not include the SJT subscores, only the SJT Total Score. The correlations of the SJT subscores with other basic criterion scores that are targeted at the supervisory aspects of the job are presented in Table 5.7 (for the total sample).

These correlation matrices were submitted to confirmatory factor analyses using the LISREL 7 computer program (Jöreskog & Sörbom, 1989b). LISREL 7 is designed to analyze covariance structural models, and is appropriate for analyzing correlation matrices only if the models to be tested are scale invariant. To determine whether the use of correlation matrices was appropriate in the present analyses, several analyses were conducted a second time using the variance-covariance matrices, as suggested by Cudeck (1989). Results indicated that correlation matrices are, in fact, appropriate for the models tested, and only the correlational results are presented here.

Table 5.5

Correlations Among the LVII Basic Criterion Scores Based on All Soldiers With Complete Data^a

Criterion Score	PFF Awards	PFF Disc	PFF Prom	PFF Phys	HO Gen	HO MDS	JK Gen	JK MDS	AMB Lead	AMB Tech	AMB Disc	AMB Phys	Overall Rat	MOS Comp	Combat Comp ^b	SE DStrc	SE DCom	SE DIS	SE CC/IS	SE CD/P	SE TSt	SE Ths	SJT Tot	
PFF-Awards	1.00																							
PFF-Discipline	-.07	1.00																						
PFF-Prom Rate	.25	-.19	1.00																					
PFF-Phys Read	.13	-.12	.18	1.00																				
HO-General ¹	.11	-.07	.19	.10	1.00																			
HO-MOS Spec	.10	-.02	.20	.08	.44	1.00																		
JK-General ¹	.15	-.03	.25	.03	.47	.37	1.00																	
JK-MOS Spec	.08	-.03	.25	.04	.36	.44	.61	1.00																
Amb-Lead/Sup	.21	-.21	.31	.12	.19	.22	.20	.27	1.00															
Amb-Tech Skill	.18	-.17	.21	.11	.17	.18	.19	.24	.78	1.00														
Amb-Discipline	.13	-.25	.23	.07	.13	.15	.17	.20	.66	.64	1.00													
Amb-Phys Fit	.16	-.24	.24	.33	.08	.10	.02	.06	.51	.48	.49	1.00												
Overall Rat	.18	-.20	.26	.13	.16	.20	.18	.25	.80	.79	.65	.53	1.00											
MOS Composite	.16	-.15	.23	.06	.18	.25	.20	.28	.74	.72	.57	.45	.72	1.00										
Combat Composite ^b	.20	-.18	.24	.13	.19	.15	.22	.25	.70	.70	.50	.48	.69	.69	1.00									
SE-Disc Struc	-.02	.01	.03	-.03	.10	.06	.03	.08	.06	.02	-.03	.01	.03	.06	.04	1.00								
SE-Disc Com	.06	.01	.08	.02	.09	.00	.13	.05	.04	.02	-.01	-.03	.03	-.01	.04	.14	1.00							
SE-Disc Int Skl	-.03	-.04	.05	.04	.07	.04	.06	.01	.04	.02	.06	.07	.04	.03	.01	.08	.14	1.00						
SE-Coun Com	.06	-.05	.14	.09	.16	.10	.17	.13	.14	.09	.09	.06	.12	.09	.09	.15	.12	1.00						
SE-Coun Diag/Fr	.08	-.08	.16	.05	.14	.07	.15	.12	.13	.06	.07	.07	.11	.07	.08	.10	.20	.52	1.00					
SE-Train Struc	.08	-.06	.13	.05	.24	.21	.24	.20	.13	.10	.09	.10	.12	.09	.13	.07	.17	.07	.26	1.00				
SE-Train Motiv	.02	-.07	.05	-.01	.17	.14	.15	.13	.10	.06	.08	.01	.07	.05	.08	.04	.20	.17	.28	.21	1.00			
SJT-Total	.07	-.06	.22	-.02	.20	.23	.42	.38	.18	.16	.17	.06	.15	.15	.15	.08	.06	.11	.18	.15	.17	.49	1.00	

Note. See Table 5.1 for the full names of the criterion scores.

^a Except the Combat Performance Prediction Scales (N = 1,144).

^b These correlations are based on the subsample of soldiers rated on the Combat Performance Prediction Scales (N = 1,101).

Table 5.6

Correlations Among the LVII Basic Criterion Scores With MOS IIB Excluded^a

Criterion Score	PFF Awards	PFF Disc	PFF Prom	PFF Phys	EW Gen	MO MOS	JK Gen	JK MOS	AMB Lead	AMB Tech	AMB Disc	AMB Phys	AMB Overall	MOS Comp	Combat Comp ^b	SE Disc	SE BCom	SE DIS	SE CC/IS	SE CD/P	SE TSt	SE ThO	SJT	
PFF-Awards	1.00																							
PFF-Discipline	-.07	1.00																						
PFF-Prom Rate	.24	.18	1.00																					
PFF-Phys Road	.11	-.13	.17	1.00																				
MO-General	.10	-.07	.18	.09	1.00																			
MO-MOS Spec	.03	-.01	.16	.05	.30	1.00																		
JK-General	.16	.01	.24	.02	.47	.33	1.00																	
JK-MOS Spec	.06	.00	.23	.03	.32	.40	.51	1.00																
AMB-Lead/Sup	.22	-.19	.30	.10	.20	.21	.18	.25	1.00															
AMB-Tech Skill	.19	-.15	.20	.10	.17	.17	.17	.22	.78	1.00														
AMB-Discipline	.17	-.25	.22	.06	.14	.16	.17	.21	.66	.66	1.00													
AMB-Phys Fit	.14	-.26	.21	.34	.07	.08	-.01	.04	.49	.47	.45	1.00												
Overall Rat	.18	-.20	.23	.12	.17	.20	.16	.24	.60	.78	.67	.52	1.00											
MOS Composite	.16	-.12	.21	.06	.18	.24	.17	.26	.73	.71	.59	.42	.72	1.00										
Combat Composite ^b	.20	-.18	.23	.13	.22	.16	.20	.23	.71	.70	.53	.47	.68	.68	1.00									
SE-Disc Struc	-.01	.00	.02	-.34	.09	.04	.02	.06	.08	.94	-.01	.63	.06	.07	.06	1.00								
SE-Disc Comm	.03	.02	.07	.02	.11	-.01	.15	.04	.63	.01	-.04	-.63	.01	-.01	.04	.17	1.00							
SE-Disc Int Skl	-.02	-.06	.02	.08	.09	.05	.09	.04	.06	.06	.07	.10	.09	.07	.04	.07	.15	1.00						
SE-Coun Comm	.06	-.06	.12	.10	.17	.09	.19	.13	.14	.07	.68	.05	.12	.08	.07	.08	.17	.11	1.00					
SE-Coun Diag/Pr	.07	-.09	.15	.06	.14	.05	.16	.12	.13	.08	.03	.10	.12	.07	.07	.07	.21	.07	.52	1.00				
SE-Frain Struc	.07	-.08	.10	.04	.24	.19	.22	.18	.10	.09	.07	.06	.11	.06	.13	.04	.21	.07	.27	.20	1.00			
SE-Frain Motiv	.00	-.06	.04	-.03	.17	.14	.17	.13	.09	.05	.09	-.01	.07	.05	.09	.06	.24	.18	.29	.19	.52	1.00		
SJT-Total	.04	-.05	.21	-.02	.20	.24	.44	.35	.16	.15	.19	.05	.13	.14	.13	.05	.12	.18	.14	.16	.15	.18	1.00	

Note. See Table 5.1 for the full names of the criterion scores.

^a Based on soldiers with complete data for all basic criterion scores except Combat Performance Prediction Scales (N = 863).

^b These correlations are based on the subsample of soldiers rated on the Combat Performance Prediction Scales (N = 821).

Table 5.7

Correlations Between Situational Judgment Test Subscores and Other Selected LVII Basic Criterion Scores

Criterion Score	SJT Subscores					
	Discipl	Focus Positive	Search	Imm/Dir Action	Chain Command	Support
SE-Disc Structure	.11	.03	.02	.07	.06	.05
SE-Disc Comm	.02	.07	.05	.08	.06	.04
SE-Disc Int Skill	.07	.03	.10	.12	.05	.05
SE-Coun Comm/IS	.05	.15	.15	.15	.10	.13
SE-Coun Diag/Prescr	.06	.11	.14	.13	.13	.08
SE-Train Structure	.05	.17	.11	.15	.14	.09
SE-Train Motiv Main	.03	.15	.12	.14	.10	.12
AWB-Leading/Sup	.10	.07	.06	.16	.13	.13
AWB-Tech Skill	.07	.06	.05	.12	.12	.11
AWB-Discipline	.07	.02	.07	.14	.11	.13
AWB-Phys Fit	.08	.02	-.01	.07	.04	.02
Overall Rating	.08	.04	.04	.13	.13	.12
Promotion Rate	.17	.11	.11	.19	.15	.13

Note. Based on all soldiers with complete data (excluding the Combat Performance Prediction Scales; N = 1,144). See Table 5.1 for the full names of the criterion scores and the SJT subscores.

LISREL 7 was used to estimate the parameters and evaluate the fit of each of the alternative models. In this program, confirmatory factor analysis parameters are organized into three matrices:

(1) The factor loadings, modeled with the Lambda X matrix, give the regressions of each observed score on the underlying factors. This matrix was tightly constrained, with each observed variable loading on only one or two factors, and these loadings were estimated by the program.

(2) The covariances among the unobserved variables or factors are represented by the Phi matrix. The diagonal elements of the Phi matrix were fixed to one in the present analyses, so that the Phi elements are actually the correlations among the unobserved variables. Methods factors were constrained to be uncorrelated with each other and with each of the substantive factors. This means that all of the "cross-method" correlation had to be explained by common loadings on substantive factors and by intercorrelations among the substantive factors. The remaining correlations were estimated by the program.

(3) The variances of and covariances among the unique components of each of the observed variables are provided in the final matrix, Theta Delta.

These values indicate the variance in the observed measures that is not accounted for by the factors (i.e., the variance that is not common, or shared, variance). In this sense, each can be viewed as a residual (or error) term arising from the prediction of the observed variable by the factor. These unique components represent the information that would be lost if the data were summarized by scores on the underlying factors and so were treated as measurement error. In the present analyses, the diagonal elements of Theta Delta (the uniquenesses) were estimated. No covariation among the unique components was postulated in the current models, and so all off-diagonal elements of Theta Delta were set to zero.

Evaluation of Model Fit

The LISREL 7 program provides a number of overall fit statistics that can be used in assessing hypotheses about the data. First, there is a chi-square fit statistic that can be used to test the hypothesis that the overall correlation matrix differs from the best-fitting model-based matrix only by sampling error. As Browne and Cudeck (in press) point out, however, the null hypothesis of exact fit is invariably false in practical situations and is likely to be rejected when using large samples. Comparison of the chi-square fit statistics for nested models allows for a test of the significance of the decrement in fit when parameters (e.g., underlying factors) are removed (Mulaik et al., 1989). Second, the Goodness of Fit Index (GFI) is the ratio of the minimum of the fit function after the model has been fitted to the fit function before any model has been fitted; it ranges from zero to one. Finally, the root mean square residual (RMSR) is a measure of the average of the fitted residuals.

One additional fit index was computed that is not provided by the LISREL 7 program. This is the root mean square error of approximation (RMSEA), which can be interpreted as a measure of the discrepancy per degree of freedom for the model (Browne & Cudeck, in press). Because these RMSEA estimates contain a certain amount of error, we also computed the 90 percent confidence interval for each of these estimates. Browne and Cudeck suggest that a value of .08 or less for the RMSEA can be interpreted as indicating a reasonable error of approximation for a model. This fit index is particularly useful because it essentially "penalizes" models that contain more parameters. Additional parameters will not necessarily improve the fit of a model as assessed by the RMSEA, so this fit index does not encourage the inclusion of unimportant or theoretically meaningless parameters just to improve model fit.

RESULTS AND DISCUSSION

Results will be discussed in terms of the confirmation of the CVII performance model, the evaluation of alternative models, and the generalizability of the models across cohorts, across MOS, and across racial subgroups.

Confirmation of the CVII Model

Indices of the overall fit for the Training and Counseling model in the LVII sample are presented in Table 5.8. The fit of this model in the LVII sample is remarkably similar to the fit of this same model in the CVII sample, especially considering that the performance data were collected several years apart using somewhat different measures. Table 5.8 also shows that the fit of

Table 5.8

LISREL Results: Overall Fit Indices for the Training and Counseling Model in the LVII and CVII Samples^a

Sample	N	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^b
<u>LVII Sample</u>						
Total Sample	1,144	652.27	185	.95	.041	.048 (.044-.052)
Excluding MOS 11B	863	562.05	185	.94	.045	.049 (.044-.053)
<u>CVII Sample^c</u>						
Total Sample	1,006	376.76	129	.96	.043	.044 (.039-.049)

^a The basic criterion scores used in modeling performance for these two samples differed somewhat.

^b The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

^c These results differ from those presented in the 1990 annual report. Some constraints on Phi have been omitted, the number-of-courses variable was excluded, and LISREL 7 (in contrast to LISREL VI) was used to estimate the parameters and fit.

this model to the LVII data with MOS 11B soldiers excluded is virtually identical to the fit for the total sample.

The parameter estimates from the LVII sample for the Training and Counseling model are shown in Tables 5.9 and 5.10. Table 5.9 includes the factor loadings and unique variance (Lambda X and Theta Delta), and Table 5.10 presents the correlations among the factors (Phi). These estimates are all very reasonable and are similar to those obtained in the CVII analyses (see Campbell & Oppler, 1990).

Evaluation of Alternative Models

Tests of the Consideration/Initiating Structure model and the variations on this model resulted in a very poor fit to the data (e.g., RMSR values greater than .09) and the program encountered a variety of problems in estimating the parameters for these models (e.g., impossible parameter values, Phi matrices not positive definite, Theta Delta elements not identified).

Table 5.9

LVII LISREL Results for the Training and Counseling Factor Model: Factor Loadings (Lambda X) and Unique Variance (Theta Delta) Parameter Estimates (Maximum Likelihood)

Criterion Score	Factor Loadings (Lambda X)										Unique Variance (Theta Delta)
	CT Core Technical	CP General Proficiency	EL Effort/ Leadership	PD Personal Discipline	PF Physical Fitness	TC Train/ Counsel	Written Method	Rating Method			
PFF-Awards	-	-	.34	-	-	-	-	-	-	-	.89
PFF-Discipline	-	-	-	.41	-	-	-	-	-	-	.83
PFF-Prom Rate	-	-	.58	-	-	-	-	-	-	-	.67
PFF-Phys Read	-	-	-	-	.41	-	-	-	-	-	.84
MO-General	-	.72	-	-	-	-	-	-	-	-	.46
MO-MDS-Specific	.68	-	-	-	-	-	-	-	-	-	.54
JK-General	-	.64	-	-	-	-	.57	-	-	-	.25
JK-MDS-Specific	.62	-	-	-	-	-	.47	-	-	-	.38
AMB-Loading/Sup	-	-	.55	-	-	-	-	.71	-	-	.19
AMB-Tech Skill	-	-	.43	-	-	-	-	.77	-	-	.23
AMB-Discipline	-	-	-	-.54	-	-	-	.59	-	-	.36
AMB-Phys Fitness	-	-	-	-	.82	-	-	.42	-	-	.16
Overall Rating	-	-	.48	-	-	-	-	.75	-	-	.29
MDS Composite	-	-	.45	-	-	-	-	.68	-	-	.34
SE-Disc Struc	-	-	-	-	-	.16	-	-	-	-	.98
SE-Disc Conn	-	-	-	-	-	.33	-	-	-	-	.89
SE-Disc Int Skill	-	-	-	-	-	.28	-	-	-	-	.96
SE-Coun Conn	-	-	-	-	-	.63	-	-	-	-	.61
SE-Coun Diag/Pr	-	-	-	-	-	.55	-	-	-	-	.70
SE-Train Struc	-	-	-	-	-	.55	-	-	-	-	.70
SE-Train Motiv	-	-	-	-	-	.54	-	-	-	-	.71
SJI-Total	-	-	.35	-	-	-	.43	-	-	-	.69

Note. See Table 5.1 for the full names of the criterion scores.

