Technical Report 1314

Validating Future Force Performance Measures (Army Class): In-Unit Performance Longitudinal Validation

Deirdre J. Knapp (Editor)
Human Resources Research Organization

Kimberly S. Owens (Editor)
U.S. Army Research Institute

Matthew T. Allen (Editor)
Human Resources Research Organization

September 2012

United States Army Research Institute for the Behavioral and Social Sciences

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

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

Peter M. Greenston, U.S. Army Research Institute
J. Douglas Dressel, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: DAPE-ARI-ZXM, 6000 6th Street (Bldg. 1464 / Mail Stop 5610), Ft. Belvoir, VA 22060-5610.

FINAL DISPOSITION: Destroy this Technical Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

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

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
September 2012	Final	December 30, 2008 – May 31, 2011

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Validating Future Force Performance Measures (Army Class): In-Unit Performance Longitudinal Validation	W91WAW-09-D-0013
	5b. PROGRAM ELEMENT NUMBER 622785
6. EDITOR(S) Deirdre J. Knapp (Human Resources Research Organization); Kimberly S. Owens (Amry Research Institute); and Matthew T. Allen (Human Resources Research Organization)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 06
	5e. WORK UNIT NUMBER 329
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Suite 700 Alexandria, Virginia 22314	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6TH Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1314

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Dr. Kimberly Owens, Contracting Officer's Representative

14. ABSTRACT *(Maximum 200 words)*: The Army needs the best personnel to meet the emerging demands of the 21st century. Accordingly, it is seeking recommendations on new predictor measures, in particular, measures of non-cognitive attributes (e.g., interests, values, temperament) that could enhance entry-level Soldier selection and classification decisions. The U. S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducted a longitudinal criterion-related validation research effort to collect data to inform these recommendations.

Data on experimental predictors were collected from about 11,000 Soldiers and criterion data were collected from these Solders at three career points—end of training, after about 12-24 months in-service, and again about a year later. This report describes the two "in-unit" criterion data collections and criterion-related validation analysis results. Overall, many of the experimental predictors significantly incremented the Armed Forces Qualification Test (AFQT) in predicting Soldier performance. They also incremented the ability to predict retention-related attitudes and behavior over Education Tier. In addition, the predictors showed potential for improving job assignment decisions above and beyond the Armed Services Vocational Aptitude Battery. As expected, due to the attrition of poor performing Soldiers which reduces variance on the outcome measures, the strength of the results is somewhat diminished using in-unit data in comparison to training criterion data.

15. SUBJECT TERMS

Personnel, Criterion-related validation, Selection and classification, Manpower

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	135	Dorothy Young (703) 545-2316

Standard Form 298

Technical Report 1314

Validating Future Force Performance Measures (Army Class): In-Unit Performance Longitudinal Validation

Deirdre J. Knapp (Editor)
Human Resources Research Organization

Kimberly S. Owens (Editor)
U.S. Army Research Institute

Matthew T. Allen (Editor)
Human Resources Research Organization

Personnel Assessment Research Unit
Tonia S. Heffner, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Bldg. 1464
Fort Belvoir, Virginia 22060

September 2012

Army Project Number
622785A790

Personnel, Performance and Training Technology

ACKNOWLEDGEMENTS

There are a large number of individuals not listed as authors who have contributed significantly to the work described in this report. In-unit performance criterion data collection support (either on-site proctoring or Help Desk support) was provided by a number of individuals from both ARI and HumRRO, including those listed below:

ARI: Elizabeth Brady, Rich Hoffman, Sharon Ardison, and Doug Dressel.

HumRRO: Mary Adeniyi, Ashley Armstrong, Doug Brown, Wade Buckland, Roy Campbell, Kristina Handy, Milt Koger, Joy Oliver, Cathy Stawarski, Matthew Trippe, Shonna Waters, and Elise Weaver.

Ms. Sharon Meyers (ARI) prepared the criterion measures for computer-based administration. Dr. Dan Putka (HumRRO) provided statistical consultation and advice.

We are, of course, also indebted to the military and civilian personnel who supported our test development and data collection efforts, particularly those Soldiers and noncommissioned officers (NCOs) who participated in the research. Of particular note are members of the Army Test Program Advisory Team (ATPAT) who provided advice and input throughout the course of this research. Members of the ATPAT are listed below.

MAJ PAUL WALTON
CSM BRIAN A. HAMM
CSM JAMES SHULTZ
SGM KENAN HARRINGTON
SGM THOMAS KLINGEL
SGM(R) CLIFFORD MCMILLAN
SGM GREGORY A. RICHARDSON
SFC WILLIAM HAYES
SFC KENNETH WILLIAMS
MR. ROBERT STEEN

VALIDATING FUTURE FORCE PERFORMANCE MEASURES (ARMY CLASS): IN-UNIT PERFORMANCE LONGITUDINAL VALIDATION

EXECUTIVE SUMMARY

Research Requirement:

The Army needs the best personnel to meet the emerging demands of the 21st century. Selecting and classifying these Soldiers requires new predictor measures that assess attributes not currently covered by the existing Armed Forces Qualification Test (AFQT), in particular measures of non-cognitive attributes (e.g., interests, values, and temperament). One of the objectives of the "Army Class" research program is to provide the Army with recommendations on which new experimental predictor measures evidence the greatest potential to enhance new Soldier selection and classification. The present report documents the in-unit performance stages of a longitudinal criterion-related validation research effort conducted to advance this objective.

Procedure:

Predictor data were collected from about 11,000 entry-level enlisted Soldiers representing all Components (Regular Army, U.S. Army Reserve, U.S. Army National Guard). Soldiers were drawn from two samples: (a) job-specific samples targeting six entry-level Military Occupational Specialties (MOS) and (b) an Army-wide sample with no MOS-specific requirements. The experimental predictor instruments were administered to new Soldiers as they entered the Army through one of four reception battalions. The predictor measures included (a) three temperament measures (Assessment of Individual Motivation [AIM], Tailored Adaptive Personality Assessment System [TAPAS], and Rational Biodata Inventory [RBI]), (b) a predictor situational judgment test (PSJT), and (c) two measures of person-environment (P-E) fit (Work Preferences Assessment [WPA] and Army Knowledge Assessment [AKA]). In addition, we obtained scores through administrative records on the Assembling Objects (AO) test, a spatial ability measure currently administered with the Armed Services Vocational Aptitude Battery (ASVAB). Two predictor measures (AIM and TAPAS) were included in the research to support a short-term requirement to identify predictors that could immediately be put into operational use by the Army (i.e., the *Expanded Enlistment Eligibility Metrics* [EEEM] initiative).

In 2008, training performance criterion measures were administered to Soldiers in six job-specific longitudinal validation samples. These measures included (a) MOS-specific and Warrior Tasks and Battle Drills (WTBD) job knowledge tests (JKTs), (b) MOS-specific and Army-wide performance ratings collected from training instructors and peers, and (c) a questionnaire (the Army Life Questionnaire [ALQ]) measuring Soldiers' experiences and attitudes towards the Army through Initial Military Training.

Next, in 2009, we collected in-unit job performance data from Soldiers in the original predictor sample, regardless of MOS, most of whom had been in the Army for 12-24 months. The criterion measures paralleled those administered at the end of training and included JKTs (including an WTBD JKT suitable for all Soldiers regardless of MOS), performance ratings, and an in-unit variation of the ALQ. We collected WTBD JKT and supervisor ratings data for all Soldiers and MOS-specific JKT and ratings data from Soldiers in the six target MOS. For all Regular Army

Soldiers, we obtained data on attrition on a quarterly basis. In 2010-2011, we conducted another in-unit data collection, this time when Soldiers would have been in the Army on average about 3 years. The same criterion measures were administered in both data collections.

This report describes the in-unit 1 and in-unit 2 data collections and analyses. Three sets of analyses were conducted. The first analyses estimated the incremental validity of the experimental predictors over AFQT scores, across multiple performance criteria. The second set of analyses examined the ability of the measures to predict various retention-related criteria. The final set of analyses looked at the potential of the experimental predictors for use in making MOS classification decisions.

Findings:

With respect to predicting in-unit Soldier performance, we found the following:

Multiple experimental measures predicted can-do (i.e., technical) in-unit criteria beyond the AFQT. As expected, AFQT predicted can-do aspects of performance (e.g., job knowledge test scores) quite well. Even so, there was evidence of incremental validity (i.e., ability to predict outcomes beyond what can be predicted with AFQT alone), particularly for the AO and PSJT. Though small in magnitude, these measures demonstrated incremental validity for most of the criteria assessed with in-unit 1 sample (the PSJT did not predict ratings of Soldiers' performance of MOS-specific tasks). All of the corrected incremental validity estimates save one (the PSJT prediction of MOS-specific job knowledge) were near zero in the in-unit 2 sample. Thus, consistent with theoretical and empirical findings, the AFQT remains the strongest predictor of can-do performance throughout a Soldier's first term of service.

Multiple experimental measures predicted will-do (i.e., non-technical) in-unit criteria, over and above the AFQT, and more strongly than they predicted can-do criteria. The AFQT demonstrated less potential to predict will-do aspects of performance (e.g., effort and discipline). Among the experimental measures, the RBI showed the most promise in predicting will-do criteria (with the exception of ratings of Soldiers' ability to work effectively with others). The RBI, TAPAS, and AIM had consistently higher incremental validity coefficients than the other measures for predicting will-do criteria (e.g., supervisor ratings of effort and discipline). The WPA, AO, and PSJT also predicted some will-do criteria over the AFQT but not as strongly as the three temperament measures. The pattern of results was similar across the in-unit 1 and in-unit 2 samples, but the estimates were weaker in the in-unit 2 sample.

Multiple experimental measures predicted deployment adjustment beyond the AFQT, but did not predict deployment performance. The AIM, RBI, and AKA predicted deployment adjustment beyond AFQT in the in-unit 1 sample, with the AIM demonstrating the largest increment in validity over the AFQT. The PSJT showed greater potential to predict deployment adjustment in the in-unit 2 sample compared to the other predictor measures; however, the RBI and AKA continued to demonstrate small incremental validity over the AFQT. No measure, including the AFQT, predicted ratings of combat/deployment performance, suggesting that this dimension may be difficult to assess outside of the operational (i.e., combat/deployment) context.