- Represents parameters that were set to zero.

Table 5.10

LVII LISREL Results for the Training and Counseling Factor Model: Factor Correlations (Phi Estimates)

Factor	Core Technical	General Proficiency	Effort/Leadership	Personal Discipline	Physical Fitness	Train/Counsel	Written Method	Ratings Method
Core Technical	1.00							
General Proficiency	.85	1.00						
Effort/Leadership	.62	.56	1.00					
Personal Discipline	-.35	-.34	-.78	1.00				
Physical Fitness	.14	.13	.50	-.57	1.00			
Train/Counsel	.33	.42	.42	-.28	.15	1.00		
Written Method	.00	.00	.00	.00	.00	.00	1.00	
Ratings Method	.00	.00	.00	.00	.00	.00	.00	1.00

To determine whether there were reasonable alternative models of second-tour soldier performance that had been overlooked, a series of exploratory analyses were initiated at this point. The LVII total sample (including MOS 11B) was randomly divided into two subsamples: 60 percent of the sample was used to develop alternative models and 40 percent was set aside for confirming new models that were identified.

The matrix of intercorrelations among the basic criterion scores for the developmental subsample was examined by project staff and several alternative models were tested for fit in the developmental sample. A number of alternatives tried different arrangements of the supervisory simulation, SJT, and rating scale basic scores, while still preserving two leadership factors. None of these alternatives resulted in a good fit with the data. However, a model that collapsed the Consideration and Initiating Structure factors into a single Leadership factor, included a single Simulation Exercise method factor, and moved the promotion rate variable to the new Leadership factor did result in a considerably better fit to the data.

Table 5.11 shows the "Leadership Factor" model that was developed based on these exploratory analyses. Note that this model is very similar to the Leadership factor model tested previously in CVII; however, in the earlier model promotion rate was not included on the Leadership factor. The new LVII model was tested on the holdout sample, and the parameter estimates were very similar to those obtained in the developmental sample. Table 5.12 shows the overall fit indices for this Leadership Factor model using the LVII sample, both with and without MOS 11B, and compares these fit indices with those obtained for the Training and Counseling model. The fit of the new Leadership Factor model to the LVII data is, for all practical purposes, identical to the fit of the Training and Counseling model to these same data. The 90 percent confidence intervals for the RMSEAs (shown in parentheses below the RMSEA estimates) overlap almost completely.

Because these models have equally good fit to the data and because the Leadership Factor model does not confound method variance with substantive variance, the Leadership Factor model was chosen as the best representation of the latent structure of second-tour performance for the LVII data.

The parameter estimates for the Leadership Factor model in the LVII sample are shown in Tables 5.13 and 5.14. A single SJT score (SJT Total Score) was used in the analyses presented on these tables, because all six of the SJT subscores loaded on the same factor (the Leadership factor). Table 5.14 shows that the correlation between the Achievement and Effort factor and the Leadership factor is very high (.94), and the correlation between Core Technical and General Soldiering Proficiency is also quite high (.85).

In retrospect, it seems likely that the high correlation between the Leadership factor and the Achievement and Effort factor is to a large extent due to the high correlation between the Army-wide Leading/Supervising rating and the Army-wide Technical Skill/Effort rating. These two variables correlated .80 with each other, and the Leading/Supervising rating is constrained to load on the Leadership factor while the Technical rating is constrained to load on the Achievement and Effort factor.

Table 5.11

Leadership Factor Model

<u>Latent Variable</u>	<u>Scores Loading on Latent Variables</u>
Core Technical Proficiency (CT)	MOS-Specific Hands-On MOS-Specific Job Knowledge
General Soldiering Proficiency (GP)	General Hands-On General Job Knowledge
Achievement and Effort (AE)	Awards and Certificates Army-Wide Ratings: Technical Skill/Effort Composite Overall Effectiveness Rating MOS Ratings: Overall Composite Combat Prediction: Overall Composite
Personal Discipline (PD)	Disciplinary Actions (reversed) Army-Wide Ratings: Personal Discipline Composite
Physical Fitness/Military Bearing (PF)	Physical Readiness Score Army-Wide Ratings: Physical Fitness/Bearing Composite
Leadership (LD)	Promotion Rate Army-Wide Ratings: Leading/Supervising Composite SE - Disciplinary Structure SE - Disciplinary Communication SE - Disciplinary Interpersonal Skill SE - Counseling Diagnosis/Prescription SE - Counseling Communication/Interpersonal Skills SE - Training Structure SE - Training Motivation Maintenance SJT - Total Score
Written Method	Job-Specific Knowledge General Job Knowledge SJT - Total Score
Ratings Method	Four Army-Wide Ratings Composites Overall Effectiveness Rating MOS Ratings: Total Composite Combat Prediction: Overall Composite
Simulation Exercise Method	All Seven Simulation Exercise Scores

Table 5.12

LVII LISREL Results: Overall Fit Indices for the Training and Counseling and the Leadership Factor Models

Sample	N	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^a
<u>Training and Counseling Model</u>						
Total Sample	1,144	652.27	185	.95	.041	.048 (.044-.052)
Excluding MOS 11B	863	562.05	185	.94	.045	.049 (.044-.053)
<u>Leadership Factor Model</u>						
Total Sample	1,144	649.27	178	.95	.043	.048 (.044-.052)
Excluding MOS 11B	863	556.35	178	.94	.047	.050 (.044-.054)

^a The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

Table 5.13

LVII LISREL Results for the Leadership Factor Model: Factor Loadings (Lambda X) and Unique Variance (Theta Delta) Estimates (Maximum Likelihood)

Criterion Score	Factor Loadings (Lambda X)										Unique Variance (Theta Delta)
	CT Core Technical	CP General Proficiency	AE Achievement/ Effort	PD Personal Discipline	PF Physical Fitness	LD Leadership	Br Written Method	Br Rating Method	SM Simulation Method		
PF-Awards	-	-	.32	-	-	-	-	-	-	-	.90
PF-Discipline	-	-	-	.40	-	-	-	-	-	-	.84
PF-Prun Rate	-	-	-	-	-	.55	-	-	-	-	.70
PF-Phys Read	-	-	-	-	.41	-	-	-	-	-	.84
MO-General	-	.71	-	-	-	-	-	-	-	-	.49
MO-MDS-Specific	.68	-	-	-	-	-	-	-	-	-	.53
JK-General	-	.65	-	-	-	-	.57	-	-	-	.24
JK-MDS-Specific	.63	-	-	-	-	-	.46	-	-	-	.38
AMB-Leading/Sup	-	-	-	-	-	.53	-	.74	-	-	.17
AMB-Tech Skill	-	-	.41	-	-	-	-	.78	-	-	.23
AMB-Discipline	-	-	-	-.52	-	-	-	.60	-	-	.36
AMB-Phys Fitness	-	-	-	-	.81	-	-	.41	-	-	.17
Overall Rating	-	-	.45	-	-	-	-	.77	-	-	.20
MDS Composite	-	-	.43	-	-	-	-	.69	-	-	.33
SE-Disc Struc	-	-	-	-	-	.10	-	-	.12	-	.98
SE-Disc Comm	-	-	-	-	-	.12	-	-	.27	-	.91
SE-Disc Int Skill	-	-	-	-	-	.12	-	-	.15	-	.97
SE-Comm Comm	-	-	-	-	-	.27	-	-	.68	-	.47
SE-Comm Organ/Pr	-	-	-	-	-	.24	-	-	.59	-	.60
SE-Train Struc	-	-	-	-	-	.36	-	-	.32	-	.77
SE-Train Motiv	-	-	-	-	-	.27	-	-	.36	-	.80
SJT-Total	-	-	-	-	-	.38	.42	-	-	-	.68

Note. See Table 5.1 for the full names of the criterion scores.

- Represents parameters that were set to zero.

Table 5.14

LVII LISREL Results for the Leadership Factor Model: Factor Correlations (Phi Estimates)

Factor	Core Technical	General Proficiency	Achievement/Effort	Personal Discipline	Physical Fitness	Training/Counsel	Written Method	Ratings Method	Simulation Method
Core Technical	1.00								
General Proficiency	.84	1.00							
Achievement/Effort	.71	.65	1.00						
Personal Discipline	-.38	-.39	-.73	1.00					
Physical Fitness	.15	.15	.52	-.57	1.00				
Leadership	.56	.63	.96	-.75	.46	1.00			
Written Method	.00	.00	.00	.00	.00	.00	1.00		
Rating Method	.00	.00	.00	.00	.00	.00	.00	1.00	
Simulation Method	.00	.00	.00	.00	.00	.00	.00	.00	1.00

Factor Assignment for Combat Prediction Scales

The Leadership Factor model was tested again with the Combat Performance Prediction Scales included. For one comparison, the Combat Prediction Score was constrained to load only on the Leadership factor and the Rating Method factor. For the second, the Combat Prediction score was constrained to load on the Achievement and Effort and the Rating Method factors only.

The second assignment (i.e., the Combat Prediction Score assigned to the Achievement and Effort factor) produced a much better fit; Table 5.15 presents the resulting overall fit indices for the total sample and for the sample with MOS 11B soldiers excluded. These results indicate that including the Combat Performance Prediction Scales did not affect the overall fit of the model and that this variable fits well on the Achievement and Effort substantive factor.

Table 5.15

LVII LISREL Results: Overall Fit Indices for the Leadership Factor Model With Combat Performance Prediction Scales Included

Sample	N	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^a
Total Sample	1,101	678.84	198	.95	.041	.051 (.047-.055)
Excluding MOS 11B	821	595.54	198	.94	.046	.049 (.045-.054)

* The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

Evaluation of Nested Models

Next, the Leadership Factor model was used as the starting point to develop a nested series of more parsimonious models, similar to those tested in the LVI sample by Oppler, Childs, and Peterson (1994). The first of these nested models was identical to the full Leadership Factor model except that the Achievement and Effort factor was collapsed with the Leadership factor. In other words, these two factors were replaced with a single factor on which all of the variables that had previously loaded on either Achievement and Effort or Leadership were constrained to load.

Similarly, the second nested model was identical to the model just described except that, in addition, the Core Technical and General Soldiering Proficiency factors were replaced with a single "can do" factor. Third, the Personal Discipline factor and the new Achievement/Leadership factor were also collapsed. The fourth model involved adding the variables from the Physical Fitness factor to this Achievement/Leadership/Personal Discipline factor, resulting in a single "will do" factor. The final model collapsed all of the substantive factors into a single overall performance factor.

Evaluating these nested models provides information concerning the extent to which fewer latent variables can account for the observed correlations. Because these more parsimonious models are nested within each other, the significance of the loss of fit can be tested by comparing the chi-square values for the various models. Again, all analyses were conducted twice, once for the total sample and once including only the seven MOS with actual general soldiering scores (i.e., excluding MOS 11B).

Fit indices obtained in testing these nested models for the total sample are shown on Table 5.16, and those obtained in testing these models with MOS 11B excluded are presented on Table 5.17. In general, as the models become more parsimonious (i.e., contain fewer underlying factors) the chi-square values become larger and the fit to the data is not as good. However, in the first nested model, which involved collapsing the Leadership factor with the Achievement and Effort factor, the resulting decrement in fit was very small, and the change in chi-square was very small (7.9 with 5 degrees of freedom). Similarly, collapsing the two "can do" factors resulted in a very small reduction in model fit. Based on these results, a model with only four substantive factors (and three method factors) can account for the data almost as well as the full Leadership Factor model.

Collapsing additional factors beyond this level resulted in larger decrements in model fit. The model with a single substantive factor has an RMSR value of .058, indicating that even this model accounts for a fair amount of the covariation among the LVII basic criterion scores. It should be remembered that this model still includes the three method factors (Written, Ratings, and Simulation Exercise), so this result is partly a reflection of the fact that a good deal of the covariation among these scores is due to shared measurement method.

The next to last model that is presented on both Table 5.16 and Table 5.17 includes two substantive factors: "can do" and "will do." Because the "will do" factor in this model contains the Leadership factor from the full model, it includes the Supervisory Simulation Exercise and SJT scores. However, both the SJT and the Supervisory Simulations are measures of maximal performance, so these measures might be better placed on the "can do" factor.

Therefore, a modified "can do/will do" model was tested that constrained the seven Simulation Exercise scores and the SJT score to load on the "can do" rather than the "will do" factor. The RMSR for this modified model was .048 and the RMSEA was .053 (compared with .050 and .056 for the original "can do/will do" model), indicating that the SJT and Simulation scores do fit somewhat better with the "can do" than with the "will do" measures.

A wide variety of additional nested analyses were also conducted to determine how the order in which the factors are collapsed affects the fit of the resulting models. These results, taken as a whole, indicated that the order in which the factors were originally collapsed (see Table 5.16) results in the smallest decrement in model fit at each stage.

Table 5.16

LVII LISREL Results: Overall Fit Indices for a Series of Nested Models That Collapse the Substantive Factors in the Leadership Factor Model, Based on Total Sample Data

Model	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^a
Full Model	649.27	178	.95	.043	.048 (.044-.052)
Single Achievement/ Leadership Factor	657.17	183	.95	.043	.048 (.044-.052)
Single "Can Do" Factor	686.58	187	.95	.043	.048 (.044-.052)
Single Achievement/Leadership/ Personal Discipline Factor	739.38	190	.94	.045	.050 (.047-.054)
Single "Will Do" Factor	875.92	192	.93	.050	.056 (.052-.060)
Single Substantive Factor	999.93	193	.92	.058	.060 (.057-.064)

Note. N = 1,144.

^a The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

For example, if the Achievement and Effort factor is first collapsed with the Personal Discipline factor rather than with the Leadership factor, the resulting model fit is much worse than the comparable model on Table 5.16 in which Achievement and Effort is collapsed with Leadership. Similarly, if the Leadership factor is collapsed with the "can do" factor rather than with the Achievement and Effort factor, the result is a much larger decrement in fit. Based on these results, the models shown on Table 5.16 appear to represent the optimal set of more parsimonious models.

Table 5.17

LVII LISREL Results: Overall Fit Indices for a Series of Nested Models That Collapse the Substantive Factors in the Leadership Factor Model, for Sample Excluding MOS 11B

Model	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^a
Full Model	556.35	178	.94	.047	.050 (.044-.054)
Single Achievement/ Leadership Factor	562.58	183	.94	.048	.049 (.044-.054)
Single "Can Do" Factor	593.14	187	.94	.049	.050 (.046-.055)
Single Achievement/Leadership/ Personal Discipline Factor	637.26	190	.94	.051	.052 (.048-.057)
Single "Will Do" Factor	764.72	192	.92	.056	.059 (.054-.063)
Single Substantive Factor	851.70	193	.91	.060	.063 (.059-.067)

Note. N = 863.

^a The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

Retrospective Re-Analysis of the CVII Data

One final approach to confirming the Leadership Factor model was to assess the fit of this new model to the CVII data. Table 5.18 shows the fit of the full Leadership Factor model to the CVII as well as the fit of the series of more parsimonious nested models. These results are virtually identical to those obtained in the LVII data (shown on Table 5.16), providing additional confirmation for the Leadership Factor model.