With respect to predicting Soldier attrition and retention intentions, we found the following:

Multiple experimental measures predicted Soldier attrition beyond Education Tier. Overall, Education Tier predicted attrition at a modest rate. Beyond Education Tier, the RBI and AIM emerged as the best predictors of attrition (in general), followed by the TAPAS and the WPA. In predicting attrition for specific reasons (i.e., moral character, performance, and medical), three experimental measures—AIM, TAPAS, and RBI—had the strongest rates of prediction. A similar pattern of results emerged for modeling attrition longitudinally. However, when using statistics that take into account the number of scales included in the model, AO provided the best fit to the data, followed by the AIM and RBI. The TAPAS also emerged as a strong predictor of moral character and performance attrition.

Multiple experimental measures showed incremental variance in Soldier retention and career intentions beyond Education Tier. Education tier was generally ineffective for predicting retention and career intentions. The experimental measures, however, showed considerable promise in predicting these outcomes. Specifically, affective commitment to the Army was predicted quite well by the RBI, WPA, and AKA in both in-unit samples. The career intentions scale was predicted by the RBI, WPA, and AKA, as well as the AIM and TAPAS. Perceived fit with the Army was predicted by all experimental measures except AO. There were minor differences across the two in-unit samples, with the most notable being the lower magnitude of the in-unit 2estimates.

Because of sample size limitations, we only evaluated the classification potential of the experimental predictors using the in-unit 1 data. We found the following:

In general, the experimental predictors exhibited non-trivial classification gains over the ASVAB for the six target MOS. This held true for both MOS-specific performance-related criteria, such as job knowledge and ratings of technical performance, and MOS-specific retention-related criteria, such as self-reported MOS fit and MOS satisfaction. Across both sets of criteria, the TAPAS, RBI, and WPA exhibited the greatest classification gains over the ASVAB for the target MOS. That being said, no single measure exhibited the greatest classification potential across the MOS (i.e., the best measure for an MOS varied by MOS).

The classification gains associated with the experimental predictor measures were somewhat higher, on average, for an expanded sample of MOS than the target MOS. Although the cross-sample differences in classification gains were generally small, these findings illustrate the point that findings of classification potential can change depending on the specific MOS included in the analysis. They also suggest that the experimental predictor measures have classification potential beyond the six target MOS. Also, the pattern of findings by predictor measure was generally the same between the expanded and target MOS samples, with the TAPAS, RBI, and WPA showing the greatest classification gains over ASVAB.

Utilization and Dissemination of Findings:

These findings provide useful information to Army personnel managers and researchers about the potential of experimental predictor measures of non-cognitive attributes to supplement the ASVAB in selecting and classifying new Soldiers.

VALIDATING FUTURE FORCE PERFORMANCE MEAUSRES (ARMY CLASS): IN-UNIT PERFORMANCE LONGITUDINAL VALIDATION

CONTENTS

Page

CHAPTER 1: INTRODUCTION 1
Deirdre J. Knapp (HumRRO) and Tonia S. Heffner (ARI)
Background 1
Overview of the Army Class Research Program 2
Overview of Report 3

CHAPTER 2: LONGITUDINAL RESEARCH DESIGN 4
Deirdre J. Knapp (HumRRO), Tonia S. Heffner, and Kimberly S. Owens (ARI)
Data Collection Points and Sample 4
Criterion Measures 5
Overview 5
In-Unit Criterion Measure Descriptions 5
Predictor Measures 7
Overview 7
Description of Predictors 7

CHAPTER 3: IN-UNIT DATA COLLECTION 12
Karen O. Moriarty, Charlotte H. Campbell (HumRRO), and Kimberly S. Owens (ARI)
Overview 12
Staff Training 12
General Procedure 12
Regular Army Data Collection 13
Reserve Component Data Collection 13

CHAPTER 4: DATABASE DEVELOPMENT 15
Karen O. Moriarty, Matthew Trippe, and Laura Ford (HumRRO)
Database Construction 15
Data Cleaning 15
Scoring the Assessments 15
Attrition Database 16
Master Longitudinal Database 16
Sample Description 16

CONTENTS (CONTINUED)

Page

CHAPTER 5: MEASURE SCORING AND PSYCHOMETRIC PROPERTIES 22
Matthew T. Allen, Tina Chang, and Michael J. Ingerick (HumRRO)

Predictor Measure Scores and Associated Psychometric Properties 22
Armed Services Vocational Aptitude Battery (ASVAB) and Education Tier 22
Assessment of Individual Motivation (AIM) 22
Tailored Adaptive Personality Assessment System (TAPAS-95s) 22
Rational Biodata Inventory (RBI) 22
Predictor Situational Judgment Test (PSJT) 23
Army Knowledge Assessment (AKA) 23
Work Preferences Assessment (WPA) 24
In-Unit Criterion Measure Scores and Associated Psychometric Properties 24
In-Unit Job Knowledge Tests (JKTs) 24
In-Unit Performance Rating Scales (PRS) 25
In-Unit Army Life Questionnaire (ALQ) 27
Attrition 28

CHAPTER 6: PREDICTING IN-UNIT SOLDIER PERFORMANCE 29
Joseph P. Caramagno, Matthew T. Allen, and Michael J. Ingerick (HumRRO)

Background 29
Incremental Validity Analysis 29
Approach 29
Findings 32
Summary 42

CHAPTER 7: PREDICTING IN-UNIT SOLDIER ATTRITION AND CONTINUANCE INTENTIONS OVER TIME 44
Matthew T. Allen (HumRRO)

Background 44
Predicting Cumulative Soldier Attrition 46
Approach 46
Findings 50
Predicting Soldier Attrition Over Time 53
Approach 53
Findings 55
Predicting Soldier Continuance 58
Approach 58
Findings 59
Summary 61

CONTENTS (CONTINUED)

Page

CHAPTER 8: EVALUATING CLASSIFICATION POTENTIAL 63
Matthew Trippe, Michael Ingerick, and Ted Diaz (HumRRO)
Overview and Background 63
Approach to Estimating the Classification Potential of the Experimental Predictors 63
Results 65
Maximizing Performance-Related Criteria 65
Maximizing Retention-Related Criteria 68
Classification Potential among an Expanded Sample of MOS 71
Conclusions 76
CHAPTER 9: SUMMARY AND CONCLUSIONS 79
Matthew T. Allen, Deirdre J. Knapp, (HumRRO), and Kimberly S. Owens (ARI)
Summary of Main Findings 79
Predicting In-Unit Soldier Performance 79
Predicting Attrition and Retention Intentions 80
Evaluating Classification Potential 80
Limitations and Issues 81
Comparing Results to Previous Army Class Findings 81
Generalizability of Findings to an Operational Setting 82
Future Research 82
REFERENCES 83
APPENDIX A: DEVELOPMENT OF THE COMBAT/DEPLOYMENT PERFORMANCE RATING SCALES A-1
APPENDIX B: DESCRIPTIVE STATISTICS AND SCORE INTERCORRELATIONS FOR SELECTED PREDICTOR MEASURES B-1
APPENDIX C: SCALE-LEVEL CORRELATIONS BETWEEN SELECTED PREDICTOR AND IN-UNIT CRITERION MEASURES C-1

Page

LIST OF TABLES

Table 2.1. Example In-Unit Performance Rating Scales 6
Table 2.2. Example In-Unit ALQ Scales 6
Table 2.3. Summary of Longitudinal Validation Predictor Measures 8

Table 3.1 In-Unit Proctored Data Collection Site Visits 14

Table 4.1. In-Unit 1 Criterion Sample by MOS and Demographic Subgroup 17
Table 4.2. In-Unit 2 Criterion Sample by MOS and Demographic Subgroup 18
Table 4.3. In-Unit 1 and In-Unit 2 Criterion Sample by Component and MOS 19
Table 4.4. Demographic Characteristics for the Predictor, Training, and In-Unit 1, and In-Unit 2 Samples 20
Table 4.5. Attrition Analysis Sample Demographics 21

Table 5.1. Descriptive Statistics and Reliability Estimates for In-Unit 1 and In-Unit 2 Job Knowledge Tests 25
Table 5.2. Descriptive Statistics and Reliability Estimates for Composite Performance Rating Scales 26
Table 5.3. Descriptive Statistics for In-Unit 1 and In-Unit 2 Performance Rating Scales (PRS) 26
Table 5.4. Descriptive Statistics and Reliability Estimates for In-Unit 1 and In-Unit 2 Army Life Questionnaire (ALQ) Scale Scores 28

Table 6.1. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 1 Can-Do Performance 33
Table 6.2. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 2 Can-Do Performance 35
Table 6.3. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 1 Will-Do Performance 37
Table 6.4. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 2 Will-Do Performance 39
Table 6.5. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit Deployment Adjustment and Performance 41
Table 6.6. Summary of Incremental Validity Estimates for Experimental Predictors over the AFQT by Criterion Domain and Months of Service 43

Page

Table 7.1. Cumulative Attrition Rates over Time by Education Tier....46
Table 7.2. Treatment of Select Interservice Separation Codes (ISC) for Different Types of Attrition Analyses....48
Table 7.3. Type of 36-Month Cumulative Attrition by Education Tier....49
Table 7.4. Incremental Validity for Experimental Predictors over Education Tier for Predicting Cumulative Attrition through 36 Months of Service....50
Table 7.5. Incremental Validity of Experimental Predictors over Education Tier for Type of Cumulative Attrition through 36 Months of Service....52
Table 7.6. Event History Analysis Assessing the Goodness-of-Fit of Nested Experimental Predictor Models Through 42 Months of Service....56
Table 7.7. Event History Analysis Assessing the Goodness-of-Fit of Non-Nested Experimental Predictor Models Through 42 Months of Service....58
Table 7.8. Incremental Validity Estimates for Experimental Predictors over the Education Tier for Predicting In-Unit 1 Retention-Related Criteria....60
Table 7.9. Incremental Validity Estimates for Experimental Predictors over the Education Tier for Predicting In-Unit 2 Retention-Related Criteria....62