Table 5.18

CVII LISREL Results: Overall Fit Indices for a Series of Nested Models That Collapse the Substantive Factors in the Leadership Factor Model

Model	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^a
Full Model	353.66	124	.96	.040	.043 (.038-.048)
Single Achievement/ Leadership Factor	370.83	129	.96	.040	.043 (.038-.048)
Single "Can Do" Factor	430.10	133	.96	.042	.047 (.042-.052)
Single Achievement/Leadership/ Personal Discipline Factor	464.80	136	.95	.043	.049 (.044-.054)
Single "Will Do" Factor	574.27	138	.94	.048	.056 (.051-.061)
Single Substantive Factor	722.83	139	.92	.054	.065 (.060-.069)

Note. N = 1,006.

^a The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

Generalizability Across MOS

Analyses were also conducted to determine whether the Leadership Factor model fits equally well for all eight MOS included in the present research. Within-MOS sample sizes were not large enough to allow for separate modeling analyses for each MOS, so clusters of similar MOS were identified on the basis of their task content. The eight MOS included in the present analyses were clustered on the basis of the results of previous research by Wise et al. (1991), in which job experts used a 96-item job analysis questionnaire to describe the task content of each Project A MOS. These MOS were then clustered according to the similarity of their job task content.

These results were used in the present research to identify three clusters of MOS. The first cluster included the 11B, 13B, 19K, and 95B MOS. As in the total sample analyses, the Leadership Factor model was tested twice for this cluster, once including MOS 11B (with the "imputed" general soldiering scores) and once excluding 11B. The second cluster included MOS 71L and 91A/B. Finally, the third cluster expanded the second cluster to also include MOS 63B and 88M.

Attempts to fit the Leadership Factor model to the LVII data for each of these clusters of MOS resulted in problems in estimating the model parameters, particularly the elements of the Phi matrix (factor correlations). Several analyses resulted in impossibly large correlations between the Leadership factor and the Achievement and Effort factor. To alleviate this problem, these analyses were run again with these two factors collapsed to form a single Achievement/Leadership factor, parallel to what was done in the evaluation of more parsimonious models.

Results of this second set of analyses are presented on Table 5.19. In general, the fit is about equally good for all of the various MOS clusters, although the fit for the cluster of 71L and 91A is somewhat worse than for the others. Although not presented here, the parameter estimates were also generally similar across MOS clusters.

Table 5.19

LVII LISREL Results: Overall Fit Indices for the Leadership Factor Model With One Factor^a Modified, for Clusters of MOS

MOS Included	N	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^b
11B, 13B, 19K 95B	633	431.30	183	.94	.050	.046 (.041-.052)
13B, 19K 95B	352	328.43	183	.92	.052	.048 (.039-.056)
71L, 91A/B	285	290.33	183	.92	.056	.045 (.035-.055)
63B, 71L, 88M, 91A/B	511	441.69	183	.93	.053	.053 (.046-.059)

^a The Achievement and Effort factor was collapsed with the Leadership factor in these analyses.

^b The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

Generalizability Across Racial Subgroups

Analyses were also conducted to determine whether this Leadership Factor model fits equally well for racial subgroups. There was not a large enough group of females in the LVII sample to conduct separate modeling analyses for males and females.

The only two racial subgroups large enough for separate modeling analyses were blacks and whites. As in the analyses for the MOS clusters, the Leadership and the Achievement and Effort factors were collapsed in order to

avoid problems in estimating the elements of the Phi matrix. Even so, the program encountered serious problems in estimating the model parameters in the black subsample. Many of these problems were related to the Physical Fitness/Military Bearing factor. The variables that load on this factor, especially the physical readiness variable, tend to have lower correlations with variables on the Achievement and Effort factor for blacks than they do for whites. Correlations between the Leadership factor variables and those on Achievement and Effort also appear somewhat lower for blacks.

The racial subgroup analyses were rerun with the two variables that load on the Fitness/Bearing factor (the Fitness/Bearing rating composite and the Physical Readiness score) and the factor itself excluded. Results are shown in Table 5.20. When the Physical Fitness/Military Bearing factor is excluded, model fit is very similar for the black and white subsamples.

Table 5.20

LVII LISREL Results: Overall Fit Indices for the Leadership Factor Model With Two Factors^a Modified, by Race

Race	N	Chi-Square	df	GFI	RMSR	RMSEA (CI) ^b
Whites	637	288.28	149	.94	.051	.046 (.038-.054)
Blacks	333	256.48	149	.93	.055	.047 (.037-.056)

^a The Achievement and Effort factor was collapsed with the Leadership factor in these analyses. The Army-wide Physical Fitness/Military Bearing rating and the Physical Readiness score were excluded.

^b The 90% confidence interval for each RMSEA estimate is shown in parentheses below the estimate.

CREATING LVII CRITERION CONSTRUCT SCORES FOR VALIDATION ANALYSES

The basic criterion construct scores for use in validation analyses are based on the full Leadership Factor model, with six substantive factors (shown in Table 5.11). The nested model with four factors (with a single Achievement/Leadership factor and a single "can do" factor combining Core Technical and General Soldiering Proficiency) fits the data almost as well and has the advantage of greater parsimony. However, it is still plausible that all six performance factors have somewhat different antecedents and could be related to different predictor constructs. Therefore, for the initial validity analyses the model that incorporates the six criterion construct scores will be retained. A description of the computation of the six performance factor scores follows.

The Core Technical Proficiency factor is composed of two basic scores: the job-specific score from the hands-on tests and the job-specific score from the job knowledge tests.

Similarly, the General Soldiering Proficiency factor is composed of two basic scores: the general soldiering score from the hands-on tests and the general soldiering score from the job knowledge tests. Soldiers from MOS 11B do not have scores on this construct because no distinction is made between core technical and general soldiering tasks for this MOS.

The Personal Discipline factor is composed of the Personal Discipline composite from the Army-wide ratings, which is the average of ratings on three different scales (Following Regulations/Orders, Integrity, and Self-Control), and the disciplinary actions score from the Personnel File Form.

The Physical Fitness and Military Bearing factor is also composed of two basic scores: the Physical Fitness and Military Bearing composite from the Army-wide ratings, which is the average of ratings made on two scales (Military Appearance and Physical Fitness) and the physical readiness score, which was collected on the Personnel File Form.

The Achievement and Effort criterion factor is composed of four composite scores and the single rating of overall effectiveness. The four composites are: (a) the Technical Skill/Effort composite from the Army-wide ratings (the average of ratings on Technical Knowledge/Skill, Effort, and Maintain Assigned Equipment); (b) the overall MOS composite, which is the average across all of the behavior-based MOS-specific rating scales; (c) the overall Combat composite which is the sum of the Combat Performance Prediction scales; and (d) the awards and certificates score from the Personnel File Form. Scores for the three rating composites (a, b, and c) were first combined, with each of the individual scores unit weighted. This score was then treated as a single subscore and combined with the two remaining subscores (i.e., the awards and certificates score, and the overall effectiveness rating).

The sixth criterion construct, Leadership, is made up of four major components. The first is the unit-weighted sum of all seven basic scores from the Personal Counseling, Training, and Disciplinary Simulation Exercises. The second is the Leading/Supervising score from the Army-wide ratings, which is the average across nine rating scales related to leadership and supervision. The third is the total score from the Situational Judgment Test, and the fourth is the Promotion Rate score from the Personnel File Form.

In computing scores for each of these factors, the major subscores were unit weighted. That is, they were combined by first standardizing each within MOS and then adding them together. These scoring procedures gave approximately equal weight to each measurement method, minimizing potential measurement bias for the resulting criterion construct scores. Table 5.21 shows the intercorrelations among these six criterion construct scores and their correlations with each of the LVII basic criterion scores.

Table 5.21

Correlations of LVII Basic Criterion Scores With Proposed Construct Scores

Criterion Scores	Constructs					
	CT Core Technical	GP General Proficiency ^a	AE Achievement/Effort	PD Personal Discipline	PF Physical Fitness	LD Leadership
JK-General ^a	.50	.85 *	.22	.10	.01	.44
JK-MOS-Specific	.85 *	.49	.26	.15	.06	.43
HO-General ^a	.37	.86 *	.20	.13	.10	.33
HO-MOS-Specific	.85 *	.37	.23	.11	.11	.32
AHB-Leading/Sup	.29	.22	.79	.55	.39	.65 *
AHB-Tech Skill	.25	.20	.82 *	.51	.36	.49
AHB-Discipline	.21	.18	.62	.79 *	.34	.45
AHB-Phys Fitness	.09	.04	.52	.46	.82 *	.35
Overall Rating	.27	.19	.86 *	.54	.41	.52
MOS Composite	.31	.21	.78 *	.45	.33	.48
Combat Composite ^b	.24	.24	.71	.49	.37	.47
PFF-Awards	.10	.15	.58 *	.13	.17	.23
PFF-Discipline	-.03	-.03	-.19	-.30 *	-.22	-.20
PFF-Prom Rate	.26	.24	.32	.26	.26	.67*
PFF-Phys Read	.07	.06	.15	.12	.81*	.13
SJT-Total	.36	.37	.17	.14	.03	.64 *
SE-Disc Struc	.08	.06	.02	-.03	-.01	.23 *
SE-Disc Comm	.03	.15	.04	-.01	-.01	.28 *
SE-Disc Int Skill	.03	.11	.02	.06	.07	.24 *
SE-Coun Comm	.14	.21	.12	.09	.09	.43 *
SE-Coun Diag/Pr	.11	.17	.12	.09	.07	.40 *
SE-Train Struc	.24	.27	.13	.09	.10	.40 *
SE-Train Motiv	.16	.20	.07	.09	.00	.37 *
CT Construct	1.00					
GP Construct	.51	1.00				
EA Construct	.29	.24	1.00			
PD Construct	.15	.13	.51	1.00		
PF Construct	.10	.06	.41	.36	1.00	
LD Construct	.44	.45	.55	.41	.30	1.00

Note: Correlations are based on a sample of 1,144 unless otherwise specified. See Table 5.1 for the full names of the criterion scores.

* Indicates the variables that were used in computing construct scores.

^a Correlations are based on all soldiers except MOS 11B (N = 863), because this MOS does not have these scores.

^b Correlations are based on the subset of soldiers who were rated on the Combat Scales (N = 1,101); the correlation with General Soldiering Proficiency excludes MOS 11B as well (N = 821).

Because Combat Performance Prediction ratings were not available for all members of the LVII sample, the Combat Prediction Performance overall composite score was not included in computing the Achievement and Effort composite score used in the correlations shown in Table 5.21. Table 5.22 shows the correlations of the other criterion construct scores with two versions of the Achievement and Effort composite: one that includes the Combat Prediction scores and one that does not. These two sets of correlations are virtually identical. Table 5.22 also shows that, as expected, the correlation of the Achievement and Effort composite score with the Combat Prediction score is higher when the Combat score is included in computing the Achievement and Effort composite.

Table 5.22

Correlations Between Two LVII Versions of the Achievement and Effort Construct Score (With and Without the Combat Prediction Score) and Other Proposed Construct Scores and the Combat Prediction Overall Composite Score

	Core Technical	General Proficiency ^a	Personal Discipline	Physical Fitness	Leadership	Combat Prediction Score
Achievement/Effort With Combat Prediction Score	.30	.26	.52	.42	.56	.77
Achievement/Effort Without Combat Prediction Score	.31	.25	.51	.41	.56	.71

Note: The correlation between the two versions of the Achievement and Effort construct score is .99. All correlations are based on the subsample of soldiers who were rated on the Combat Performance Prediction Scales (N = 1,101).

^a Correlations are based on all soldiers except MOS 11B (N = 821), because the 11B MOS does not have this score.

Results of the nested analyses were used to form more parsimonious sets of criterion construct scores as well. This was done by first standardizing each of the six construct scores described above (based on the full Leadership model). These were then added together in the order shown on Figure 5.1 to form sets of five, four, three, two and finally one criterion composite construct score.

CONCLUDING COMMENTS

Results of the LVII modeling analyses reported in this chapter show that both the Training and Counseling model and the Leadership Factor model fit the LVII data quite well. Further, retrospective reanalysis of the CVII data showed that these two models had a similarly good fit in the CVII sample.

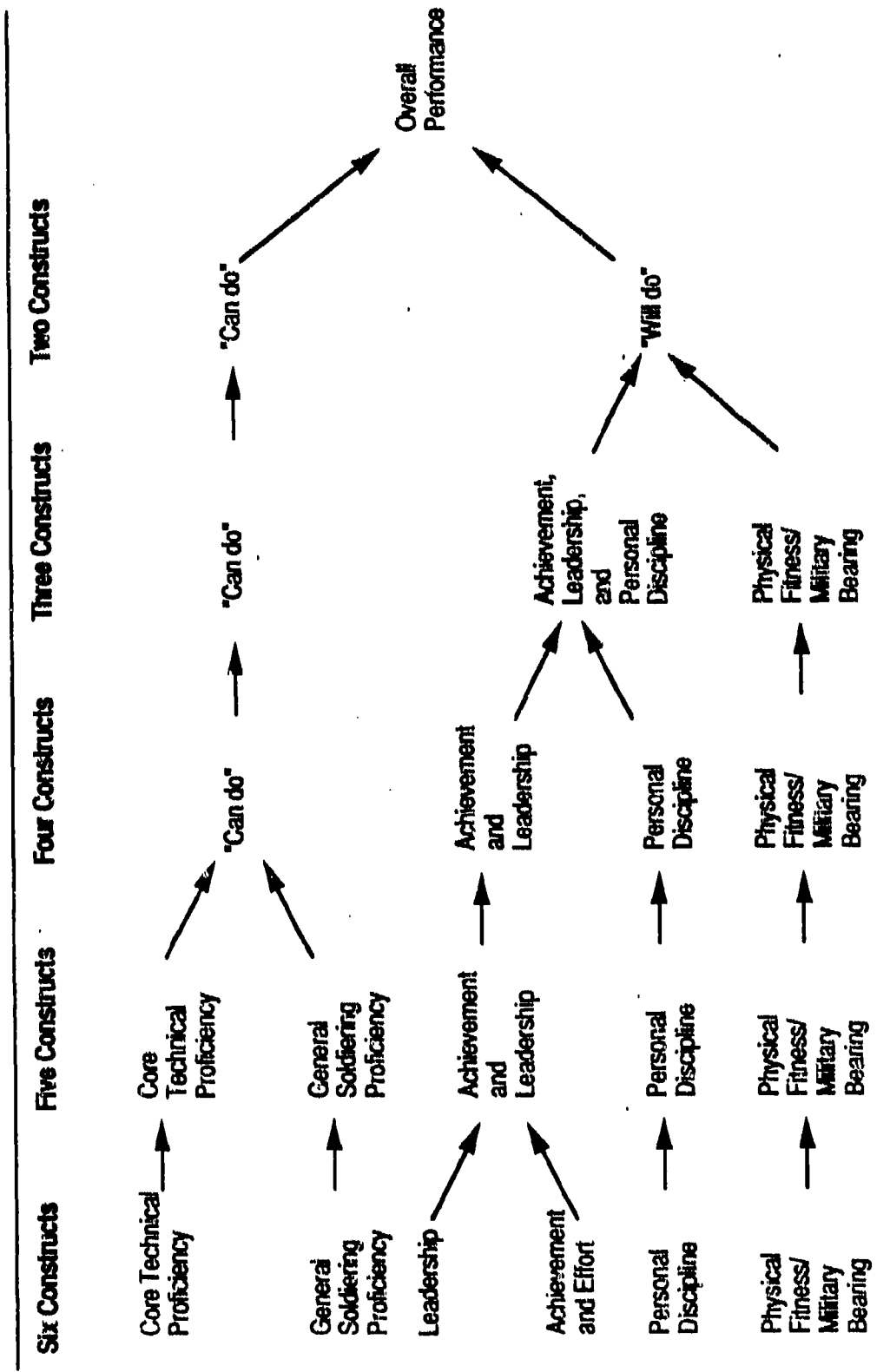


Figure 5.1. Final LVII Criterion and Alternate Criterion Constructs based on more parsimonious models.

Because the factors in the Leadership Factor model do not confound method and substantive variance, this model was chosen as the best representation of the latent structure of second-tour soldier performance.

Results of the modeling analyses conducted on subgroups identified on the basis of race and MOS provide evidence that, in general, the model fits equally well for soldiers from different MOS and for black and white soldiers. However, the variables loading on the Physical Fitness/Military Bearing construct behave much differently for blacks than for whites. When these variables are excluded, the Leadership Factor model fits about equally well for blacks and whites.

Efforts to identify more specific leadership components within the general leadership factor were not successful, even though the LVII contained a greater variety of basic criterion scores related to leadership than did the CVII. This could indicate that the current performance measures are not sensitive to the latent structure of leadership performance or that leadership responsibilities at the junior NCO level are not yet well differentiated, or that the latent structure is actually unidimensional. Given the robust findings from the previous literature that argue for multidimensionality, the explanation is most likely some combination of the first two reasons.

The promotion rate variable was included on the Leadership construct mainly because it was expected to share a great deal of variance with leadership and supervisory performance. Soldiers with more leadership potential are more likely to be promoted, and soldiers who have been promoted more are likely to have obtained more experience in leading and supervising other soldiers. The fact that promotion rate fit very well on the Leadership factor confirmed the expectation.

The new six-factor Leadership Factor model of second-tour performance is also consistent with the CVI/LVI model of first-tour soldier performance. In addition to including performance factors that are parallel to those identified for first-tour soldiers, the LVII second-tour model includes a Leadership factor that contains all measures that were in fact targeted at the leadership/supervision aspects of the job. This is consistent with the results of the second-tour job analyses which indicated that second-tour soldiers perform many of the same tasks as the first-tour soldiers in addition to their supervisory responsibilities. In sum, the Leadership Factor model provides the starting point for the LVII validity analyses and further enhances our understanding of second-tour soldier performance.

Chapter 6 OVERALL SUMMARY AND FUTURE PLANS

John P. Campbell

During the third year of the Career Force Project, the major emphases were on (a) completing the second-tour Longitudinal Validation (LVII) data collection, (b) preparing the LVII data files for analysis, and (c) analyzing the covariance structure of the second-tour performance measures using the LVII sample data. The LVII sample is the major data source for estimating the validity with which NCO performance during the second tour can be predicted from selection and classification tests administered at the time of accession, from performance during training, and from job performance during the first tour of duty.

SUMMARY OF YEAR THREE

In one sense, much of the work described in this third annual report is a replication of a similar data collection and data analysis in the second-tour Concurrent Validation sample (CVII). The same basic array of criterion measures was used to collect performance data from junior NCOs who had been in the Army from 5 to 6 years. Using the CVII sample, the scale- and task-level data were used to define a set of "basic" criterion scores for each type of instrument (e.g., four "scores" were obtained from the individual Army-wide rating scales), and alternative models for the latent structure of second-tour performance were evaluated in terms of their fit to the observed covariance of the basic criterion scores.

However, the LVII sample and its subsequent analyses are more than a replication of CVII. First, the lessons learned in the CVII data collection were used to improve the LVII data collection. For example, selected members of the project staff were carefully trained as role players and scorers for the Supervisory Simulation Exercises. Also, the Situational Judgment Test was item analyzed, revised, and expanded. Second, the sample sizes for MOS were designed to be larger, and much greater effort was expended to include as many individuals from the LVI sample as possible. Third, the LVII sample was intended as a true confirmatory test for the basic criterion score definitions and the model of second-tour performance that were proposed on the basis of the CVII analysis. In this sense, the LVII data collection and criterion analyses were very much not a replication of the CVII results. They were a relatively stringent test of the validity of the hypothesized structure of NCO job performance.

The LVII Data Collection

During year three, the first major order of business was to complete the LVII sample data collection. The original plan called for assessing at least 150 soldiers in each of the nine Batch A MOS who had also been in the first-tour longitudinal sample (LV), and who had been assessed on the Experimental Predictor Battery (EB), the training performance (EOT) measures, and the first-tour job performance (LVI) measures as part of Project A. The original data collection plan called for data collection teams to visit 15 sites between May 1991 and February 1992. However, in this instance, the best-laid plans were influenced by more than the usual number of perturbations. The

principal unanticipated factor was Operation Desert Shield/Desert Storm, which prevented some data collections and significantly delayed a number of others. The LVII data collection was finally concluded in July 1992.

The dates for the final data collections were such that the available "window" for assessing second-tour NCOs who had also been in the LVI sample was pushed to its limit. That is, the people in the LVI sample who had reenlisted for a second tour were beginning either to leave the Army or to reenlist for a third tour of duty. As a consequence, the proportion of the LVII sample who also participated in LVI was somewhat smaller than it otherwise might have been. The delay created by Desert Shield/Desert Storm also lengthened the average tenure for individuals in the LVII sample, as compared to the CVII sample. On average, they had been in the Army about 4-6 months longer. A longer tenure adds to the quality of the sample, losing people who had also participated in LVI detracts from it.

In sum, and although the data collection schedule was delayed, the Project succeeded in using improved data collection methods to produce, in comparison to CVII, a larger sample that incorporated a higher percentage of people who had also been assessed during their first tour of duty.

Analysis of Basic Criterion Scores

The revision of the criterion measures of second-tour performance benefitted greatly from the analysis of the CVII data. The Situational Judgment Test and the supervisory role-play simulations were the most extensively revised. However, the currency and content validity of the hands-on, job knowledge, MOS rating, and the Personnel File Form measures were also improved.

The psychometric characteristics of the criterion measures were the same or better in LVII than in CVII. In general, they tended to yield somewhat greater variance. The way in which the item, scale, step, and task scores were aggregated into a more manageable set of "basic scores" was virtually identical to CVII.

In some cases, this was by design. For the hands-on and job knowledge measures there were no compelling reasons to alter the scoring rules used in CVII. For the ratings measures, the LVII factor analytic results were virtually identical to those obtained in the CVII sample. For example, the individual factor loadings of the Army-wide rating scales on each of the four factors seldom differed between the two samples except in the second decimal place. As reported previously, the same result was obtained when the CVI and LVI factor structures were compared (Campbell & Zook, 1990). In both instances there is great stability in the factor structure of the ratings, in spite of the presence of a large general factor.

In total, the second-tour performance measures showed great consistency between CVII and LVII in terms of their psychometric properties, their content validity, and their intercorrelations. Behavioral science data does replicate.

Editing the LVII Data File

One major consequence of using the CVII experience to improve the data collection procedures for LVII is that the quality of the data did improve commensurately. That is, there was comparatively less incorrect, incomplete, or missing data. As noted in Chapter 4, relatively few cases were dropped when the same decision rules that were used in CVII were used in LVII to set scores to missing. Consequently, it was not necessary to use regression-based imputation procedures to obtain scores for individuals with partially missing data. In the very small percentage of cases where missing data treatments were applied, the individual's mean score on the available items or steps was used. Based on previous evaluations of imputation procedures (Campbell & Zook, 1994), the partial data treatments applied to the LVII data file should not alter the distributions or intercorrelations of the criterion scores.

Analysis of the Second-Tour Performance Model

The results of the LVII modeling analyses were gratifying. Even though some changes were made in the criterion measures, the CVII performance model fit the LVII data as well as it fit the CVII sample data on which it had been developed 3 years before. Improvements in the Situational Judgment Test and the supervisory role-play simulation exercises also permitted an expanded set of basic scores to be computed from these two measures. This permitted the specification of an alternative LVII performance model that was able to unconfound substantive and method variance regarding the basic scores that depended on the role-play measurement method.

Method variance attributable to the role-play method could not be accounted for by CVII basic scores. However, when the LVII model was retrospectively fit to the CVII sample data (recognizing that the item composition for the SJT scores is not identical), the LVII model fit the CVII sample data as well as it did the LVII data. Consequently, two models have been identified that fit the data equally well, and equally well in both samples. The LVII model was selected as the validation model of choice because it provides a multiple-method definition of the leadership factor, rather than confounding substantive and method variance for that factor.

The LVII modeling analysis also showed that a hierarchically nested model collapsed into either five or four latent factors fit the data almost as well as the full six-factor LV model. However, nested models that collapsed the six factors into three or two factors did not fit the data nearly as well. Future validation analyses will use factor scores from both the six-factor model and composite factor scores from the hierarchical collapsed models to determine whether the full six-factor version will yield differential prediction information that is covered up by the aggregated factors.

FUTURE PLANS

All the major data collections that were designed as part of Project A and the Career Force Project are now complete, and all major data files are edited and in place. Consequently, during year four the Career Force Project will concentrate exclusively on a number of data analysis objectives.

The first order of business will be to carry out the basic LVII validation analyses. This will entail estimating (a) the zero-order validities of the ASVAB and Experimental Battery tests for predicting each of the LVII performance factors and composite factors, (b) the validities of each of the regression-weighted predictor domains, and (c) the incremental validities (over ASVAB) for each of the Experimental Battery predictor domains. Results will be compared to those obtained in CVI, LVI, and CVII.

Completion of the LVII data file also makes it possible to estimate the validity of prediction of second-tour job performance from first-tour job performance and from training performance. Future analyses will examine these relationships in terms of their convergent and divergent patterns across performance factors. That is, if performance really has a multidimensional latent structure, and if the latent structure is consistent across cohorts and across organizational levels (i.e., first tour vs. second tour), then scores on a particular performance factor in LVI should have a higher correlation with that factor in LVII than with any of the non-correspondent factors.

The final step in this sequence of analyses will be to consider the accuracy with which (a) training performance can be predicted from the test battery, (b) first-tour job performance can be predicted from the test battery plus training performance, and (c) second-tour performance can be predicted from the test battery plus training performance plus first-tour performance. This is the full "Roll-Up Model" originally envisioned by the framers of the Project A Statement of Work.

Attrition is also a criterion variable of major interest for the military services. Attrition data are now available for the first-tour Longitudinal Validation sample (LVI) and are part of the Career Force Project data file. The validity of ASVAB and the Experimental Battery for predicting attrition is being examined, using both traditional regression methods and survival analysis. The latter provides information about how accurately the time at which attrition will take place can be predicted.

The LVI data file also includes the data from the administration of the Army Job Satisfaction Questionnaire (AJSQ). A series of analyses are being conducted that focus on (a) job satisfaction as a criterion outcome to be predicted from accession data, the ASVAB, and the Experimental Battery, and (b) job satisfaction as a correlate of performance and attrition. The results of these analyses will be presented in the next annual report.

Finally, during its last year, the Career Force Project will be concerned with a number of analyses that focus on the utility of the Experimental Battery for making classification decisions. These will include an analysis of the optimal set of prediction equations that best reflect the level of differential prediction across performance factors and across MOS, an examination of using empirical keying to maximize classification validity, and an exploration of how the specificity of the criterion measure influences estimates of differential prediction and classification validity.

When the above analyses have been finished, the Career Force Project will be concluded and the information base that is necessary for building a model, or test bed, of the Army job assignment system will have been completed.

References

- Browne, M. W., & Cudeck, R. (in preparation). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models. Beverly Hills, CA: Sage.
- Campbell, C. H., Ford, P., Rumsey, M. G., Pulakos, E. D., Borman, W. C., Felker, D. B., de Vera, M. V., & Riegelhaupt, B. J. (1990). Development of multiple job performance measures in a representative sample of jobs. Personnel Psychology, 43, 277-300.
- Campbell, J. P. (Ed.) (1987). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year (ARI Technical Report 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A193 343)
- Campbell, J. P. (Ed.) (1988). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year (ARI Technical Report 792). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A198 856)
- Campbell, J. P. (Ed.) (1989). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1987 fiscal year (ARI Technical Report 862). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A219 046)
- Campbell, J. P. (Ed.) (1991). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1988 fiscal year (ARI Research Note 91-34). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 750)
- Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. Personnel Psychology, 43, 313-333.
- Campbell, J. P., & Oppler, S. (1990). Modeling of second-tour performance. In J. P. Campbell & L. M. Zook (Eds.), Building and retaining the Career Force: New procedures for accessing and assigning Army enlisted personnel--Annual report, 1990 fiscal year (ARI Technical Report 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A252 675)
- Campbell, J. P., & Zook, L. M. (Eds.) (1990). Building and retaining the Career Force: New procedures for accessing and assigning Army enlisted personnel--Annual report, 1990 fiscal year (ARI Technical Report 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A252 675)

- Campbell, J. P., & Zook, L. M. (Eds.) (1991). Improving the selection, classification, and utilization of Army enlisted personnel: Final report on Project A (ARI Research Report 1597). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A242 921)
- Campbell, J. P., & Zook, L. M. (Eds.) (1994). Building and retaining the Career Force: New procedures for accessing and assigning Army enlisted personnel--Annual report, 1991 fiscal year (ARI Research Note 94-10). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A278 726)
- Campbell, R. C. (1985). Scorer training materials (ARI WP-SC-85-06). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Claudy, J. G. (1978). Multiple regression and validity estimation in one sample. Applied Psychological Measurement, 2, 295-601.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. Psychological Bulletin, 105, 317-327.
- Cureton, E. E. (1965). Reliability and validity: Basic assumptions and experimental designs. Educational and Psychological Measurement, 25, 327-346.
- Fleishman, E. A. (1973). Twenty years of consideration and structure. In E. A. Fleishman & J. G. Hunt (Eds.), Current developments in the study of leadership. Carbondale, IL: U. of Southern Illinois Press.
- Fleishman, E. A., Zaccaro, S. J., & Mumford, M. D. (1991). Individual differences and leadership: An overview. Leadership Quarterly, 2(4), 237-243.
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), Review of Personality and Social Psychology (Vol. 2, pp. 141-165). Beverly Hills, CA: Sage.
- Hanson, M. A., & Borman, W. B. (In preparation). Development and construct validation of the Situational Judgment Test (SJT) (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Harris, C. W., & Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. Psychometrika, 29, 347-362.
- Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. British Journal of Statistical Psychology, 17, 65-70.