Table 8.1. Criterion Measures Used in Classification Potential Analyses....64
Table 8.2. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted MOS-Specific Job Knowledge Across and Within MOS....66
Table 8.3. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted MOS-Specific Performance Ratings Across and Within MOS....67
Table 8.4. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted Retention-Related Outcomes Averaged Across and Within MOS....69
Table 8.5. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted IMT Performance Outcomes Averaged Across and Within MOS....70
Table 8.6. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted IMT Performance Outcomes (Number of Times Restarted during IMT) Averaged Across and Within MOS (Expanded Sample)....72

Page

Table 8.7. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted In-Unit Perceived MOS Fit (ALQ) Averaged Across and Within MOS (Expanded Sample) 73
Table 8.8. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted In-Unit MOS Satisfaction (ALQ) Averaged Across and Within MOS (Expanded Sample) 74
Table 8.9. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted MOS-Specific Performance Ratings Averaged Across and Within MOS (Expanded Sample) 75
Table 8.10. Summary of the Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted Training Outcomes (Target MOS) 77
Table 8.11. Summary of the Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted In-Unit Outcomes (Target MOS) 78

Table A.1. Results of the AW and Source Document Dimension Sort A-2
Table A.2. Summary of In-House Critical Incident Retranslation Results A-4
Table A.3. Mapping of AW and Combat /Deployment Rating Scales A-4
Table A.4. Combat/Deployment Dimension Definitions Based on Critical Incident Workshops A-5
Table A.5. Percentage of Anchors Categorized as Intended A-7

Table B.1. Descriptive Statistics for Education Tier, Armed Services Vocational Aptitude Battery (ASVAB) Subtests, and Armed Forces Qualification Test (AFQT) B-1
Table B.2. Intercorrelations among Education Tier, ASVAB Subtest, and AFQT Scores B-1
Table B.3. Descriptive Statistics and Reliability Estimates for Assessment of Individual Motivation (AIM) Scales B-2
Table B.4. Intercorrelations among AIM Scales B-2
Table B.5. Descriptive Statistics for Tailored Adaptive Personality Assessment System (TAPAS-95s) Scales B-3
Table B.6. Intercorrelations among TAPAS-95s Scales B-3
Table B.7. Descriptive Statistics and Reliability Estimates for Rational Biodata Inventory (RBI) Scale Scores B-4
Table B.8. Intercorrelations among RBI Scale Scores B-5

CONTENTS (CONTINUED)

Page

Table B.9. Descriptive Statistics and Reliability Estimates for Army Knowledge Assessment (AKA) Scales B-5
Table B.10. Intercorrelations among AKA Scales B-6
Table B.11. Descriptive Statistics and Reliability Estimates for Work Preferences Assessment (WPA) Dimension and Facet Scores B-6
Table B.12. Intercorrelations among WPA Dimension and Facet Scores B-7

Table C.1. Correlations between all Experimental Predictors and Select In-Unit 1 Performance-Related Can-Do Criteria C-1
Table C.2. Correlations between all Experimental Predictors and Select In-Unit 1 Performance-Related Will-Do Criteria C-3
Table C.3. Correlations between all Experimental Predictors and Select In-Unit 2 Performance-Related Can-Do Criteria C-5
Table C.4. Correlations between all Experimental Predictors and Select In-Unit 2 Performance-Related Will-Do Criteria C-7
Table C.5. Correlations between all Experimental Predictors and In-Unit Combat Performance and Deployment Adjustment Criteria C-9

VALIDATING FUTURE FORCE PERFORMANCE MEASURES (ARMY CLASS): IN-UNIT LONGITUDINAL VALIDATION

CHAPTER 1: INTRODUCTION

Deirdre J. Knapp (HumRRO) and Tonia S. Heffner (ARI)

Background

The Personnel Assessment Research Unit (PARU) of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is responsible for conducting research to optimize the potential of the individual Soldier through maximally effective selection, classification, and retention strategies, with an emphasis on the changing needs of the Army as it transforms into the future force.

The *Validating Future Force Performance Measures (Army Class)* research program is a continuation of separate but related efforts that ARI has pursued since 2000 to ensure the Army is provided with the best personnel to meet the emerging demands of the 21st century. This research program is intended to support changes to the Army enlisted personnel selection and classification system that will result in improved performance, increased Soldier satisfaction, and extended service continuation. The current selection and classification system relies primarily on the Armed Services Vocational Aptitude Battery (ASVAB), which is a cognitive aptitude test battery.

Army Class builds on three prior research efforts. These are *Maximizing Noncommissioned Officer (NCO) Performance for the 21st Century* (NCO21; Knapp, McCloy, & Heffner, 2004)*; New Predictors for Selecting and Assigning Future Force Soldiers* (Select21; Knapp, Sager, & Tremble, 2005); and *Performance Measures for 21st Century Soldier Assessment* (PerformM21; Knapp & Campbell, 2006). The NCO21 research identified and validated non-cognitive predictors of NCO performance for use in the junior NCO promotion system. The Select21 research provided new personnel tests to improve the ability to select and assign first-term Soldiers with the highest potential for future jobs. The Select21 effort validated new and adapted individual difference measures against criteria representing both technical and non-technical (i.e., can-do and will-do) aspects of performance. Finally, the PerformM21 research examined the feasibility of instituting routine competency assessments for enlisted personnel. Accordingly, the researchers focused on developing cost-effective job knowledge assessments and examining the role of assessment within the overall structure of Army operational, education, and personnel systems. Because of their unique but complementary emphases, these three research efforts provided a strong theoretical and empirical foundation (including the identification of potential predictors and criteria) for the current project.

Overview of the Army Class Research Program

The Army Class effort began in 2006 with contract support from the Human Resources Research Organization (HumRRO). In the first year (2006), there were three distinct activities—one supporting military occupational specialty (MOS) reclassification of experienced Soldiers and two supporting pre-enlistment MOS classification. The first activity explored the idea that job knowledge tests (JKTs) could potentially be used to facilitate reclassification of experienced Soldiers by assessing knowledge and skills applicable to their new MOS, then focusing retraining on areas of deficiency. The project team thus developed prototype JKTs for several MOS (Moriarty, Campbell, Heffner, & Knapp, 2009). The banks of test items developed for this demonstration effort also were used to construct the performance criterion JKTs used in the Army Class validation research.

Given the resources required to conduct classification research in the Army that supports the needs of over 200 MOS, the second Year 1 activity involved obtaining recommendations for performing large-scale classification research from an expert panel (Campbell et al., 2007). The third Year 1 activity was a concurrent validation of the battery of experimental pre-enlistment predictor and criterion measures developed in Select21 (Knapp et al., 2005). The goal of the concurrent validation was to supplement the Select21 database to better support classification analyses because the Select21 job-specific samples were insufficient for this purpose. Although the classification analyses using the combined Select21/Army Class concurrent validation database were still based on a relatively small sample of incumbent Soldiers in the target MOS, results indicated that the experimental predictor measures showed promise for enhancing the classification of entry-level Soldiers (Ingerick, Diaz, & Putka, 2009).

In Year 2 (2007), the planned longitudinal criterion-related validation effort was initiated with the administration of experimental predictor measures to over 11,000 new Soldiers. At the same time, the emphasis of the Army Class research shifted to more fully focus on initial Soldier selection—a topic of great interest to Army policymakers. This heightened interest in immediate improvements to the Soldier selection process was also reflected in the initiation of a companion ARI project entitled *Expanded Enlistment Eligibility Metrics (EEEM)*. The EEEM effort had a shorter timeframe for making recommendations to the Army about the use of new pre-enlistment tests to supplement the ASVAB. The EEEM project capitalized on the Army Class longitudinal validation and led to the addition of two experimental pre-enlistment measures to the research predictor set—an experimental version of the Assessment of Individual Motivation (AIM) and the Tailored Adaptive Personality Assessment System (TAPAS).

In Year 3 of the research program (2008), training performance criterion data were collected for the longitudinal validation sample as Soldiers completed Advanced Individual Training (AIT) or One-Station Unit Training (OSUT). Some data were collected using assessments developed for Army Class and other data were obtained from archival databases, on variables like attrition and training course scores. For the Army Class effort, the analyses examined the extent to which the experimental pre-enlistment measures from Select21 predicted training criteria using the full training criterion sample (Knapp & Heffner, 2009). The EEEM analyses were conducted earlier in the year using training criteria collected to that point (Knapp & Heffner, 2010) with the goal of identifying predictors to recommend to the Army for immediate use in an Initial Operational Test and Evaluation (IOT&E) starting in 2009.

In Year 4 (2009) of Army Class, we collected in-unit job performance data on Soldiers from the longitudinal validation sample in an effort to get them when most would have been working in their units for 12 to 24 months. Years 5 and 6 (2010 and 2011) included a second round of in-unit job performance data collection from Soldiers in the longitudinal validation sample. Collection and analysis of the two rounds of in-unit performance criterion data is the subject of the present report.

Year 6 also will include additional analysis work based on the full longitudinal database. Final documentation of the method, findings, and recommendations coming out of the Army Class research program will be produced in the form of two capstone reports scheduled for publication in early 2012. One report will be geared primarily to a technical audience and the other report will be geared to a general Army audience.

Regarding future plans for EEEM, the program has transitioned into a multi-year IOT&E of the *Tier One Performance Screen (TOPS)*. In the TOPS program the TAPAS is being administered to Army applicants as part of the computerized test platform used by the Military Entrance Processing Command (MEPCOM). The Work Preferences Assessment (WPA) will be added to the IOT&E initiative in 2011. Similar to Army Class, the TOPS IOT&E calls for the collection of training and in-unit performance data but from Soldiers who are administered the predictors during pre-enlistment testing on an operational, rather than experimental, basis. The IOT&E work will be documented in a separate series of reports.

Overview of Report

The present report describes the collection of in-unit performance criterion data from the Army Class longitudinal validation sample. It details the measures, data collection strategy, sample, and psychometric characteristics of the in-unit criterion measures. Selection and administration of the predictor and training criterion measures is documented in Knapp and Heffner (2009) and they are briefly described here, along with the results of analyses using the in-unit criterion data. Note that this report focuses on purely empirical evaluations of the experimental measures. Other considerations pertaining to suitability for operational implementation of these measures (e.g., administration time, redundancy in content across measures, and potential for response distortion) are discussed in Knapp and Heffner (2010).

The remainder of this report is organized as follows: Chapter 2 describes the Army Class research design and measures. Chapter 3 describes the in-unit performance data collection. Chapter 4 discusses the database work and Chapters 5 through 8 describe the results of various sets of data analyses. Finally, the report ends with Chapter 9 which summarizes the latest Army Class research findings and next steps.

CHAPTER 2: LONGITUDINAL RESEARCH DESIGN

Deirdre J. Knapp (HumRRO), Tonia S. Heffner, and Kimberly S. Owens (ARI)

This chapter describes the research design for the Army Class longitudinal validation, beginning with the sample selection strategy and plan for collecting data from participating Soldiers at up to four points in time. We then provide descriptions of both the criterion and predictor measures.

Data Collection Points and Sample

In 2007 through early 2008, predictor data were collected from new Soldiers as they entered the Army through one of four Army reception battalions. Training performance criterion data were subsequently obtained on participating Soldiers at the completion of their Initial Military Training (IMT)[1]—either Advanced Individual Training (AIT) or One-Station Unit Training (OSUT), as applicable to the MOS. The training criterion data collection included only Soldiers who were in one of the six MOS listed below. In 2009, in-unit *job* performance criterion data were collected from over 1,500 Soldiers in the longitudinal validation sample. We collected in-unit performance data again in 2010-2011 when most Soldiers will have about 3 years of service. This plan should thus yield data collected from at least a subset of the participating Soldiers at four different points in their Army careers.

Soldiers were drawn from two types of samples: (a) MOS-specific samples targeting six entry-level jobs and (b) an Army-wide sample with no MOS-specific membership requirements. The six MOS-specific samples targeted the following occupations:

- 11B (Infantryman)
- 19K (Armor Crewman)
- 31B (Military Police)
- 68W (Health Care Specialist)
- 88M (Motor Transport Operator)
- 91B[2] (Light Wheeled Vehicle Mechanic)

These six MOS, individually and collectively, were selected on the basis of multiple considerations, especially their importance to the Army's mission (e.g., as measured by the number of Soldiers in the MOS) and the feasibility of developing MOS-specific criterion measures for use in the research within the specified timeframe.

The resulting sample includes Soldiers from all Army components—Regular Army (RA), U.S. Army Reserve (USAR), and the U.S. Army National Guard (ARNG).

[1] Formerly known as Initial Entry Training (IET).
[2] During the course of this research, the designation for Light Wheeled Vehicle Mechanic was changed from 63B to 91B.

Criterion Measures

Overview

Across the three criterion measurement points, we operationally defined success in the Army as scores using four types of indices: (a) job knowledge tests (JKTs), (b) performance rating scales, (c) attitudinal variables, and (d) administrative data. Development and descriptive details for the most of the performance criterion measures can be found in Moriarty et al. (2009). Information on scoring is provided in Chapter 5 of this report. With the exception of a set of combat-oriented rating scales developed specifically for the in-unit 2 data collection, the same measures were used for both the in-unit 1 and in-unit 2 data collections.

In-Unit Criterion Measure Descriptions

Job Knowledge Tests (JKTs)

Depending upon the MOS, many JKT items were drawn from items originally developed in PerformM21 (Knapp & Campbell, 2006), Select21 (Collins, Le, & Schantz, 2005), and Project A (Campbell & Knapp, 2001). Most of the JKT items were in a multiple-choice format with two to four response options; however, other formats, such as multiple response (i.e., check all that apply), rank ordering, and matching were also used. Many items referred to images in order to reduce reading requirements. Each in-unit JKT (WTBD and MOS-specific) comprised approximately 40 items.

Performance Rating Scales (PRS)

The behaviorally anchored PRS also had roots in previous research (see Moriarty et al., 2009 for details). Table 2.1 provides example scales from both the AW and MOS-specific PRS. The number of dimensions per set of scales ranged from four to eight for the MOS-specific PRS and the AW PRS had 14 dimensions. Each dimension was assessed through one item. The in-unit scales were completed by supervisors of the target Soldiers. Response options ranged from 1 (lowest) to 7 (highest) and included a "not applicable" option as well. The 14 AW PRS scales are shown below:

- Performing Core Warrior Tasks
- Performing MOS-Specific Tasks
- Communicating with Others
- Processing Information
- Solving Problems
- Exhibiting Effort
- Exhibiting Personal Discipline
- Contributing to the Team
- Exhibiting Fitness and Bearing
- Interactions with Indigenous People and Soldiers from other Countries
- Following Safety Procedures
- Developing Own Skills
- Managing Personal Matters
- Leadership Potential

Anticipating that Soldiers in the in-unit 2 data collection would generally have experience working under deployment conditions, we developed a supplemental set of rating scales for rater-ratee pairs who had been jointly deployed. The Combat/Deployment Performance Rating Scales (CDPRS) used the same format as the AW PRS, and included the following scales:

- Field/Combat Judgment
- Field Readiness
- Physical Endurance
- Physical Courage
- Awareness and Vigilance

Details on development of the CDPRS are provided in Appendix A.

Army Life Questionnaire (ALQ)

The ALQ was designed to measure Soldiers' self-reported attitudes and experiences. The original form of the ALQ was developed in the Select21 project (Van Iddekinge, Putka, & Sager, 2005). The in-unit ALQ yields 13 scale scores that cover (a) deployment adjustment, (b) objective performance, and (c) commitment and fit attitudes. Table 2.2 provides example scales and items.

Table 2.1. Example In-Unit Performance Rating Scales

Focus	Name	Description
MOS-Specific	Responds to Emergency Situations	Responds to life-threatening situations at accident sites, in the field, or in emergency rooms (performs triage, determines and applies treatment).
Army-Wide	Solves Problems	Adapts to new problem situations; applies prior training, rules, and strategies correctly; weighs alternatives when making decisions; develops novel solutions to problems; completes tasks despite major changes.

Table 2.2. Example In-Unit ALQ Scales

General Category	Name	Description	Example Item
Attitudinal Measures (In-unit)	Affective Commitment	Seven-item scale measuring Soldiers' emotional attachments to the Army	*I feel like I am part of the Army 'family.'*
Deployment	Deployment History/Tempo	Three-item scale measuring Soldiers' deployment history.	*How many total months have you been deployed?*

Attrition

Attrition data were obtained on Soldiers from the original longitudinal validation predictor sample at quarterly intervals throughout the course of the research, with a final data

capture scheduled for the end of CY2011. Attrition information is extracted for participating Soldiers from the Two Tier Attrition Screen (TTAS) database maintained by the U.S. Army Accessions Command. The attrition analyses were limited to Regular Army Soldiers due to difficulties in obtaining accurate separation data on Soldiers in the Reserve Components.

Predictor Measures

Overview

The starting point for the identification and preparation of experimental predictor measures for the longitudinal validation was the Army's Select21 project. Given the Army Class project's initial emphasis on classification, the original primary goal was to identify predictors likely to prove useful for classification purposes. The secondary goal was to assess selection-oriented predictors that needed additional research in a predictive validation (as opposed to concurrent validation) context. Two logistical constraints—a 2-hour administration time limit and the requirement for paper-based administration (because of the large numbers of Soldiers to be tested in single sittings)—made selection of the predictors very simple. Several desirable predictor measures *requiring* computer administration (notably the Work Suitability Inventory [WSI], Work Values Inventory [WVI], and the Record of Pre-Enlistment Training and Experience [REPETE]) could not be included in the longitudinal administration plan, thus permitting all remaining measures to be selected.

After the Army Class predictor data collection was underway, the ARI EEEM project was initiated and resulted in the addition of two additional predictor measures—the AIM and TAPAS—to the data collection plan. To do this without violating administration time restrictions, we temporarily suspended administration of some of the originally selected predictors while data from a sufficient number of new Solders were collected on the AIM and TAPAS. Thus, the sample sizes for several predictor measures are noticeably smaller. Table 2.3 summarizes the predictor measures included in the Army Class research.

Description of Predictors

Current Army Selection and Classification Instruments

Three metrics for selecting and classifying Soldiers were used in the present research as baseline measures for evaluating the experimental measures, depending on the criterion of interest. They are (a) the full ASVAB, (b) the Armed Forces Qualification Test (AFQT), and (c) Education Tier. Given their various purposes, as described below, the ASVAB was used as the primary measure for evaluating the experimental measures as classification instruments, the AFQT was used as the primary basis of comparison for evaluating the experimental measures as predictors of Soldier performance, and Education Tier was used as the primary basis for comparison for evaluating the experimental measures as predictors of first-term Soldier attrition.