- Hough, L. M. (1992). The "Big Five" personality variables--construct confusion: Description versus prediction. Human Performance, 5, 139-155.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those variables (Monograph). Journal of Applied Psychology, 75, 581-595.
- Jöreskog, K. G., & Sörbom, D. (1986). LISREL VI: Analysis of linear structural relationships by the method of medium likelihood. Uppsala, Sweden: University of Uppsala.
- Jöreskog, K. G., & Sörbom, D. (1989a). LISREL 7: A guide to the program and applications (2nd ed.). Chicago: SPSS.
- Jöreskog, K. G., & Sörbom, D. (1989b). LISREL 7: User's Reference Guide (1st ed.). Scientific Hardware, Inc.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. Personnel Psychology, 43, 335-354.
- Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 66, 574-583.
- Oppler, S. H., Childs, R. A., & Peterson, N. G. (1994). Development of the longitudinal validation sample first-tour performance model. In J. P. Campbell & L. M. Zook (Eds.), Building and retaining the Career Force: New procedures for accessing and assigning Army enlisted personnel--Annual report, 1991 fiscal year (ARI Research Note 94-10). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A278 726)
- Peterson, N., Russell, T., Hallam, G., Hough, L., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1990). Analysis of the experimental predictor battery: LV sample. In J. P. Campbell & L. M. Zook (Eds.), Building and retaining the Career Force: New procedures for accessing and assigning Army enlisted personnel--Annual report, 1990 fiscal year (ARI Technical Report 952, pp. 73-199). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A252 675)
- Pulakos, E. D., & Borman, W. C. (Eds.) (1986). Development and field test of Army-wide rating scales and the rater orientation and training program (ARI Technical Report 716). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD B112 857)

- Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. Psychological Bulletin, 85, 1348-1351.
- Wise, L. L., McHenry, J. J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. Personnel Psychology, 43, 355-366.
- Wise, L. L., McHenry, J. J., & Young, W. Y. (1986). Project A Concurrent Validation: Treatment of Missing Data. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. Unpublished manuscript.
- Wise, L. L., Peterson, N. G., Hoffman, R. G., Campbell, J. P., & Arabian, J. M. (1991). The Army Synthetic Validity Project: Report of Phase III Results, Vol. I (ARI Technical Report 922). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A235 635)

Appendix A

Tasks Comprising the Hands-On and Job Knowledge Components by MOS (LVII)

Table A-1

Tasks Tested: 11B

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>11B Infantryman</u>					
Evaluate a casualty		x	First Aid	Safety/Survival	MOS-Specific
Put on a field or pressure dressing	x	x	First Aid	Safety/Survival	MOS-Specific
Practice preventive medicine		x	First Aid	Safety/Survival	MOS-Specific
Navigate on the ground	x	x	Navigate	Basic Soldiering	MOS-Specific
Determine grid coordinates		x	Navigate	Basic Soldiering	MOS-Specific
Orient map by terrain assoc		x	Navigate	Basic Soldiering	MOS-Specific
Decontaminate your skin	x	x	Nuc/Bio/Chem	Safety/Survival	MOS-Specific
Check soldiers in MOPP4		x	Nuc/Bio/Chem	Safety/Survival	MOS-Specific
Conduct unmasking procedures		x	Nuc/Bio/Chem	Safety/Survival	MOS-Specific
Engage target w/M16		x	Weapons	Basic Soldiering	MOS-Specific
Zero M249 machinegun		x	Weapons	Basic Soldiering	MOS-Specific
Engage target w/M72A2 LAW	x	x	Weapons	Basic Soldiering	MOS-Specific
Engage target w/M60		x	Weapons	Basic Soldiering	MOS-Specific
Engage target w/.50		x	Weapons	Basic Soldiering	MOS-Specific
Prepare M47 for firing		x	Weapons	Basic Soldiering	MOS-Specific
Operate AN/PVS-4	x	x	Weapons	Basic Soldiering	MOS-Specific
Zero AN/PVS-4		x	Weapons	Basic Soldiering	MOS-Specific
Call/adjust indirect fire		x	Field Techniques	Basic Soldiering	MOS-Specific
Select overwatch position		x	Field Techniques	Basic Soldiering	MOS-Specific
React to ambush		x	Field Techniques	Basic Soldiering	MOS-Specific
Conduct defense by squad		x	Field Techniques	Basic Soldiering	MOS-Specific
Perform movement MOUT	x	x	Field Techniques	Basic Soldiering	MOS-Specific
Control fire team		x	Field Techniques	Basic Soldiering	MOS-Specific
Control organic fires		x	Field Techniques	Basic Soldiering	MOS-Specific
Use an automated CEOI	x	x	Communications	Communications	MOS-Specific
Send a radio message	x	x	Communications	Communications	MOS-Specific
Identify armored vehicles		x	Visual Identification	Identify Targets	MOS-Specific
Install/fire M18 claymore		x	Mines/Traps	Basic Soldiering	MOS-Specific
Install/remove M21 mine	x	x	Mines/Traps	Basic Soldiering	MOS-Specific

^a Short task titles are given.

Table A-2

Tasks Tested: 13B

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>13B Cannon Crewmember</u>					
Evaluate a casualty		x	First Aid	Safety/Survival	General
Administer nerve agent antidote-self	x	x	First Aid	Safety/Survival	General
ID terrain features on map		x	Navigate	Basic Soldiering	General
Select movement route using map		x	Navigate	Basic Soldiering	General
Locate unknown point on map		x	Navigate	Basic Soldiering	General
Decontaminate your skin	x	x	Nuc/Bio/Chem	Safety/Survival	General
Recognize/react to chem/bio		x	Nuc/Bio/Chem	Safety/Survival	General
Use M256 chemical kit		x	Nuc/Bio/Chem	Safety/Survival	General
Maintain M16-series rifle		x	Weapons	Basic Soldiering	General
Engage targets w/M72A2 LAW		x	Weapons	Basic Soldiering	General
Headspace/timing on .50	x	x	Weapons	Basic Soldiering	General
Practice noise/light/litter		x	Field Techniques	Basic Soldiering	General
Select temp fighting position		x	Field Techniques	Basic Soldiering	General
React to indirect fire		x	Field Techniques	Basic Soldiering	General
Use automated CEOD		x	Communications	Communications	General
Report enemy information-SALUTE		x	Visual Identification	Identify Targets	General
Install/fire/recover M18A1	x	x	Mines/Traps	Basic Soldiering	General
Locate mines by probing		x	Mines/Traps	Basic Soldiering	General
Operate vehicle in a convoy		x	Drive Vehicles	Vehicles	MOS-Specific
Perform PMCS ^b		x	Maintain Vehicles	Vehicles	MOS-Specific
Perform prefire checks ^b	x	x	Operate/Maintain Howitzer	Technical	MOS-Specific
Prepare separate-loaded ammo ^b		x	Operate/Maintain Howitzer	Technical	MOS-Specific
Prepare howitzer for firing		x	Operate/Maintain Howitzer	Technical	MOS-Specific
Record firing data (DA Form-4513)	x	x	Operate/Maintain Howitzer	Technical	MOS-Specific
Determine howitzer safe-to-fire		x	Operate/Maintain Howitzer	Technical	MOS-Specific
Direct cannon crew during firing		x	Operate/Maintain Howitzer	Technical	MOS-Specific
Prepare range card ^b	x	x	Operate Sights	Technical	MOS-Specific
Establish aiming points ^b	x		Operate Sights	Technical	MOS-Specific
Determine site/range to crest		x	Operate Sights	Technical	MOS-Specific
Lay howitzer ^b		x	Operate Sights	Technical	MOS-Specific
Lay howitzer for initial direction		x	Operate Sights	Technical	MOS-Specific
Boresight DAP ^b	x	x	Operate Sights	Technical	MOS-Specific
Set/lay for deflection ^b	x	x	Operate Sights	Technical	MOS-Specific

^aShort task titles are given.^bTracked for M109, M110, and M198 howitzers.

Table A-3

Tasks Tested: 19K

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>19K Tank Crewman</u>					
Administer nerve agent antidote-self	x	x	First Aid	Safety/Survival	General
Put on a field or pressure dressing		x	First Aid	Safety/Survival	General
Evacuate wounded crewman		x	First Aid	Safety/Survival	General
Determine location on ground		x	Navigate	Basic Soldiering	General
Analyze terrain using five aspects		x	Navigate	Basic Soldiering	General
Use the latrine while in MOPP4		x	Nuc/Bio/Chem	Safety/Survival	General
Prepare NBC-1 reports	x	x	Nuc/Bio Chem	Safety/Survival	General
Prepare vehicle for nuclear		x	Nuc/Bio/Chem	Safety/Survival	General
Conduct unmasking procedures		x	Nuc/Bio/Chem	Safety/Survival	General
Maintain M240 coax	x	x	Weapons	Basic Soldiering	General
Maintain cal .50 M2 HB machinegun	x	x	Weapons	Basic Soldiering	General
Call for/adjust indirect fire		x	Field Techniques	Basic Soldiering	General
Establish tank firing position		x	Field Techniques	Basic Soldiering	General
Encode/decode using KTC 600		x	Communications	Communications	General
Use KTC 1400D system	x	x	Communications	Communications	General
Identify armored vehicles		x	Visual Identification	Identify Targets	General
Use visual signals	x	x	Visual Identification	Identify Targets	General
Recognize minefield markers		x	Mines/Traps	Basic Soldiering	General
Power-up gunner's station	x	x	Operate Tanks	Technical	MOS-Specific
Inspect and stow ammo		x	Operate Tanks	Technical	MOS-Specific
Recover a mired tank (M1 series)		x	Operate Tanks	Technical	MOS-Specific
Troubleshoot tank system		x	Operate Tanks	Technical	MOS-Specific
Perform computer self test	x	x	Tank Gunnery	Technical	MOS-Specific
Update MRS (M1A1)	x	x	Tank Gunnery	Technical	MOS-Specific
Boresight M1A1 tank		x	Tank Gunnery	Technical	MOS-Specific
Perform lead system check		x	Tank Gunnery	Technical	MOS-Specific
Engage target with main gun	x	x	Tank Gunnery	Technical	MOS-Specific
Conduct movement using wing man		x	Tank Gunnery	Technical	MOS-Specific

^aShort task titles are given.

Table A-4

Tasks Tested: 31C

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>31C Single Channel Radio Operator</u>					
Put on a field or pressure dressing	x		First Aid	Safety/Survival	General
Prevent shock	x		First Aid	Safety/Survival	General
Perform mouth-to-mouth resuscitation	x		First Aid	Safety/Survival	General
Determine grid coordinates	x		Navigate	Basic Soldiering	General
Determine location on ground	x		Navigate	Basic Soldier	General
Decontaminate your skin	x		Nuc/Bio/Chem	Safety/Survival	General
Put on/wear/remove M17 mask	x		Nuc/Bio/Chem	Safety/Survival	General
Recognize/react to chem/bio	x		Nuc/Bio/Chem	Safety/Survival	General
Maintain M17 protective mask	x		Nuc/Bio/Chem	Safety/Survival	General
Maintain an M16 rifle	x		Weapons	Basic Soldiering	General
Load/reduce/clear M16 rifle	x		Weapons	Basic Soldiering	General
Battlesight zero M16A1/M16A2 ^b	x		Weapons	Basic Soldiering	General
Camouflage equipment	x		Field Techniques	Basic Soldiering	General
Practice noise/light/litter discipline	x		Field Techniques	Basic Soldiering	General
Use an automated CEOI	x		Communications	Communications	General
Establish/enter/leave radio net	x		Communications	Communications	General
Visually identify threat aircraft	x		Visual Identification	Identify Targets	General
Drive/maintain vehicle	x		Drive Vehicles	Vehicles	General
Inspect operational generator	x		Generators	Technical	MOS-Specific
Troubleshoot PU-620 generator	x		Generators	Technical	MOS-Specific
Troubleshoot AN/GRC-106	x		Maintain/Operate TTY Equipment	Technical	MOS-Specific
Operate radio teletypewriter	x		Maintain/Operate TTY Equipment	Technical	MOS-Specific
Troubleshoot radio teletypewriter	x		Maintain/Operate TTY Equipment	Technical	MOS-Specific
Direct install doublet antenna	x		Install TTY Equipment	Technical	MOS-Specific
Select team radio site	x		Install TTY Equipment	Technical	MOS-Specific
Install radio set AN/GRC-106	x		Install TTY Equipment	Technical	MOS-Specific
Install radio teletypewriter	x		Install TTY Equipment	Technical	MOS-Specific
Prepare/maintain records/logs	x		Operations	Technical	MOS-Specific
Inventory radio equipment	x		Operations	Technical	MOS-Specific

^a Short task titles are given.

^b Tracked for M16A1 and M16A2 rifles.

Table A-5

Tasks Tested: 63B

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>63B Light Wheel Vehicle Mechanic</u>					
Administer nerve agent antidote-self	x	x	First Aid	Safety/Survival	General
Prevent shock		x	First Aid	Safety/Survival	General
Navigate on the ground	x	x	Navigate	Basic Soldiering	General
Plan route reconnaissance		x	Navigate	Basic Soldiering	General
Decontaminate your skin		x	Nuc/Bio/Chem	Safety/Survival	General
Put on/wear MOPP		x	Nuc/Bio/Chem	Safety/Survival	General
React to nuclear hazard		x	Nuc/Bio/Chem	Safety/Survival	General
Maintain M16A1/M16A2 rifle ^b	x	x	Weapons	Basic Soldiering	General
Perform maintenance on M60		x	Weapons	Basic Soldiering	General
Camouflage self and equipment	x	x	Field Techniques	Basic Soldiering	General
Use automated CEOI		x	Communications	Communications	General
Report enemy information-SALUTE		x	Visual Identification	Identify Targets	General
Prepare DA Form 2404	x	x	Maintain Vehicles	Vehicles	MOS-Specific
Perform annual PMCS	x	x	Maintain Vehicles	Vehicles	MOS-Specific
Replace hydraulic master cylinder		x	Brake/Suspension	Technical	MOS-Specific
Troubleshoot service brake		x	Brake/Suspension	Technical	MOS-Specific
Troubleshoot air system		x	Brake/Suspension	Technical	MOS-Specific
Troubleshoot air-hydraulic brake		x	Brake/Suspension	Technical	MOS-Specific
Inspect/replace suspension		x	Brake/Suspension	Technical	MOS-Specific
Troubleshoot charging system		x	Power Train	Technical	MOS-Specific
Troubleshoot engine	x	x	Power Train	Technical	MOS-Specific
Troubleshoot fuel system malfunctions	x	x	Fuel/Coolant	Technical	MOS-Specific
Troubleshoot liquid cooling system		x	Fuel/Coolant	Technical	MOS-Specific
Recon terrain/route to recovery		x	Vehicle Recovery	Technical	MOS-Specific
Recover disabled vehicles		x	Vehicle Recovery	Technical	MOS-Specific
Inventory tools/equipment		x	Motor Pool Operations	Technical	MOS-Specific
Use oxygen acetylene torch		x	Motor Pool Operations	Technical	MOS-Specific

^a Short task titles are given.^b Hands-on test tracked for M16A1 and M16A2 rifles.

Table A-6

Tasks Tested: 71L

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>71L Administrative Specialist</u>					
Evaluate a casualty		x	First Aid	Safety/Survival	General
Prevent shock	x	x	First Aid	Safety/Survival	General
Perform mouth-to-mouth resuscitation		x	First Aid	Safety/Survival	General
Determine grid coordinates	x	x	Navigate	Basic Soldiering	General
Identify terrain features		x	Navigate	Basic Soldiering	General
Decontaminate your skin		x	Nuc/Bio/Chem	Safety/Survival	General
Put on/wear/remove M17 mask	x	x	Nuc/Bio/Chem	Safety/Survival	General
Put on/wear MOOPP		x	Nuc/Bio/Chem	Safety/Survival	General
Recognize/react to chem/bio		x	Nuc/Bio/Chem	Safety/Survival	General
Maintain an M16A1/M16A2 rifle ^b	x	x	Weapons	Basic Soldiering	General
Load/reduce/clear M16 rifle		x	Weapons	Basic Soldiering	General
Battlesight zero M16A1/M16A2 ^c		x	Weapons	Basic Soldiering	General
Camouflage self and equipment	x	x	Field Techniques	Basic Soldiering	General
Practice noise/light/litter discipline		x	Field Techniques	Basic Soldiering	General
Use challenge and password		x	Field Techniques	Basic Soldiering	General
Send a radio message	x	x	Communications	Communications	General
Operate FM radio set	x	x	Communications	Communications	General
Report enemy information-SALUTE		x	Visual Identification	Identify Targets	General
Identify armored vehicles		x	Visual Identification	Identify Targets	General
Request resupply of pubs/forms	x	x	Forms/Files Management	Technical	MOS-Specific
File documents and correspondence	x	x	Forms/Files Management	Technical	MOS-Specific
File using MARKS system		x	Forms/Files Management	Technical	MOS-Specific
Assemble correspondence	x	x	Correspondence	Technical	MOS-Specific
Type a memorandum		x	Correspondence	Technical	MOS-Specific
Proofread/edit correspondence/reports		x	Correspondence	Technical	MOS-Specific
Type straight copy	x		Correspondence	Technical	MOS-Specific
Type endorsement to memorandum	x	x	Correspondence	Technical	MOS-Specific
Rec/Trans classified material	x	x	Classified Materials	Technical	MOS-Specific
Inventory classified material	x	x	Classified Materials	Technical	MOS-Specific
Receive/control office equipment		x	Supervision/Coordination	Technical	MOS-Specific
Control supplies		x	Supervision/Coordination	Technical	MOS-Specific

^a Short task titles are given.^b Hands-on test tracked for M16A1 and M16A2 rifles.^c Tracked for M16A1 and M16A2 rifles.