Table 2.3. Summary of Longitudinal Validation Predictor Measures

Predictor Measure	Description
Baseline Predictors	
Armed Services Vocational Aptitude Battery (ASVAB) and Armed Forces Qualification Test (AFQT)	The ASVAB contains nine subtests, which are formed into composites used for Soldier selection and classification. The AFQT measures general cognitive ability and is a rationally weighted composite based on four ASVAB subtests (Arithmetic Reasoning, Mathematics Knowledge, Word Knowledge, and Paragraph Comprehension). Applicants must meet a minimum score on the AFQT to enter the Army. Applicants must meet a minimum score on various Aptitude Area (AA) composites in order to be classified into particular MOS.
Education Tier	Education Tier classifies an applicant's educational credential into one of three categories (Tier 1, 2, and 3). Tier 1 constitutes a high school diploma or more (e.g., a college degree), while Tier 2 constitutes a non-high school diploma (e.g., a General Educational Development [GED] credential).Tier 3 applicants (no high school credential) are not allowed to enlist and the number of Tier 2 Soldiers allowed to enlist is restricted.
Cognitive Predictor	
Assembling Objects (AO)	Measures spatial ability. AO is currently administered as part of the ASVAB, but until recently had not been used to screen or select Army applicants. AO is now included in the Two Tier Attrition Screen (TTAS) used to screen applicants who have not earned a high school diploma.
Temperament Predictors	
Assessment of Individual Motivation (AIM) – EEEM	Measures six temperament characteristics predictive of first-term Soldier attrition and performance (e.g., work orientation, dependability, adjustment). Each item consists of four behavioral statements. Respondents are asked to self-select the statement that is most descriptive of them and the statement that is least descriptive of them.
Tailored Adaptive Personality Assessment System (TAPAS-95s) – EEEM	Measures 12 dimensions or temperament characteristics predictive of first-term attrition and performance (e.g., dominance, attention-seeking, intellectual efficiency, physical conditioning). Uses a multidimensional pairwise preference (MDPP) format in which respondents indicate which of two statements is most like them.
Rational Biodata Inventory (RBI)	Measures 14 temperament and motivational characteristics important for entry-level Soldier performance and retention. Items ask respondents about their past behavior, experiences, and reactions to previous life events (e.g., the extent to which they enjoyed thinking about the "plusses and minuses" of alternative approaches to solving a problem).
Predictor Situational Judgment Test (PSJT)	Measures respondents' judgment and decision-making proficiency across situations commonly encountered prior to or during the first enlistment term (e.g., dealing with a difficult co-worker). Each item consists of a description of a problem situation and a list of four alternative actions that the respondent might take in that situation. Respondents rate the effectiveness of each action.
Person-Environment (P-E) Fit Predictors	
Work Preferences Assessment (WPA)	Measures respondents' preferences for different kinds of work activities and settings offered by different jobs (e.g., working with others, repairing machines or equipment). Items ask respondents to rate how important a series of characteristics is to their ideal job. Content is based on Holland's (1997) theory of vocational personality and work environment.
Army Knowledge Assessment (AKA)	Measures respondents' understanding or expectations about the kinds of work activities and settings typically offered by the Army. Respondents are asked to read a brief description of six work settings and then rate the extent to which they think each setting describes the Army. Like the WPA, content is based on Holland's (1997) theory of vocational personality and work environment.

Aptitude Area (AA) composites composed of ASVAB subtests are used to classify Soldiers into their MOS. For this reason, the full ASVAB was used as the baseline for evaluating the experimental measures as classification instruments.

The AFQT is a rationally weighted composite of four ASVAB subtests (Arithmetic Reasoning, Math Knowledge, Word Knowledge, and Paragraph Comprehension). Scores on the AFQT provide an assessment of an applicant's general cognitive ability. AFQT is used in conjunction with high school degree status and medical and moral screens to evaluate applicants for enlistment. Examinees are classified into categories based on their AFQT percentile scores (Category I = 93–99, Category II = 65–92, Category IIIA = 50–64, Category IIIB = 31–49, Category IV = 10–30, Category V = 1–9).

Finally, Education Tier classifies individuals with a high school diploma or equivalent into Education Tier 1, those with an alternative high school credential (e.g., General Educational Development) into Education Tier 2, and those with no educational credential into Education Tier 3. The number of Tier 2 Soldiers allowed to enlist is restricted because previous research has shown that they are much more likely to attrit in their first term of service than Tier 1 Soldiers (Knapik, Jones, Haurik, Darakjy, & Piskador, 2004).

Assembling Objects (AO)

We included scores on the Assembling Objects (AO) portion of the ASVAB as an experimental predictor to be evaluated in the Army Class research.[3] The AO subtest is administered to U.S. military applicants as part of the ASVAB but until recently had not been used to screen or select Army applicants. AO measures spatial ability and was first developed in Project A (Russell et al., 2001). The items are graphical in nature, requiring respondents to visualize how an object will look when its parts are put together correctly. Past research has shown that AO could supplement one or more of the existing ASVAB subtests in predicting entry-level Soldier performance, while potentially yielding lower gender differences than subtests measuring comparable abilities (Peterson et al., 1992; Russell, Reynolds, & Campbell, 1994).

Assessment of Individual Motivation (AIM)

The AIM was added to the Army Class longitudinal validation as part of the EEEM initiative. The original AIM was developed to address faking concerns with the otherwise promising Assessment of Background and Life Experiences (ABLE) developed in Project A (White & Young, 1998; White, Young, & Rumsey, 2001). The AIM uses a forced-choice format to reduce fakability. Each item consists of four behavioral statements (i.e., tetrads). Respondents are asked to select, among four alternative statements, (a) the statement that is most descriptive of them and (b) the statement that is least descriptive of them. The AIM measures six temperament characteristics predictive of first-term Soldier attrition and performance: Dependability (Non-Delinquency), Adjustment, Physical Conditioning, Leadership, Work Orientation, and Agreeableness. The version of AIM administered in this research has 30 items.

[3] AO is now included in the Two Tier Attrition Screen (TTAS) used to screen applicants who have not earned a high school diploma.

Currently, the AIM is used operationally by the Army in the TTAS program to screen Tier 2 applicants.

Tailored Adaptive Personality Assessment System (TAPAS-95s)

The TAPAS-95s was also added to the Army Class project as part of the EEEM effort. Developed by the Drasgow Consulting Group under the Army's Small Business Innovation Research (SBIR) program (Drasgow, Stark, & Chernyshenko, 2006; Stark, Drasgow, & Chernyshenko, 2008), the TAPAS-95s assesses 12 personality dimensions over 95 items. The instrument builds on the AIM's ability to measure a host of narrow personality constructs (facets) known to predict success in the military while incorporating features designed to enhance resistance to faking. Examples of the constructs assessed by the TAPAS include Dominance, Attention-Seeking, Intellectual Efficiency, and Physical Conditioning. Soldiers taking the TAPAS must select which of two statements is more descriptive of them. The version of the TAPAS administered in this research was a static, non-adaptive surrogate for an item response theory (IRT)-based computerized adaptive personality assessment system capable of measuring up to 22 facets of potential interest to the Army.

Rational Biodata Inventory (RBI)

The RBI measures multiple temperament or motivational characteristics important to entry-level Soldier performance and retention (Kilcullen, Putka, McCloy, & Van Iddekinge, 2005). The measure has evolved in various ways depending on the application but grew out of the Assessment of Right Conduct (Kilcullen, White, Sanders, & Hazlett, 2003) and the Test of Adaptable Personality (Kilcullen, Mael, Goodwin, & Zazanis, 1999). Thus, with varying sets of items, it has been used in prior Army research and operational applications (e.g., for selection into Special Forces) for almost a decade. Items on the RBI ask respondents about their past behavior, experiences, and reactions to previous life events using Likert-style response options (e.g., the extent to which they enjoyed thinking about the plusses and minuses of alternative approaches to solving a problem). The RBI yields scores on a range of attributes (e.g., Achievement Motivation, Cognitive Flexibility, Fitness Motivation, Hostility to Authority, Peer Leadership, Self-Efficacy, and Stress Tolerance). The RBI used in the Army Class longitudinal validation has 101 items covering 14 attributes and is the same version used in the Select21 research (Kilcullen et al., 2005).

Predictor Situational Judgment Test (PSJT)

The PSJT is a 20-item paper-and-pencil measure designed to assess an individual's judgment and decision-making proficiency in challenging situations (e.g., working with uncooperative peers to accomplish a task; determining when to handle a problem alone versus consulting a supervisor; Waugh & Russell, 2005). The situations presented in the PSJT are civilian counterparts to the kinds of situations typically encountered by Soldiers during their first few months in the Army. These situations (and their underlying dimensions) were identified through collection of critical incidents from Soldiers in IMT. Each item consists of a description of a situation followed by four actions that might be taken in that situation. Respondents rate the effectiveness of each action on a 1 to 7 scale (from "Ineffective" to "Very Effective"). The PSJT targets five kinds of situations or dimensions important to first-term Soldier performance: (a)

Adaptability to Changing Conditions, (b) Relating to and Supporting Peers, (c) Teamwork, (d) Self-Management, and (e) Self-Directed Learning. Although the PSJT items were written to reflect these dimensions, the measure is designed to yield a single total score.

Work Preferences Assessment (WPA)

The Work Preferences Assessment (WPA) is designed to assess an individual's preferences (or fit) for different kinds of work activities and environments (Van Iddekinge et al., 2005). The 72 items comprising the WPA were written to measure each of the six dimensions and their subfacets underlying Holland's (1997) theory of vocational personality and work environment. According to Holland's theory, work interests are expressions of personality that can be used to categorize individuals and work environments into six types (or dimensions): Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C). For each dimension or facet, the WPA contains three types of items: (a) interests in work activities (e.g., "A job that requires me to teach others"), (b) interests in work environments or settings (e.g., "A job that requires me to work outdoors"), and (c) interests in learning opportunities (e.g., "A job in which I can learn how to lead others"). Respondents are asked to rate each item in terms of its importance to their ideal job using a 5-point Likert-type scale (1 = "Extremely unimportant to have in my ideal job" to 5 = "Extremely important to have in my ideal job") (Putka & Van Iddekinge, 2007).

The WPA yields six dimension scores (corresponding to each of the six RIASEC dimensions) and 14 facet scores (corresponding to facets underlying the six RIASEC dimensions). These raw scores can then be combined or modified based on additional data to obtain multiple, alternative sets of scores for use in one or more of the Army's personnel management objectives.

Army Knowledge Assessment (AKA)

The Army Knowledge Assessment (AKA) is a 30-item instrument that assesses Soldiers' knowledge about the extent to which the current Army (in general) supports each RIASEC dimension (Van Iddekinge et al., 2005). Respondents read a brief description of six work settings and then rate the extent to which they think each setting describes the Army. The AKA yields six dimension scores, corresponding to the six RIASEC dimensions defined by Holland (1997). These raw scores can then be combined or modified based on additional data to obtain alternative sets of scores for use in one or more of the Army's personnel management objectives. Conceptually, the AKA differs from the WPA in that it indicates whether respondents have realistic expectations about the interests that would be satisfied with Army life whereas the WPA indicates whether respondents are interested in what Army life offers. Both are strategies for predicting person-environment fit.

CHAPTER 3: IN-UNIT DATA COLLECTION

Karen O. Moriarty, Charlotte H. Campbell (HumRRO), and Kimberly S. Owens (ARI)

Overview

The research plan called for two rounds of in-unit job performance criterion data collection from all Soldiers in the longitudinal validation predictor sample (n = 11,065). The first round occurred between January and August of 2009, when most Soldiers in the sample had 12 to 24 months time-in- service (TIS). The second round, when Soldiers had about 3 years TIS, began in July 2010 and concluded in March 2011.