Table A-7

Tasks Tested: 88M

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>88M Motor Transport Operator</u>					
Administer nerve agent antidote-self	x	x	First Aid	Safety/Survival	General
Prevent shock		x	First Aid	Safety/Survival	General
Perform mouth-to-mouth resuscitation		x	First Aid	Safety/Survival	General
Determine grid coordinates	x	x	Navigate	Basic Soldiering	General
Identify terrain features	x	x	Navigate	Basic Soldiering	General
Determine location on ground		x	Navigate	Basic Soldiering	General
Analyze terrain using five mil aspects		x	Navigate	Basic Soldiering	General
Decontaminate your skin	x	x	Nuc/Bio/Chem	Safety/Survival	General
Mark NBC contaminated area		x	Nuc/Bio/Chem	Safety/Survival	General
Recognize/react to chem/bio		x	Nuc/Bio/Chem	Safety/Survival	General
Decontaminate equipment w/ABC M11		x	Nuc/Bio/Chem	Safety/Survival	General
Cross a contaminated area in truck		x	Nuc/Bio/Chem	Safety/Survival	General
Maintain an M16A1/M16A2 rifle ^b	x	x	Weapons	Basic Soldiering	General
Perform maintenance on M60		x	Weapons	Basic Soldiering	General
Make water safe for drinking		x	Field Techniques	Basic Soldiering	General
Camouflage equipment		x	Field Techniques	Basic Soldiering	General
Move under direct fire		x	Field Techniques	Basic Soldiering	General
Camouflage defensive position		x	Field Techniques	Basic Soldiering	General
Use proper ambushed defense		x	Field Techniques	Basic Soldiering	General
Send a radio message	x	x	Communications	Communications	General
Identify armored vehicles		x	Visual Identification	Identify Targets	General
Neutralize booby traps		x	Mines/Traps	Basic Soldiering	General
Install/fire M18 claymore		x	Mines/Traps	Basic Soldiering	General
Transport general cargo	x	x	Drive Vehicles	Vehicles	MOS-Specific
Operate truck/semitrailer	x	x	Drive Vehicles	Vehicles	MOS-Specific
Operate vehicle in convoy		x	Drive Vehicles	Vehicles	MOS-Specific
Drive vehicle in convoy		x	Drive Vehicles	Vehicles	MOS-Specific
Perform PMCS (M915/M916/M931A2)	x	x	Maintain Vehicles	Vehicles	MOS-Specific
Process vehicle commitment order	x	x	Dispatch Vehicles	Technical	MOS-Specific
Perform vehicle emergency procedures		x	Recover Vehicles	Technical	MOS-Specific

^aShort task titles are given.

^bHands-on test tracked for M16A1 and M16A2 rifles.

Table A-8

Tasks Tested: 91A

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
<u>91A Medical Specialist</u>					
Evaluate a casualty		x	First Aid	Safety/Survival	General
Prevent shock		x	First Aid	Safety/Survival	General
Triage	x	x	First Aid	Safety/Survival	General
Navigate on the ground	x	x	Navigate	Basic Soldiering	General
Put on/wear MOPP	x	x	Nuc/Bio/Chem	Safety/Survival	General
Supervise fitting of M17 mask		x	Nuc/Bio/Chem	Safety/Survival	General
Replace filters on M17 mask		x	Nuc/Bio/Chem	Safety/Survival	General
Maintain an M16A1/M16A2 rifle ^b	x	x	Weapons	Basic Soldiering	General
Load/reduce/clear M16 rifle		x	Weapons	Basic Soldiering	General
Camouflage self and equipment		x	Field Techniques	Basic Soldiering	General
Move under direct fire		x	Field Techniques	Basic Soldiering	General
Select and mark evacuation		x	Field Techniques	Basic Soldiering	General
Pitch and strike tents		x	Field Techniques	Basic Soldiering	General
Request MEDEVAC	x	x	Communications	Communications	General
Use automated CEOI	x	x	Communications	Communications	General
Report enemy information-SALUTE	x	x	Visual Identification	Identify Targets	General
Perform PMCS (M998/M1010)	x	x	Maintain Vehicles	Vehicles	General
Initiate field medical card	x	x	Clinic/Ward Treatment	Technical	MOS-Specific
Initiate IV	x	x	Clinic/Ward Treatment	Technical	MOS-Specific
Administer an injection	x	x	Clinic/Ward Treatment	Technical	MOS-Specific
Initiate treatment for shock		x	Clinic/Ward Treatment	Technical	MOS-Specific
Establish an ET tube airway		x	Clinic/Ward Treatment	Technical	MOS-Specific
Apply MAST	x	x	Clinic/Ward Treatment	Technical	MOS-Specific
Treat a suspected spine injury		x	Clinic/Ward Treatment	Technical	MOS-Specific
Treat impalement	x	x	Clinic/Ward Treatment	Technical	MOS-Specific
Immobilize a dislocated hip		x	Clinic/Ward Treatment	Technical	MOS-Specific
Carry out rescue/evacuation		x	Clinic/Ward Treatment	Technical	MOS-Specific
Attend to casualties		x	Clinic/Ward Treatment	Technical	MOS-Specific
Request/control medical supplies		x	Clinic/Ward Management	Technical	MOS-Specific
Maintain medical kits		x	Clinic/Ward Management	Technical	MOS-Specific

^aShort task titles are given.^bHands-on test tracked for M16A1 and M16A2 rifles.

Table A-9

Tasks Tested: 95B

Task ^a	HO	JK	Functional Category	Task Factor	Task Construct
95B Military Police					
Evaluate a casualty		x	First Aid	Safety/Survival	General
Navigate on ground	x	x	Navigate	Basic Soldiering	General
Determine grid coordinates		x	Navigate	Basic Soldiering	General
Conduct hasty route reconnaissance		x	Navigate	Basic Soldiering	General
Decontaminate your skin	x	x	Nuc/Bio/Chem	Safety/Survival	General
Recognize/react to chem/bio		x	Nuc/Bio/Chem	Safety/Survival	General
Prepare NBC-1 reports		x	Nuc/Bio/Chem	Safety/Survival	General
Engage target with M16		x	Weapons	Basic Soldiering	General
Perform maintenance on M60	x	x	Weapons	Basic Soldiering	General
Camouflage self and equipment		x	Field Techniques	Basic Soldiering	General
Call/adjust indirect fire		x	Field Techniques	Basic Soldiering	General
Conduct defense by squad		x	Field Techniques	Basic Soldiering	General
Move around obstacles	x	x	Field Techniques	Basic Soldiering	General
Direct fire/maneuver		x	Field Techniques	Basic Soldiering	General
Use automated CEOI		x	Communications	Communications	General
Report enemy information-SALUTE		x	Visual Identification	Identify Targets	General
Locate mines by probing	x	x	Mines/Traps	Basic Soldiering	General
Perform PMCS (M998)	x	x	Maintain Vehicles	Vehicles	General
Collect/process evidence	x	x	Patrol Duties	Technical	MOS-Specific
Perform patrol duties		x	Patrol Duties	Technical	MOS-Specific
Prepare MP reports and forms	x	x	Patrol Duties	Technical	MOS-Specific
Enforce traffic regulations	x	x	Patrol Duties	Technical	MOS-Specific
Advise Miranda		x	MP Procedures	Technical	MOS-Specific
Decide when to use force		x	MP Procedures	Technical	MOS-Specific
Control restricted area	x	x	Security	Technical	MOS-Specific
Plan/supervise security operation		x	Security	Technical	MOS-Specific
Perform EPW/CI activities		x	Security	Technical	MOS-Specific
Prepare operations overlay		x	Operations	Technical	MOS-Specific
Operate a CCP		x	Operations	Technical	MOS-Specific

^aShort task titles are given.

Appendix B

Task, Functional Category, Task Factor, and Task Construct Scores
Descriptive Statistics by MOS (LVII)

Table B-1

Descriptive Statistics: Hands-On Tests: 11B

Level	N	Mean (Percent GO)	SD
Task Level			
Put on a field or pressure dressing	340	89.29	15.47
Navigate on the ground	336	74.00	18.99
Decontaminate your skin	340	84.02	17.85
Engage target w/M72A2 LAW	340	45.59	16.79
Operate AN/PVS-4	339	85.84	14.39
Perform movement MOUT	337	88.97	13.96
Use an automated CEOI	332	63.38	32.89
Send a radio message	337	88.16	16.73
Install/remove M21 mine	330	94.07	12.15
<i>Across Tasks</i>	<i>341</i>	<i>79.23</i>	<i>7.72</i>
Functional Category Level			
First Aid	340	89.29	15.47
Navigate	336	74.00	18.99
Nuc/Bio/Chem	340	84.02	17.85
Weapons	341	65.65	11.87
Field Techniques	337	88.97	13.96
Communications	338	75.93	19.65
Mines/Traps	330	94.07	12.15
Task Factor Level			
Safety/Survival	341	86.69	12.43
Basic Soldiering	341	77.57	7.92
Communications	338	75.93	19.65
Task Construct Level			
MOS-Specific	341	79.23	7.72

Note. Tasks are standardized by test site.

Table B-2

Descriptive Statistics: Hands-On Tests: 13B

Level	N	Mean (Percent GO)	SD
Task Level			
Administer nerve agent antidote-self	173	79.20	22.13
Decontaminate your skin	173	77.39	20.77
Headspace/timing on .50	172	79.16	24.91
Install/fire/recover M18A1	154	88.79	17.26
Perform prefire checks ^a	172	71.34	30.61
Record firing data (DA Form-4513)	173	50.98	20.29
Prepare range card ^a	172	27.19	29.94
Establish aiming points ^a	172	80.48	32.07
Determine site/range to crest	172	76.25	30.65
Lay howitzer ^a	168	79.30	29.23
Boresight DAP ^a	170	62.41	36.52
Set/lay for deflection ^a	173	82.38	28.94
<i>Across Tasks</i>	174	70.99	16.27
Functional Category Level			
First Aid	173	79.20	22.132
Nuc/Bio/Chem	173	77.39	20.77
Weapons	172	79.16	24.91
Mines/Traps	154	88.79	17.26
Operate/Maintain Howitzer	174	61.05	19.37
Operate Sights	173	76.12	25.07
Task Factor Level			
Safety/Survival	173	78.30	16.11
Basic Soldiering	172	83.47	17.67
Technical	174	66.14	20.36
Task Construct Level			
General	173	80.81	13.64
MOS-Specific	174	66.15	20.36

Note. Tasks are standardized by test site and track.

^aTracked for M109, M110, and M198 howitzers.

Table B-3

Descriptive Statistics: Hands-On Tests: 19K

Level	N	Mean (Percent GO)	SD
Task Level			
Administer nerve agent antidote-self	166	82.03	18.43
Prepare NBC-1 reports	166	44.21	20.70
Maintain M240 coax	164	96.80	8.96
Maintain cal .50 M2 HB machinegun	164	92.90	13.80
Use KTC 1400D system	166	42.61	25.29
Use visual signals	165	39.55	27.21
Power-up gunner's station	160	92.81	12.29
Perform computer self test	160	78.45	19.99
Update MRS (M1A1)	160	82.33	26.89
Engage target with main gun	160	77.84	18.96
<i>Across Tasks</i>	<i>166</i>	<i>72.43</i>	<i>9.11</i>
Functional Category Level			
First Aid	166	82.03	18.43
Nuc/Bio/Chem	166	44.21	20.70
Weapons	164	94.85	9.29
Communications	166	42.61	25.29
Visual Identification	165	39.55	27.21
Operate Tanks	160	92.81	12.29
Tank Gunnery	160	79.55	15.33
Task Factor Level			
Safety/Survival	166	63.12	14.19
Basic Soldiering	164	94.85	9.29
Communications	166	42.61	25.29
Identify Targets	165	39.55	27.21
Technical	160	82.86	11.89
Task Construct Level			
General	166	66.15	9.06
MOS-Specific	160	82.86	11.89

Note. Tasks are standardized by test site.

Table B-4

Descriptive Statistics: Hands-On Tests: 63B

Level	N	Mean (Percent GO)	SD
Task Level			
Administer nerve agent antidote-self	187	83.34	18.56
Navigate on the ground	187	62.74	18.58
Maintain M16A1/M16A2 rifle*	187	83.89	12.35
Camouflage self and equipment	183	73.05	15.33
Prepare DA Form 2404	171	62.33	17.68
Perform annual PMCS	171	77.09	20.13
Troubleshoot engine	171	78.59	30.95
Troubleshoot fuel system malfunctions	177	84.36	17.41
<i>Across Tasks</i>	187	75.67	8.31
Functional Category Level			
First Aid	187	83.34	18.56
Navigate	187	62.74	18.58
Weapons	187	83.89	12.35
Field Techniques	183	73.05	15.33
Maintain Vehicles	171	69.71	15.20
Power Train	171	78.59	30.95
Fuel/Coolant	177	84.36	17.41
Task Factor Level			
Safety/Survival	187	83.34	18.56
Basic Soldiering	187	73.17	10.13
Vehicles	171	69.71	15.20
Technical	184	81.31	18.77
Task Construct Level			
General	187	75.75	8.76
MOS-Specific	185	75.79	13.05

Note. Tasks are standardized by test site and track.

*Tracked by rifle type.

Table B-5

Descriptive Statistics: Hands-On Tests: 71L

Level	N	Mean (Percent GO)	SD
Task Level			
Prevent shock	156	69.67	25.27
Determine grid coordinates	155	73.05	25.19
Put on/wear/remove M17 mask	152	81.13	22.79
Maintain an M16A1/M16A2 rifle*	156	72.11	18.40
Camouflage self and equipment	156	70.05	16.80
Send a radio message	154	69.48	36.67
Operate FM radio set	153	45.26	40.05
Request resupply of pubs/forms	156	53.63	19.55
File documents and correspondence	152	52.16	25.85
Assemble correspondence	155	27.58	27.20
Type straight copy	156	53.56	16.44
Type endorsement to memorandum	156	55.95	17.58
Rec/tran classified material	151	41.17	23.20
Inventory classified documents	153	56.17	19.61
<i>Across Tasks</i>	<i>156</i>	<i>58.66</i>	<i>9.36</i>
Functional Category Level			
First Aid	156	69.67	25.27
Navigate	155	73.05	25.19
Nuc/Bio/Chem	152	81.13	22.79
Weapons	156	72.11	18.40
Field Techniques	156	70.05	16.80
Communications	154	57.39	29.33
Forms/Files Management	156	52.94	16.88
Correspondence	156	45.76	13.51
Classified Materials	156	49.17	17.05
Task Factor Level			
Safety/Survival	156	75.19	19.21
Basic Soldiering	156	71.72	13.93
Communications	154	57.39	29.33
Technical	156	48.63	10.37
Task Construct Level			
General	156	68.64	13.61
MOS-Specific	156	48.63	10.37

Note. Tasks are standardized by test site and track.

*Tracked by rifle type.

Table B-6

Descriptive Statistics: Hands-On Tests: 88M

Level	N	Mean (Percent GO)	SD
Task Level			
Administer nerve agent antidote-self	88	74.13	18.83
Determine grid coordinates	88	64.97	30.75
Identify terrain features	88	66.68	23.85
Decontaminate your skin	87	78.31	21.42
Maintain an M16A1/M16A2 rifle*	88	87.58	12.33
Send a radio message	88	75.60	19.33
Transport general cargo	85	43.68	31.87
Operate truck/semitrailer	86	58.13	35.73
Perform PCMS (M915/M916/M931A2)	86	77.59	12.99
Process vehicle commitment order	88	73.34	12.13
<i>Across Tasks</i>	<i>88</i>	<i>70.19</i>	<i>9.14</i>
Functional Category Level			
First Aid	88	74.13	18.83
Navigate	88	65.82	21.52
Nuc/Bio/Chem	87	78.31	21.42
Weapons	88	87.58	12.33
Communications	88	75.60	19.33
Drive Vehicles	86	50.99	25.02
Maintain Vehicles	86	77.59	12.99
Dispatch Vehicles	88	73.34	12.13
Task Factor Level			
Safety/Survival	88	76.03	15.22
Basic Soldiering	88	73.07	15.33
Communications	88	75.60	19.33
Vehicles	88	60.53	17.05
Technical	88	73.34	12.13
Task Construct Level			
General	88	74.50	11.80
MOS-Specific	88	63.63	13.20

Note. Tasks are standardized by test site and track.