All of the criterion measures (described in Chapter 2) were administered via the Internet, in either proctored or unproctored sessions. After logging on to the Army Class Soldier web page, Soldiers proceeded through the assessment at their own pace. They were first presented with project background information and the Privacy Act Statement, which provided the authority and procedures under which ARI can collect research data. Soldiers were assured of confidentiality and advised their participation was voluntary. They answered several background questions (e.g., pay grade) and provided contact information for their supervisors. All Soldiers completed the ALQ and WTBD JKT, and those who were in one of the six target MOS also completed an MOS-specific JKT. The complete assessment required 40–90 minutes per Soldier.

Similarly, supervisors logged on to the Army Class Supervisor web page, where the project background information and Privacy Act Statement were presented. Supervisors received a short rater training primer (e.g., how to avoid halo error) and were assured that Soldiers would not see their ratings nor would the ratings become a part of Soldiers' Army records. Supervisors first completed the AW rating scales, and if their Soldier(s) was in one of the target MOS, they also completed MOS-specific rating scales.

Prior to collecting data, all measures and procedures were reviewed and approved by both HumRRO's and ARI's, Institutional Review Boards.

Staff Training

A data collection procedure manual was developed as a basis for training staff involved in data collection activities. The manual included information on the project and the measures, instructions for setting up the computers and rooms, and procedures for documenting data and quality control issues, such as identification (ID) number errors or Soldiers progressing through the assessment too quickly. It also included guidance on coordinating unproctored data collections, although those procedures were frequently modified as the result of command-specific coordination decisions. This manual was updated during the course of the data collection periods to reflect lessons learned.

General Procedure

Planning of the data collection process began with review of Soldier rosters obtained from the U.S. Army Human Resources Command (HRC), with quarterly updates. These Master

Rosters told us where the target Soldiers were assigned, and other information regarding their status and availability. Based on the number of Soldiers at a given location, the decision was made to conduct either proctored or unproctored data collections. Generally, if fewer than 25 eligible Soldiers were stationed at a single location, an unproctored data collection was planned. For the Reserve Component (RC), proctored assessments were generally infeasible because the Soldiers are widely distributed throughout the country and assemble at single locations (usually without digital classrooms) only intermittently.

ARI made contact with the major commands by means of Research Support Requests (RSRs) to identify testing dates and points of contact (POCs) with whom subsequent arrangements would be coordinated. HumRRO and ARI worked together on coordination with the POCs to negotiate for time, space, and computers. Frequent communications were critical in preparing for data collection visits.

For the unproctored data collections, coordination involved determining the dates for sending the materials to the POC for distribution to subordinate units and to individual Soldiers and supervisors (or directly to the Soldiers and supervisors), and a suspense date for completion. We recognized that there were potential problems with this approach, including lack of control over the testing environments and lack of on-site assistance to answer participants' questions. Therefore, we established a Help Desk which Soldiers or supervisors could call or email for help with problems.

Regular Army Data Collection

Across the two rounds of data collection, teams made 37 visits to 17 locations to proctor on-site data collections (see Table 3.1). These visits typically involved multiple 2-hour testing sessions each day. Soldiers and supervisors were provided background information and login instructions, and then allowed to complete the assessment at their own pace. Instructions for participation in unproctored assessments were left or emailed to the post POC to distribute to Soldiers or supervisors who were unable to attend any of the proctored sessions. In all, we obtained data from about 19% of the Regular Army (RA) target Soldiers from the predictor sample in in-unit 1 and 15% of the RA target Soldiers in in-unit 2.

Reserve Component Data Collection

Since there were almost 5,500 RC Soldiers on the Master Roster (i.e., predictor sample), we made diligent efforts to reach these Soldiers and seek their participation, as well as that of their supervisors. ARI submitted RSRs to both the USAR and ARNG requesting support in communicating the project to RC Soldiers. After discussions with USAR and ARNG representatives, it was determined that it was generally infeasible to use troop support tasking processes to reach RC Soldiers and a more direct approach was advisable. We therefore emailed RC Soldiers and Supervisors directly and individually. This strategy was not as successful as the in-person visits we used for most of the RA data collections. We obtained data from about 10% of the target RC Soldiers in the first round and just over 2% in the second round of in-unit data collections. Final RA and RC sample sizes and sample descriptions are provided in Chapter 4.

Table 3.1 In-Unit Proctored Data Collection Site Visits

Site	In-Unit 1	In-Unit 2
Fort Benning, GA	51	27
Fort Bliss, TX	117	35
Fort Bragg, NC	83	137
Fort Campbell, KY	69	-
Fort Carson, CO	64	46
Fort Drum, NY	85	-
Fort Eustis, VA	14	-
Fort Hood, TX	92	-
Fort Knox, KY	-	31
Fort Lee, VA	6	-
Fort Lewis, WA	40	44
Fort Myer, VA	32	-
Fort Polk, LA	47	56
Fort Riley, KS	38	33
Schofield Barracks, HI (USARPAC)	30	70
Fort Stewart, GA	65	49
Fort Wainwright, AK	-	38
Hunter AAF – 3rd & RSTB (USAR)	12	-
Germany (USAREUR)	90	-
TOTAL	935	566

Note. These numbers are based on trip reports provided by data collection staff and do not reflect final, post-data collection cleaning sample sizes. The numbers do not include supervisor raters.

CHAPTER 4: DATABASE DEVELOPMENT

Karen O. Moriarty, Matthew Trippe, and Laura Ford (HumRRO)

During the course of the in-unit data collections, HumRRO received and cleaned raw data from ARI on a regular basis. At the conclusion of the data collections, the criterion data were analyzed and scored. Attrition data were updated quarterly. Finally, the in-unit data were added to the predictor and training criterion data to create a master longitudinal database with over 5,000 variables.

Database Construction

Data Cleaning

The in-unit 1 and in-unit 2 criterion data captured on the ARI server were sent to HumRRO weekly. One file contained Soldier data and a second file contained supervisor ratings. Once the data were processed, we merged Soldier-level records with the predictor data records, matching on Soldier ID. Data cleaning followed the same rules and protocols implemented in Select21 with regard to treatment of missing data and identification of Soldiers with questionable or suspect data (e.g., when more than 10% of data for a score was missing, the score was set to missing) (Knapp & Tremble, 2007).

We evaluated cumulative time completing the assessment, pattern responding, and missing data as well. As mentioned, the different Soldier measures in this data collection were actually presented as one assessment. So, the "time-taken" variable includes the time the Soldier took to complete *all* measures. Soldiers with a cumulative time taken value of fewer than roughly 5 minutes were flagged. Our reasoning was that the ALQ (the first measure administered in the assessment) could be completed in that amount of time. However, we put the most emphasis in determining whether a case should be excluded from analyses on the proportion of missing items rather than a time-taken variable because the proportion of missing items could be calculated individually by measure.

The supervisor ratings data were cleaned in a similar fashion. A modification to the "missing 10% or more responses" was made for the ratings. Ratings were flagged as unusable where supervisors were missing more than 10% of their ratings *or* selected "cannot rate" for more than 50% of the scales.

Scoring the Assessments

Measures were scored following the same rules and procedures used in previous research (Ingerick et al., 2008; Knapp & Tremble, 2007). We examined item-level statistics (e.g., frequencies, item-total correlations, item difficulties) to determine if there were poorly performing items that should be dropped when computing a total score. From there, the criteria were scored and the data provided to the analysis team.

Attrition Database

To support the attrition analyses, we obtained quarterly extracts of attrition data from the Tier Two Attrition Screen (TTAS) database starting in the first quarter of FY09. No additional preparation or cleaning of these data was required. All attrition data through the first quarter of FY11 extract were included in the longitudinal database.

Master Longitudinal Database

The master longitudinal database, formatted in SPSS, consists of the following data elements: (a) predictor data, (b) training criterion data, (c) first-round in-unit criterion data, (d) second-round in-unit criterion data, and (e) administrative data from Army personnel databases at the item level. The predictor data, training criterion data, and administrative data were collected during previous data collections/data analyses and were simply added here. The full database has 11,068 records and 5,658 variables. The database documentation includes copies of all measures and syntax used for scoring them.

Sample Description

Tables 4.1 through 4.5 provide summary demographic information on the in-unit 1 and 2 samples. Comparing Tables 4.1 and 4.2, we see that overall percentages of subgroups were fairly consistent between the first and second in unit samples. The in-unit 2 sample had a slightly higher percentage of males (86% vs. 81%) than the in-unit 1 sample. Across the two samples, approximately 78% of the Soldiers were White, approximately 13% Black, and 15% Hispanic.[4] The largest percentage of female Soldiers was found in the 68W (Health Care Specialist) MOS. About half of the sample was in our six target MOS for both in-unit samples. Although not shown in Tables 4.1 or 4.2, the percentage of Soldiers in each AFQT category varies somewhat by MOS. By definition, most Soldiers fall in the middle AFQT categories, but the percentage of Soldiers in Category I ranges from 2% to 25% across the target MOS. The percentage of Soldiers in Category IV ranges from 0 to nearly 15%. Similarly, education tier varies somewhat across MOS, with as few as 65% and as many as 87% in Tier 1. These numbers are consistent with the predictor sample of which these samples were a subset. Time in service (TIS) targets for the in-unit 1 and 2 samples were 12-24 months and 36 months, respectively. The mean TIS for the in-unit 1 sample was 20 months and the mean TIS for the in-unit 2 sample was 35 months.

[4] Note that race and ethnicity are independent demographic variables.