*Tracked by rifle type.

Table B-7

Descriptive Statistics: Hands-On Tests: 91A

Level	N	Mean (Percent GO)	SD
Task Level			
Triage	214	58.87	26.28
Navigate on the ground	212	63.93	17.18
Put on/wear MOPP	210	84.87	16.43
Maintain an M16A1/M16A2 rifle*	211	78.38	15.33
Request MEDEVAC	210	25.87	29.53
Use automated CEOI	205	34.43	32.77
Report enemy information-SALUTE	214	81.63	21.45
Perform PMCS (M998/M1010)	212	64.71	17.83
Initiate field medical card	214	70.14	16.14
Initiate IV	211	89.14	15.69
Administer an injection	210	90.49	13.29
Apply MAST	208	81.53	17.77
Treat impalement	213	53.78	25.31
<i>Across Tasks</i>	<i>215</i>	<i>67.62</i>	<i>10.50</i>
Functional Category Level			
First Aid	214	58.87	26.27
Navigate	212	63.93	17.18
Nuc/Bio/Chem	210	84.87	16.43
Weapons	211	78.38	15.33
Communications	212	30.25	27.15
Visual Identification	214	81.63	21.45
Maintain Vehicles	212	64.71	17.83
Clinic/Ward Treatment	214	76.83	10.19
Task Factor Level			
Safety/Survival	214	71.71	16.71
Basic Soldiering	214	71.10	12.75
Communications	212	30.25	27.15
Identify Targets	214	81.63	21.45
Vehicles	212	64.71	17.83
Technical	214	76.83	10.19
Task Construct Level			
General	215	61.82	13.15
MOS-Specific	214	76.83	10.19

Note. Tasks are standardized by test site and track.

*Tracked by rifle type.

Table B-8

Descriptive Statistics: Hands-On Tests: 95B

Level	N	Mean (Percent GO)	SD
Task Level			
Navigate on the ground	166	67.45	19.93
Decontaminate your skin	166	75.40	16.28
Perform maintenance on M60	163	79.23	24.21
Move around obstacles	162	75.09	20.36
Locate mines by probing	165	65.23	20.83
Perform PMCS (M998)	160	65.93	15.27
Collect/process evidence	166	67.68	15.34
Prepare MP reports and forms	166	82.91	11.98
Enforce traffic regulations	165	73.90	15.36
Control restricted area	166	72.24	19.15
<i>Across Tasks</i>	<i>168</i>	<i>72.57</i>	<i>8.30</i>
Functional Category Level			
Navigate	166	67.45	19.93
Nuc/Bio/Chem	166	75.40	16.28
Weapons	163	79.23	24.21
Field Techniques	162	75.09	20.36
Mines/Traps	165	65.23	20.83
Maintain Vehicles	160	65.93	15.27
Patrol Duties	168	74.85	10.34
Security	166	72.24	19.15
Task Factor Level			
Safety/Survival	166	75.40	16.28
Basic Soldiering	166	71.70	13.15
Vehicles	160	65.93	15.27
Technical	168	74.26	9.15
Task Construct Level			
General	167	71.47	11.19
MOS-Specific	168	74.26	9.15

Note. Tasks are standardized by test site.

Table B-9

Descriptive Statistics: Job Knowledge Tests: 11B

Level	N	Mean (Percent Correct)	SD
Task Level			
Evaluate a casualty	345	87.57	25.61
Put on a field or pressure dressing	345	90.24	19.14
Practice preventive medicine	345	54.65	23.73
Navigate on the ground	345	72.99	26.11
Determine grid coordinates	345	85.33	25.18
Orient map by terrain assoc	345	92.03	20.21
Decontaminate your skin	345	78.99	22.97
Check soldiers in MOPP4	345	56.73	28.97
Conduct unmasking procedures	345	44.35	27.89
Engage target w/M16	345	27.86	20.02
Zero M249 machinegun	345	43.50	23.19
Engage target w/M72A2 LAW	345	52.65	21.11
Engage target w/M60	345	73.85	22.96
Engage target w/.50	345	54.20	26.82
Prepare M47 for firing	345	83.01	19.25
Operate AN/PVS-4	345	75.31	23.28
Zero AN/PVS-4	345	71.01	26.48
Call/adjust indirect fire	345	63.30	23.75
Select overwatch position	345	59.65	17.79
React to ambush	345	88.40	17.77
Conduct defense by squad	345	79.20	21.83
Perform movement MOUT	345	78.25	17.33
Control fire team	345	75.07	24.15
Control organic fires	345	63.81	30.37
Use an automated CEOI	345	54.71	21.40
Send a radio message	345	38.40	25.20
Identify armored vehicles	345	64.83	20.40
Install/fire M18 claymore	345	51.45	20.85
Install/remove M21 mine	345	55.39	14.34
<i>Across Tasks</i>	345	64.90	8.34

(table continues)

Table B-9 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	345	73.83	14.31
Navigate	345	81.33	18.27
Nuc/Bio/Chem	345	58.45	18.16
Weapons	345	53.88	9.33
Field Techniques	345	61.09	10.69
Communications	345	83.80	15.07
Visual Identification	345	78.25	17.33
Mines/Traps	345	70.25	21.52
Task Factor Level			
Safety/Survival	345	65.74	12.61
Basic Soldering	345	61.18	9.13
Communications	345	83.80	15.07
Identify Targets	345	78.25	17.33
Task Construct Level			
MOS-Specific	345	64.90	8.34

Table B-10

Descriptive Statistics: Job Knowledge Tests: 13B

Level	N	Mean (Percent Correct)	SD
Task Level			
Evaluate a casualty	179	84.35	28.16
Administer nerve agent antidote-self	179	84.41	23.65
ID terrain features on map	179	76.19	22.70
Select movement route using map	179	47.64	25.13
Locate unknown point on map	179	53.25	32.67
Decontaminate your skin	179	79.97	27.25
Recognize/react to chem/bio	179	79.88	24.83
Use M256 chemical kit	179	55.87	28.59
Maintain M16-series rifle	179	80.15	19.30
Engage targets w/M72A2 LAW	179	47.67	19.97
Headspace/timing on .50	179	61.49	28.27
Practice noise/light/litter	179	83.43	23.80
Select temp fighting position	179	63.22	20.67
React to indirect fire	179	55.87	31.47
Use automated CEOI	179	77.97	25.59
Report enemy information-SALUTE	179	90.25	17.90
Install/fire/recover M18A1	179	61.63	25.85
Locate mines by probing	179	44.93	30.43
Operate vehicle in a convoy	179	45.93	21.87
Perform PMCS ^a	179	76.82	25.00
Perform prefire chks ^a	179	65.43	27.01
Prepare separate-loaded ammo ^a	179	68.89	22.95
Prepare howitzer for firing	179	63.61	20.78
Record firing data (DA-4513)	179	40.92	32.06
Determine howitzer safe-to-fire	179	79.10	22.51
Direct cannon crew during firing	179	70.46	25.97
Prepare range card ^a	179	63.85	23.53
Lay howitzer for initial direction	179	49.44	24.00
Boresight DAP	179	52.15	33.60
Set/lay for deflection	179	55.62	27.57
<i>Across Tasks</i>	179	65.19	10.90

(table continues)

Table B-10 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	179	84.39	19.41
Navigate	179	59.55	18.04
Nuc/Bio/Chem	179	70.29	18.95
Weapons	179	65.80	14.69
Field Techniques	179	67.39	14.43
Communications	179	77.97	25.59
Visual Identification	179	90.25	17.89
Mines/Traps	179	54.47	21.96
Drive Vehicles	179	45.93	21.87
Maintain Vehicles	179	65.43	27.01
Operate/Maintain Howitzer	179	59.75	16.19
Operate Sights	179	71.75	17.26
Task Factor Level			
Safety/Survival	179	74.99	16.30
Basic Soldiering	179	62.62	10.92
Communications	179	77.97	25.59
Identify Targets	179	90.25	17.90
Vehicles	179	55.68	18.49
Technical	179	63.47	15.09
Task Construct Level			
General	179	68.87	10.71
MOS-Specific	179	60.67	14.25

Note. Tasks are standardized by track.

*Tracked for M109, M110, and M198 howitzers.

Table B-11

Descriptive Statistics: Job Knowledge Tests: 19K

Level	N	Mean (Percent Correct)	SD
Task Level			
Administer nerve agent antidote-self	168	87.10	24.72
Put on a field or pressure dressing	168	86.30	21.99
Evacuate wounded crewman	168	64.48	20.10
Determine location on ground	168	67.85	29.39
Analyze terrain using five aspects	168	54.17	24.57
Use the latrine while in MOPP4	168	38.87	23.76
Prepare NBC-1 reports	168	48.60	22.15
Prepare vehicle for nuclear	168	69.30	21.64
Conduct unmasking procedures	168	51.49	29.30
Maintain M240 coax	168	66.42	20.92
Maintain cal .50 M2 HB machinegun	168	64.05	22.13
Call for/adjust indirect fire	168	50.47	27.50
Establish tank firing position	168	63.28	29.03
Encode/decode using KTC 600	168	22.65	21.02
Use KTC 1400D system	168	55.17	33.76
Identify armored vehicles	168	91.10	10.65
Use visual signals	168	40.39	22.45
Recognize minefield markers	168	34.72	18.09
Power-up gunner's station	168	65.89	20.35
Inspect and stow ammo	168	51.25	21.37
Recover a mired tank (M1 series)	168	39.52	20.05
Troubleshoot tank system	168	74.40	20.83
Perform computer self test	168	63.69	23.19
Update MRS (M1A1)	168	46.45	18.11
Boresight M1A1 tank	168	21.92	17.07
Perform lead system check	168	44.07	20.16
Engage target with main gun	168	56.41	26.41
Conduct movement using wing man	168	43.00	23.83
<i>Across Tasks</i>	168	56.89	8.85

(table continues)

Table B-11 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	168	75.59	16.70
Navigate	168	60.03	20.77
Nuc/Bio/Chem	168	53.73	16.19
Weapons	168	65.23	16.75
Field Techniques	168	56.17	22.03
Communications	168	40.72	20.56
Visual Identification	168	76.19	11.16
Mines/Traps	168	34.72	18.09
Operate Tanks	168	52.92	11.57
Tank Gunnery	168	48.77	11.85
Task Factor Level			
Safety/Survival	168	61.68	13.70
Basic Soldiering	168	56.37	12.45
Communications	168	40.72	20.56
Identify Targets	168	76.19	11.16
Technical	168	50.29	9.74
Task Construct Level			
General	168	60.37	10.27
MOS-Specific	168	50.29	9.75

Table B-12

Descriptive Statistics: Job Knowledge Tests: 31C

Level	N	Mean (Percent Correct)	SD
Task Level			
Put on a field or pressure dressing	70	83.82	21.92
Prevent shock	70	57.35	22.87
Perform mouth-to-mouth resuscitation	70	83.82	17.15
Determine grid coordinates	70	81.13	28.05
Determine location on ground	70	68.25	30.59
Decontaminate your skin	70	72.55	22.99
Put on/wear/remove M17 mask	70	73.71	19.42
Recognize/react to chem/bio	70	75.98	23.63
Maintain M17 protective mask	70	86.47	16.73
Maintain an M16A1 rifle	70	88.53	14.37
Load/reduce/clear M16 rifle	70	80.51	21.53
Battlesight zero M16A1/M16A2 ^a	70	49.05	18.35
Camouflage equipment	70	68.13	22.63
Practice noise/light/litter discipline	70	87.25	21.57
Use an automated CEOI	70	84.80	20.30
Establish/enter/leave radio net	70	90.44	16.17
Visually identify threat aircraft	70	69.11	22.07
Drive/maintain vehicle	70	78.92	22.97
Inspect operational generator	70	43.75	25.69
Troubleshoot PU-620 generator	70	67.77	26.85
Troubleshoot AN/GRC-106	70	62.13	24.62
Operate radio teletypewriter	70	56.62	25.04
Troubleshoot radio teletypewriter	70	54.41	22.41
Direct install doublet antenna	70	60.59	24.43
Select team radio site	70	87.43	20.90
Install radio set AN/GRC-106	70	31.34	22.89
Install radio teletypewriter	70	69.94	21.33
Prepare/maintain records/logs	70	50.88	23.79
Inventory radio equipment	70	53.53	27.52
<i>Across Tasks</i>	70	68.35	7.79

(table continues)

Table B-12 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	70	74.19	12.72
Navigate	70	74.69	24.98
Nuc/Bio/Chem	70	78.19	12.10
Weapons	70	68.07	11.74
Field Techniques	70	77.69	14.57
Communications	70	79.77	13.66
Visual Identification	70	43.75	25.69
Drive Vehicles	70	84.80	20.30
Generators	70	57.84	18.37
Maintain/Operate TTY Equipment	70	46.24	17.79
Install TTY Equipment	70	73.75	13.00
Operations	70	64.55	19.45
Task Factor Level			
Safety/Survival	70	76.49	9.30
Basic Soldiering	70	71.43	11.13
Communications	70	79.77	13.66
Identify Targets	70	43.75	25.69
Vehicles	70	84.80	20.30
Technical	70	60.87	11.19
Task Construct Level			
General	70	73.25	8.19
MOS-Specific	70	60.87	11.19

Note. Tasks are standardized by track.

*Tracked by rifle type.