Table 4.1. In-Unit 1 Criterion Sample by MOS and Demographic Subgroup

Subgroup	Army-Wide		MOS												Subgroup Totals	
			11B		19K		31B		68W		88M		91B			
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender																
Male	593	73.8	311	100.0	95	100	161	75.9	23	59.0	41	67.2	56	86.2	1,280	80.7
Female	205	25.5	0	0.0	0	0.0	51	24.1	16	41.0	20	32.8	9	13.8	301	19.0
Race																
White	577	71.9	272	87.5	72	75.8	184	86.8	32	82.1	46	75.4	55	84.6	1,238	78.1
Black	141	17.6	15	4.8	8	8.4	13	6.1	2	5.1	11	18.0	7	10.8	197	12.4
Other	81	10.1	24	7.7	13	14.0	15	7.1	5	12.8	3	5.0	3	4.6	144	9.1
Ethnicity																
White Non-Hispanic	508	63.3	238	76.5	69	72.6	169	79.7	32	82.1	41	67.2	46	70.8	1,103	69.5
Hispanic	128	15.9	48	15.4	10	10.5	29	13.7	5	12.8	8	13.1	11	16.9	239	15.1
Totals	803	50.6	311	19.6	95	6.0	212	13.4	39	2.5	61	3.8	65	4.1	1,586	100.0

Note. The figures reported by subgroup and MOS do not add up to the totals due to missing data. Soldiers indicating more than one race (e.g., White and Black) or those identifying as American Indian/Alaska Native, Asian, or Native Hawaiian/Other Pacific Islander are coded as "Other." The sample sizes for individual criterion measures vary due to missing data. These data exclude Soldiers with prior military service.

Table 4.2. In-Unit 2 Criterion Sample by MOS and Demographic Subgroup

Subgroup	Army-Wide		MOS												Subgroup Totals	
			11B		19K		31B		68W		88M		91B			
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender																
Male	435	79.1	215	100.0	82	100.0	114	88.4	14	53.9	21	80.8	33	94.3	914	86.0
Female	111	20.2	0	0.0	0	0.0	15	11.6	12	46.2	5	19.2	2	5.7	145	13.6
Race																
White	386	70.2	183	85.1	69	84.2	112	86.8	23	88.5	19	73.1	30	85.7	822	77.3
Black	115	20.9	17	7.9	2	2.4	7	5.4	0	0.00	5	19.2	4	11.4	150	14.1
Other	47	8.6	14	6.5	10	12.2	10	7.8	3	11.5	2	7.7	1	2.9	87	8.2
Ethnicity																
White Non-Hispanic	323	58.7	162	75.4	63	76.8	106	82.2	23	88.5	20	76.9	27	77.1	724	68.1
Hispanic	99	18.0	34	15.8	10	12.2	15	11.6	2	7.7	0	0.0	4	11.4	164	15.4
Totals	550	51.7	215	20.2	82	7.7	129	12.1	26	2.45	26	2.5	35	3.3	1,063	100.0

Note. The figures reported by subgroup and MOS do not add up to the totals due to missing data. Soldiers indicating more than one race (e.g., White and Black) or those identifying as American Indian/Alaska Native, Asian, or Native Hawaiian/Other Pacific Islander are coded as "Other." The sample sizes for individual criterion measures vary due to missing data. These data exclude Soldiers with prior military service.

Table 4.3 shows that well over half of the in-unit 1 sample is from the Regular Army, which is higher than the percentage in the predictor sample (67% vs. 50%). This is even more pronounced in the in-unit 2 sample, in which 88% is Regular Army. This can be attributed to the low participation among reserve component Soldiers discussed in Chapter 3.

Table 4.3. In-Unit 1 and In-Unit 2 Criterion Sample by Component and MOS

MOS/Sample	Component						MOS Totals	
	Regular		ARNG		USAR			
	n	%	*n*	%	*n*	%	*n*	%
In-Unit 1								
11B/X	277	89.1	31	10.0	0	0.0	308	19.4
19K	89	93.7	6	6.3	0	0.0	95	6.0
31B	128	60.4	53	25.0	31	14.6	212	13.4
68W	17	43.6	14	35.9	8	20.5	39	2.5
88M	33	54.1	17	27.9	11	18.0	61	3.9
91B	35	53.8	13	20.0	17	26.2	65	4.1
Army-Wide	475	59.2	188	23.4	140	17.4	803	50.7
Totals	1,054	66.5	322	20.3	207	13.1	1,583	100.0
In-Unit 2								
11B/X	211	22.5	4	4.9	0	0.0	215	20.2
19K	80	8.5	2	2.5	0	0.0	82	7.7
31B	114	12.2	8	9.9	7	15.6	129	12.1
68W	18	1.9	5	6.2	3	6.7	26	2.5
88M	19	2.0	5	6.2	2	4.4	26	2.5
91B	32	3.4	3	3.7	0	0.0	35	3.3
Army-Wide	463	49.4	54	66.7	33	73.3	550	51.7
Totals	937	88.2	81	7.6	45	4.23	1,063	100.0

Note. One Soldier is missing component information. The figures reported do not add up to the totals due to missing data. These data exclude Soldiers with prior military service.

Table 4.4 shows the demographic information for the predictor, training, and both in-unit samples. The demographic characteristics of the predictor and in-unit samples closely parallel each other whereas the training sample is different because it includes data only from Soldiers in the six target MOS. In particular, it is more heavily male because of the relatively large proportion of combat MOS included in the target MOS sample. Given the relationship between education and attrition, it is also not surprising that the ratio of Tier 1 (high school degree or higher graduates) to Tier 2 (nongraduates) is higher in the in-unit samples than in the full predictor sample. Additionally, the percentage of Tier 2 Soldiers may appear unexpectedly high because the Army officially only admits 10% of these individuals. The difference is that some of those otherwise designated as Tier 2 are treated as Tier 1 for enlistment purposes if they pass the TTAS screen. We have useable archival attrition data for approximately 96% (5,174 of 5,370) of the original RA sample. Demographic information for the attrition analysis sample is in Table 4.5.

Table 4.4. Demographic Characteristics for the Predictor, Training, and In-Unit 1, and In-Unit 2 Samples

	Predictor Sample		Training Sample		In-Unit 1 Sample		In-Unit 2 Sample	
Subgroup	*N*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
Gender								
Male	8,646	80.0	2,083	90.8	1,280	80.7	914	86.0
Female	2,113	19.5	207	9.0	301	19.0	145	13.6
Race								
White	8,431	78.0	1,976	86.1	1,239	78.1	822	77.3
Black	1,527	14.1	157	6.8	197	12.4	150	14.1
Other	818	7.6	154	6.7	144	9.1	87	8.2
Ethnicity								
White Non-Hispanic	7,541	69.7	1,776	77.4	1,104	69.6	724	68.1
Hispanic	1,527	14.1	323	14.1	239	15.1	164	15.4
AFQT Category								
I	470	4.3	83	3.6	123	7.8	55	5.2
II	3,009	27.8	661	28.8	474	29.9	331	31.1
IIIA	2,676	24.7	637	27.8	350	22.1	276	26.0
IIIB	4,167	38.5	834	36.4	564	35.5	349	32.8
IV	414	3.8	72	3.1	65	4.1	45	4.2
Highest Education Level (at Entry)[a]								
Tier 1	8,103	74.9	1,667	72.7	1,234	77.8	827	77.8
Tier 2	2,682	24.8	625	27.2	353	22.2	236	22.2
MOS								
11B/X	1,790	16.6	671	29.3	311	19.6	215	20.2
19K	581	5.4	471	20.5	95	6.0	82	7.7
31B	1,484	13.7	716	31.2	212	13.4	129	12.1
68W	307	2.8	136	5.9	39	2.5	26	2.5
88M	512	4.7	72	3.1	61	3.8	26	2.5
91B	472	4.4	219	9.5	65	4.1	35	3.3
Army-Wide	5,654	52.3	9	0.4	803	50.6	550	51.7
Component								
Regular Army	5,370	49.7	1,387	60.5	1,054	66.4	937	88.2
ARNG	3,793	35.1	694	30.3	322	20.3	81	7.6
USAR	1,651	15.3	213	9.3	211	13.3	45	4.2
Totals	10,814	100.0	2,294	21.2	1,587	14.7	1,063	9.8

Note. The Training Sample reflects the number of Soldiers that participated in the training data collection, not the number for which we had archival training data. The "%" figures in the "Totals" row represent percent of the predictor sample. Soldiers indicating more than one race (e.g., White and Black) or those identifying as American Indian/Alaska Native, Asian, or Native Hawaiian/Other Pacific Islander are coded as "Other." The sample sizes for individual criterion measures vary due to missing data. These data exclude Soldiers with prior military service.

[a]The percentage of Tier 2 Soldiers may appear unexpectedly high because some of those otherwise designated as Tier 2 are treated as Tier 1 for enlistment purposes if they pass the TTAS screen. The Army officially only admits 10% of the Tier 2 Soldiers.

Table 4.5. Attrition Analysis Sample Demographics

	Valid Attrition Data	
Subgroup	*n*	%
Gender		
Male	4,386	84.8
Female	776	15.0
Race		
White	4,049	78.3
Black	686	13.3
Other	421	8.1
Ethnicity		
White Non-Hispanic	3,605	69.7
Hispanic	761	14.7
AFQT Category		
I	250	4.8
II	1,464	28.3
IIIA	1,320	25.5
IIIB	1,917	37.1
IV	204	3.9
Highest Education Level (at Entry)		
Tier 1	3,692	71.4
Tier 2	1,482	28.6
MOS		
11B/X	1,112	21.5
19K	416	8.0
31B	604	11.7
68W	107	2.1
88M	155	3.0
91B	182	3.5
Army-Wide	2,598	50.2
Total	5,174	100.0

Note. Sample excludes Soldiers with prior military service and those serving in the Army National Guard or the Army Reserves. The figures reported do not add up to the totals due to missing data.

CHAPTER 5: MEASURE SCORING AND PSYCHOMETRIC PROPERTIES

Matthew T. Allen, Tina Chang, and Michael J. Ingerick (HumRRO)

In this chapter we describe the scoring of the predictor and criterion measures and their psychometric properties as estimated in the in-unit 1 and in-unit 2 Army Class samples. The predictor measures are presented first, followed by the in-unit criterion measures. The Army Class training longitudinal validation report summarized the scoring procedures and psychometric properties for the training criterion measures (Knapp & Heffner, 2009).

Predictor Measure Scores and Associated Psychometric Properties

Armed Services Vocational Aptitude Battery (ASVAB) and Education Tier

Soldiers' AFQT, ASVAB, and Education Tier data were extracted from MEPCOM administrative records. Descriptive statistics and score intercorrelations are provided in Appendix B (Tables B.1 and B.2, respectively).