Table B-13

Descriptive Statistics: Job Knowledge Tests: 63B

Level	N	Mean (Percent Correct)	SD
Task Level			
Administer nerve agent antidote-self	192	88.37	19.83
Prevent shock	192	52.21	25.03
Navigate on the ground	192	56.70	27.75
Plan route reconnaissance	192	72.27	25.77
Decontaminate your skin	192	78.17	26.66
Put on/wear MOPP	192	56.33	25.66
React to nuclear hazard	192	86.45	15.42
Maintain M16A1/M16A2 rifle	192	62.89	33.81
Perform maintenance on M60	192	44.90	16.71
Camouflage self and equipment	192	43.83	28.39
Use automated CEOI	192	66.84	28.63
Report enemy information-SALUTE	192	69.53	38.38
Prepare DA Form 2404	192	89.08	18.47
Perform annual PMCS	192	48.26	28.07
Replace hydraulic master cylinder	192	88.93	17.79
Troubleshoot service brake	192	38.71	26.85
Troubleshoot air system	192	51.90	32.30
Troubleshoot air-hydraulic brake	192	41.35	23.51
Inspect/replace suspension	192	80.59	21.65
Troubleshoot charging system	192	63.02	27.99
Troubleshoot engine	192	73.17	31.07
Troubleshoot fuel system malfunctions	192	56.11	25.87
Troubleshoot liquid cooling system	192	80.47	23.77
Recon terrain/route to recovery	192	45.36	26.40
Recover disabled vehicles	192	79.61	25.15
Inventory tools/equipment	192	78.34	22.19
Use oxygen acetylene torch	192	66.38	20.82
<i>Across Tasks</i>	192	65.01	8.73

(table continues)

Table B-13 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	192	67.70	17.20
Navigate	192	51.19	22.51
Nuc/Bio/Chem	192	68.92	17.85
Weapons	192	77.61	17.81
Field Techniques	192	44.90	16.71
Communications	192	69.53	28.38
Visual Identification	192	89.08	18.47
Maintain Vehicles	192	59.19	16.75
Brake/Suspension	192	62.40	12.94
Power Train	192	80.04	18.55
Fuel/Coolant	192	71.69	16.38
Vehicle Recovery	192	53.41	21.38
Motor Pool Operations	192	50.09	21.94
Task Factor Level			
Safety/Survival	192	68.39	12.89
Basic Soldiering	192	60.19	13.44
Communications	192	69.53	28.38
Identify Targets	192	89.08	18.47
Vehicles	192	59.19	16.75
Technical	192	63.37	10.75
Task Construct Level			
General	192	67.62	10.03
MOS-Specific	192	62.74	9.95

Table B-14

Descriptive Statistics: Job Knowledge Tests: 71L

Level	N	Mean (Percent Correct)	SD
Task Level			
Evaluate a casualty	155	83.55	25.55
Prevent shock	155	57.90	21.65
Perform mouth-to-mouth resuscitation	155	78.87	20.57
Determine grid coordinates	155	68.81	33.26
Identify terrain features	155	73.75	25.37
Decontaminate your skin	155	63.44	25.13
Put on/wear/remove M17 mask	155	57.58	24.57
Put on/wear MOPP	155	65.16	29.75
Recognize/react to chem/bio	155	74.67	24.27
Maintain an M16A1/M16A2 rifle	155	81.67	19.86
Load/reduce/clear M16 rifle	155	58.92	29.15
Battlesight zero M16A1/M16A2*	155	38.05	13.69
Camouflage self and equipment	155	41.20	20.80
Practice noise/light/litter discipline	155	77.85	25.84
Use challenge and password	155	76.13	24.83
Send a radio message	155	63.70	20.97
Operate FM radio set	155	76.97	21.51
Report enemy information-SALUTE	155	80.75	25.77
Identify armored vehicles	155	45.49	17.09
Request resupply of pubs/forms	155	78.27	22.97
File documents and correspondence	155	74.87	22.87
File using MARKS system	155	70.97	23.91
Assemble correspondence	155	52.09	33.11
Type a memorandum	155	72.74	20.01
Proofread/edit correspondence/reports	155	68.97	22.01
Type endorsement to memorandum	155	55.52	27.83
Rec/Trans classified material	155	65.97	25.39
Inventory classified documents	155	72.69	24.17
Receive/control office equipment	155	59.84	27.53
Control supplies	155	63.75	22.63
<i>Across Tasks</i>	155	64.26	8.09

(table continues)

Table B-14 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	155	71.41	12.97
Navigate	155	71.63	22.22
Nuc/Bio/Chem	155	64.63	15.77
Weapons	155	55.79	12.83
Field Techniques	155	60.72	14.32
Communications	155	70.34	15.78
Visual Identification	155	57.25	14.30
Forms/Files Management	155	74.42	17.11
Correspondence	155	63.71	15.17
Classified Materials	155	68.20	21.05
Supervision/Coordination	155	62.07	17.99
Task Factor Level			
Safety/Survival	155	67.58	11.54
Basic Soldiering	155	60.51	10.91
Communications	155	70.34	15.78
Identify Targets	155	57.25	14.30
Technical	155	67.21	12.16
Task Construct Level			
General	155	62.68	8.84
MOS-Specific	155	67.21	12.16

Note. Tasks are standardized by track.

*Tracked by rifle type.

Table B-15

Descriptive Statistics: Job Knowledge Tests: 88M

Level	N	Mean (Percent Correct)	SD
Task Level			
Administer nerve agent antidote-self	89	89.89	19.72
Prevent shock	89	51.97	25.90
Perform mouth-to-mouth resuscitation	89	77.81	21.47
Determine grid coordinates	89	61.80	36.77
Identify terrain features	89	70.51	26.80
Determine location on ground	89	55.81	32.41
Analyze terrain using five mil aspects	89	46.35	23.40
Decontaminate your skin	89	73.78	21.01
Mark NBC contaminated area	89	36.90	23.06
Recognize/react to chem/bio	89	73.41	29.38
Decontaminate equipment w/ABC M11	89	55.43	28.40
Cross a contaminated area in truck	89	43.82	25.92
Maintain an M16A1/M16A2 rifle	89	84.57	18.36
Perform maintenance on M60	89	58.05	31.59
Make water safe for drinking	89	43.33	25.31
Camouflage equipment	89	54.68	22.95
Move under direct fire	89	44.10	29.68
Camouflage defensive position	89	28.09	22.97
Use proper ambushed defense	89	82.25	19.22
Send a radio message	89	64.33	21.61
Identify armored vehicles	89	54.69	20.19
Neutralize booby traps	89	24.09	22.95
Install/fire M18 claymore	89	50.59	23.80
Transport general cargo	89	67.64	22.86
Operate truck/semitrailer	89	67.74	19.93
Operate vehicle in convoy	89	59.13	27.08
Drive vehicle in convoy	89	35.46	26.78
Perform PCMS (M915/M916/M931A2)	89	83.15	19.13
Process vehicle commitment order	89	39.04	23.82
Perform vehicle emergency procedures	89	42.32	32.48
<i>Across Tasks</i>	<i>89</i>	<i>57.99</i>	<i>8.81</i>

(table continues)

Table B-15 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	89	71.71	15.83
Navigate	89	58.59	19.86
Nuc/Bio/Chem	89	55.87	13.20
Weapons	89	74.63	18.55
Field Techniques	89	53.12	13.24
Communications	89	64.33	21.61
Visual Identification	89	54.69	20.19
Mines/Traps	89	39.23	17.14
Drive Vehicles	89	58.09	13.95
Maintain Vehicles	89	83.15	19.13
Dispatch Vehicles	89	39.04	23.82
Recover Vehicles	89	42.32	32.48
Task Factor Level			
Safety/Survival	89	62.32	11.20
Basic Soldiering	89	56.27	11.85
Communications	89	64.33	21.61
Identify Targets	89	54.69	20.19
Vehicles	89	62.86	12.18
Technical	89	40.45	19.29
Task Construct Level			
General	89	58.21	9.55
MOS-Specific	89	57.26	10.39

Table B-16

Descriptive Statistics: Job Knowledge Tests: 91A

Level	N	Mean (Percent Correct)	SD
Task Level			
Evaluate a casualty	220	95.89	15.34
Prevent shock	220	50.46	18.54
Triage	220	74.66	30.89
Navigate on the ground	220	61.53	30.47
Put on/wear MOPP	220	81.77	24.48
Supervise fitting of mask (M17)	220	79.35	24.81
Replace filters on M17 mask	220	75.30	25.80
Maintain an M16A1/M16A2 rifle	220	85.05	17.75
Load/reduce/clear M16 rifle	220	75.91	25.89
Camouflage self and equipment	220	47.18	23.74
Move under direct fire	220	50.74	31.47
Select and mark evacuation	220	53.20	24.68
Pitch and strike tents	220	37.36	24.22
Request MEDEVAC	220	48.77	28.08
Use automated CEOI	220	77.88	24.63
Report enemy information-SALUTE	220	88.49	19.21
Perform PMCS (M998/M1010)	220	52.14	29.50
Initiate field medical card	220	71.39	19.66
Initiate IV	220	81.11	20.79
Administer an injection	220	66.76	27.79
Initiate treatment for shock	220	45.36	26.18
Establish an ET tube airway	220	40.24	33.64
Apply MAST	220	57.31	26.41
Treat a suspected spine injury	220	48.09	27.10
Treat impalement	220	64.73	26.84
Immobilize a dislocated hip	220	79.68	27.27
Carry out rescue/evacuation	220	74.91	22.53
Attend to casualties	220	62.25	32.34
Request/control medical supplies	220	45.35	25.73
Maintain medical kits	220	90.30	18.48
<i>Across Tasks</i>	220	65.45	10.59

(table continues)

Table B-16 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	220	69.22	16.11
Navigate	220	61.53	30.47
Nuc/Bio/Chem	220	78.46	17.19
Weapons	220	80.99	16.27
Field Techniques	220	47.12	16.43
Communications	220	63.33	21.44
Visual Identification	220	88.49	29.21
Maintain Vehicles	220	52.14	29.50
Clinic/Ward Treatment	220	64.34	11.16
Clinic/Ward Management	220	64.61	17.44
Task Factor Level			
Safety/Survival	220	73.84	13.25
Basic Soldiering	220	59.20	14.52
Communications	220	63.33	21.44
Identify Targets	220	88.49	19.21
Vehicles	220	52.14	29.50
Technical	220	64.38	10.62
Task Construct Level			
General	220	66.16	12.40
MOS-Specific	220	64.38	10.62

Table B-17

Descriptive Statistics: Job Knowledge Tests: 95B

Level	N	Mean (Percent Correct)	SD
Task Level			
Evaluate a casualty	168	90.18	23.38
Navigate on the ground	168	68.94	29.86
Determine grid coordinates	168	80.51	28.91
Conduct hasty route reconnaissance	168	47.62	29.76
Decontaminate your skin	168	79.23	21.77
Recognize/react to chem/bio	168	78.82	24.70
Prepare NBC-1 reports	168	44.34	26.62
Engage target with M16	168	28.16	22.26
Perform maintenance on M60	168	75.45	27.95
Camouflage self and equipment	168	46.19	23.47
Call/adjust indirect fire	168	50.36	24.76
Conduct defense by squad	168	51.44	20.27
Move around obstacles	168	77.83	28.05
Direct fire/maneuver	168	65.33	27.15
Use automated CEOI	168	79.80	24.15
Report enemy information-SALUTE	168	94.35	14.69
Locate mines by probing	168	64.88	28.37
Perform PMCS (M998)	168	39.19	22.70
Collect/process evidence	168	82.92	20.31
Perform patrol duties	168	82.74	21.88
Prepare MP reports & forms	168	86.16	17.47
Enforce traffic regulations	168	74.60	22.26
Advise Miranda	168	88.10	18.34
Decide when to use force	168	85.91	21.73
Control restricted area	168	66.27	26.30
Plan/supervise security operation	168	41.15	20.83
Perform EPW/CI activities	168	55.78	21.75
Prepare operations overlay	168	42.86	28.97
Operate a CCP	168	72.32	32.26
<i>Across Tasks</i>	168	64.87	09.30

(table continues)

Table B-17 (continued)

Level	N	Mean (Percent Correct)	SD
Functional Category Level			
First Aid	168	90.18	23.38
Navigate	168	66.01	21.08
Nuc/Bio/Chem	168	63.30	17.40
Weapons	168	48.27	18.03
Field Techniques	168	55.26	13.82
Communications	168	79.80	24.15
Visual Identification	168	94.35	14.69
Mines/Traps	168	65.33	27.15
Maintain Vehicles	168	39.19	22.70
Patrol Duties	168	82.03	12.79
MP Procedures	168	87.00	15.74
Security	168	52.81	14.55
Operations	168	52.68	24.22
Task Factor Level			
Safety/Survival	168	67.44	15.39
Basic Soldiering	168	57.35	11.98
Communications	168	79.80	24.15
Identify Targets	168	94.35	15.69
Vehicles	168	39.19	22.70
Technical	168	68.14	10.04
Task Construct Level			
General	168	62.98	10.70
MOS-Specific	168	68.14	10.04

Appendix C

Army-Wide and MOS-Specific Rating Scale Contents

Army-Wide Rating Dimensions

Section I: Army-Wide Performance Categories

1. Technical Knowledge/Skill
2. Effort
3. Supervising
4. Following Regulations and Orders
5. Integrity
6. Training/Developing
7. Maintaining Assigned Equipment
8. Physical Fitness
9. Self-Development
10. Consideration for Subordinates
11. Military Appearance/Bearing
12. Self-Control

Section II: Supervisor Performance Requirements

1. Acting as a Role Model for Subordinates
2. Communication
3. Personal Counseling
4. Monitoring Subordinate Performance
5. Organizing Missions/Operations
6. Personnel Administration
7. Performance Counseling/Correcting

Section III: Overall Effectiveness

Section IV: Senior NCO Potential

MOS-Specific Rating Dimensions

11B: Infantryman

1. Maintaining and Accounting for Equipment and Weapons
2. Supervising Soldiers in the Field
3. Leading the Team
4. Navigation
5. Use of Organic Weapons and Equipment
6. Personal Safety, Field Sanitation, and Personal Hygiene
7. Fighting Positions
8. Avoiding Enemy Detection
9. Operating a Radio Set
10. Reconnaissance
11. Guard and Security Duties
12. Prisoners of War
13. Proficiency in Battle
14. Overall MOS Performance

13B: Cannon Crewmember

1. Loading Out Equipment
2. Driving and Maintaining Vehicles, Howitzers, and Equipment
3. Transporting/Sorting/Storing and Preparing Ammunition for Fire
4. Preparing for Occupation/Emplacing Howitzer
5. Setting Up Communications
6. Gunnery
7. Loading/Unloading Howitzer
8. Receiving and Relaying Communications
9. Recording/Record Keeping
10. Position Improvement
11. Assuming Supervisory Duties in Absence of the Section Chief
12. Overall MOS Performance

19K: Tank Crewman

1. Maintaining Tank, Tank Systems, and Associated Equipment
2. Driving/Recovering Tanks
3. Stowing Ammunition Aboard Tanks
4. Loading/Unloading Weapons
5. Maintaining Weapons
6. Engaging Targets with Tank Weapon Systems
7. Operating Communications Equipment
8. Preparing Tanks for Field Problems
9. Assuming Supervisory Duties in Absence of the Tank Commander
10. Overall MOS Performance

31C: Single Channel Radio Operator

1. Inspecting and Servicing Equipment
2. Installing Equipment
3. Operating Communications Devices
4. Preparing Reports
5. Maintaining Security
6. Preparing for Movement
7. Providing Safe Transportation
8. Managing the RATT Rig
9. Overall MOS Performance

63B: Light Wheel Vehicle Mechanic

1. Inspecting and Testing Equipment Problems
2. Checking Repairs Made by Other Mechanics
3. Troubleshooting
4. Performing Preventive Maintenance Checks and Services
5. Repair
6. Using/Accounting for Tools and Test Equipment
7. Using Technical References
8. Equipment Operation
9. Safety Mindedness
10. Administrative Duties
11. Determining Task Requirements
12. Recovery
13. Overall MOS Performance

71L: Administrative Specialist

1. Preparing, Typing, and Proofreading Documents
2. Processing and Distributing Documents
3. Maintaining Office Resources
4. Establishing and/or Maintaining File IAW MARKS
5. Correspondence Management
6. Preparing and Safeguarding Classified Materials
7. Providing Customer Service
8. Overall MOS Performance

88M: Motor Transport Operator

1. Driving Vehicles
2. Vehicle Coupling
3. Checking and Maintaining Vehicles
4. Using Maps/Following Proper Routes
5. Loading and Transporting Cargo
6. Loading and Transporting Personnel
7. Parking and Securing Vehicles
8. Performing Administrative Duties
9. Self-Recovering Vehicles
10. Safety-Mindedness
11. Performing Dispatcher Duties
12. Overall MOS Performance

91A/B: Medical Specialist

1. Maintaining and Operating Army Medical Vehicles and Equipment
2. Maintaining Accountability of Medical Supplies and Equipment
3. Keeping Medical Records
4. Arranging for Transport and/or Transporting Injured Personnel
5. Dispensing Medications
6. Preparing/Maintaining Field Site or Clinic Facilities in the Field
7. Providing Routine and Ongoing Patient Care
8. Responding to Emergency Situations
9. Providing Health Care & Health Maintenance Instruction to Army Personnel
10. Overall MOS Performance

95B: Military Police

1. Traffic Control and Enforcement
2. Providing Security
3. Investigating Crimes and Making Apprehensions
4. Patrolling
5. Leading the Team in a Tactical Environment
6. Promoting the Public Image of the Military Police
7. Dealing with Difficult Interpersonal Situations
8. Responding to Medical Emergencies
9. Navigation
10. Avoiding Enemy Detection
11. Use of Weapons and Other Equipment
12. Fighting Positions
13. Battlefield Circulation Control
14. Enemy Prisoners of War
15. Overall MOS Performance