Assessment of Individual Motivation (AIM)

For each AIM item tetrad, respondents provided two responses—one indicating the statement that is *most* like them and one indicating the statement that is *least* like them. A quasi-ipsative scoring method generated four construct scores for each item (i.e., one score for each stem) based on whether the respondents indicated the stem was most like them, least like them, or neither. Scale scores were obtained by averaging (across items) the scores for stems measuring the same construct. A minimum of 80% of the items for any given construct must have been completed in order to obtain a score for that scale. Descriptive statistics and reliability estimates for the AIM scales are presented in Appendix B (Table B.3). The reliability estimates were all acceptable (ranging from .70 to .77). The mean validity (or lie scale) score was low, suggesting response distortion due to socially desirable responding was minimal.

Tailored Adaptive Personality Assessment System (TAPAS-95s)

For each TAPAS item pair, respondents selected the item that is most like them. TAPAS-95s scoring was based on multidimensional pairwise preference (MDPP) in which items were created by pairing statements subject to similarity constraints on social desirability and/or location (extremity). Item Response Theory (IRT) was used to determine the dimension scores using the model originally proposed by Stark (2002). A detailed presentation of the scoring procedure is provided in the EEEM technical report (Knapp & Heffner, 2009). Descriptive statistics are shown in Appendix B (Table B.5) and scale intercorrelations are shown in Table B.6.

Rational Biodata Inventory (RBI)

RBI scores were computed by summing responses to the items applicable to each scale and dividing by the number of items in the scale. A minimum of 75% of the items for any given

construct must have been completed in order to obtain a score for that scale. To ensure comparable results across the experimental measures, substantive scale scores were not adjusted using the "Lie" scale score. Descriptive statistics and reliability estimates are shown in Appendix B (Table B.7). Most of the reliability estimates approached or exceeded .70. The substantive scales with fairly low internal consistency reliability estimates were Narcissism (.55) and Gratitude (. 43). These reliability estimates, as well as the mean scores, are generally similar to results from the same version of the RBI used in the Select21 concurrent validation (Knapp & Tremble, 2007), with the highest score in both samples being Self-Efficacy and the lowest score being Hostility to Authority. Scale intercorrelations are provided in Table B.8.

Predictor Situational Judgment Test (PSJT)

For each PSJT item, the respondents rated the effectiveness of four possible actions in response to a hypothetical situation. The ratings were made on a 1 (ineffective) to 7 (very effective) response scale. The PSJT was scored in the manner developed and described by Waugh and Russell (2005). An initial judgment score for each response option was calculated using Equation 1 below.

$$Judgment\ Score_{Option\ x} = 6 - |\ Soldiers Rating_{Option\ x} - keyedEffectiveness_{Option\ x}\ | \qquad (1)$$

The keyed effectiveness ratings were based on judgments made by 67 subject matter experts during the Select21 project (Knapp & Tremble, 2007). We subtracted the difference between the respondent's rating and keyed effectiveness values from 6 to reflect the scores, so that higher values represented better scores. The judgment score for the entire test was the mean of the 80 option scores across the 20 scenarios. To minimize effects of a response pattern that recognizes that the keyed score will rarely be 1 or 7, the key was stretched as shown in Equations 2 and 3.

For original key values above 4.0, newValue = oldValue + 0.5 * (oldValue – 4). (2)
For original key values below 4.0, newValue = oldValue – 0.5 * (4 – oldValue). (3)

Finally, after stretching the key, we rounded the new value to the nearest integer. If the new value was less than one, we rounded it up to one; if the new value was greater than 7, we rounded it down to 7.

The mean PSJT score for the total sample was 4.67 (*SD* = .41, *n* = 4,970) and the coefficient alpha reliability estimate was .86. These results are consistent with those obtained from the Army Class and Select21 concurrent validation samples (Ingerick et al., 2009; Waugh & Russell, 2005).

Army Knowledge Assessment (AKA)

The AKA yields six dimension scores corresponding to each of Holland's (1997) six RIASEC dimensions. Items for each scale were averaged to create a total score for that scale. Total scores on each facet ranged from one to five. Descriptive statistics and reliability estimates for the AKA scales are shown in Table B.9. With the exception of Realistic Interests, which had a reliability estimate of .76, estimates for the remaining scales were high, ranging from .81 to

.89. The scale with the highest mean score, not surprisingly for a sample of Soldiers, was Realistic Interests. AKA scale intercorrelations are shown in Table B.10.

Work Preferences Assessment (WPA)

The WPA yields six raw dimension scores (corresponding to each of the six RIASEC dimensions) and 14 facet scores (corresponding to the subfacets underlying the six RIASEC dimensions). Raw scale scores were computed by obtaining the average of the scores across the items constituting each dimension or facet. Total raw scale scores range from one to five.

Descriptive statistics and reliability estimates for both the dimension and facet scores are shown in Table B.11. Most reliability estimates are relatively high (mid-.70s to .90). Several of the facet scores were a bit lower, with Clear Procedures (a facet of Conventional Interests) being the score with the lowest estimated reliability (.64). The WPA score intercorrelations are shown in Table B.12.

In-Unit Criterion Measure Scores and Associated Psychometric Properties

In-Unit Job Knowledge Tests (JKTs)

The in-unit 1 and in-unit 2 JKTs were developed and scored the same way as the training JKTs (Allen, Cheng, Ingerick, & Caramagno, 2009). One overall score was computed for each test corresponding to the six target MOS and another overall score was computed for the Warrior Tasks and Battle Drills (WTBD) test. Poorly performing items were eliminated using diagnostic analyses such as item-total correlations. As with the training JKTs, the final set of items for each test was used to compute overall scores in two ways: (a) a percent correct score, computed by dividing the number of points the Soldier received by the total number of points possible on the test and (b) a raw total score, computed by summing the total number of points Soldiers earned across all of the retained items. A standardized total score was also computed for the in-unit MOS-specific JKTs by taking the *z*-score of the raw total score within each MOS.

The same MOS-specific and WTBD JKTs were used for both the in-unit 1 and in-unit 2 data collections. They were also scored the exact same way. The descriptive statistics for these tests can be found in Table 5.1. Both the in-unit 1 and in-unit 2 JKTs exhibited good internal consistency reliability for research purposes, despite low sample sizes in many cases. Relatively low reliability estimates were associated with the WTBD JKT and the in-unit 1 68W JKT. The average mean percent correct (M = 68.3%) was a bit higher than what was observed with the training JKTs. The mean percent correct scores for the MOS-specific JKTs were generally higher for the in-unit 2 sample than the in-unit 1 sample, which is consistent with the additional maturity of that sample. The exception to this was the MOS-specific JKT scores for 91B.

Table 5.1. Descriptive Statistics and Reliability Estimates for In-Unit 1 and In-Unit 2 Job Knowledge Tests

JKT Type	*n*	Min	Max	Max Possible	*M*	*SD*	Mean Percent Correct	*α*
In-Unit 1 Job Knowledge Tests								
11B – Infantryman	246	20	62	71	46.13	8.82	65.0	.82
19K – Armor Crewman	83	18	51	57	37.80	8.15	66.3	.81
31B – Military Police	168	36	92	107	71.43	11.23	66.8	.81
68W – Health Care Specialist	34	30	46	53	39.47	4.22	74.5	.61
88M – Motor Transport Operator	47	38	80	94	62.04	11.15	66.0	.87
91B – Light Wheel Vehicle Mechanic	50	21	49	54	34.54	6.97	64.0	.79
WTBD	1,374	4	26	26	18.53	3.56	72.2	.65
In-Unit 2 Job Knowledge Tests								
11B – Infantryman	190	18	63	71	47.04	9.91	66.3	.86
19K – Armor Crewman	62	17	53	57	38.53	7.82	73.3	.87
31B – Military Police	108	40	90	107	73.13	10.98	68.3	.83
68W – Health Care Specialist	18	35	51	53	40.56	4.09	76.5	--
88M – Motor Transport Operator	22	42	81	94	63.09	11.11	67.1	.85
91B – Light Wheel Vehicle Mechanic	30	23	44	54	33.57	5.81	58.9	.76
WTBD	928	2	26	26	18.50	3.70	71.2	.68

Note. Max Possible = Maximum possible score on JKT; Mean Percent Correct = Average percent correct received on JKT [M / Max Possible]; α = coefficient alpha, which was not computed for 68W in the in-unit 2 sample due to low sample size. The mean percent correct provides information about the characteristics of the test and does not indicate the readiness or skill of those tested.

In-Unit Performance Rating Scales (PRS)

The in-unit performance rating scales (PRS) consisted of several behaviorally anchored scales (BARS). The rating options ranged from 1 (low performance) to 7 (high performance). Raters also had the option of checking "cannot rate" when they had not observed the Soldier on the targeted behaviors. The scales were scored by first dropping ratings from supervisors who were missing more than 10% of their ratings or had selected "cannot rate" for more than 50% of the scales. In the rare cases where there was more than one rater for a particular Soldier, the average was taken of the two supervisors' ratings.

The same MOS-Specific and AW PRS were implemented in both the in-unit 1 and in-unit 2 data collections. The AW PRS consisted of 14 BARS, while the MOS-specific PRS consisted of four to nine BARS, depending on the target MOS. In addition, the Combat/Deployment Performance Rating Scales (CDPRS), which included five scales, were administered with the in-unit 2 rating scales. The AW scales were combined into three unit-weighted composites based on previous research (Campbell, Hanson, & Oppler, 2001): Cognitive Performance PRS (a composite of the Processing Information and Solving Problems scales), Effort and Discipline PRS (a composite of the Exhibiting Effort, Exhibiting Personal Discipline, Managing Personal Matters, and Following Safety Procedures scales), and Working Effectively with Others PRS (a composite of the Communicating with Others, Contributing to the Team, and Leadership Potential scales). Similar to the AW PRS composites, the in-unit MOS-specific PRS and the CDPRS were scored by taking a unit-weighted average of the individual component scales.

The descriptive statistics for the AW PRS composites can be found in Table 5.2. Both the in-unit 1 and in-unit 2 AW PRS exhibited acceptable internal consistency reliabilities, with the Working Effectively with Others PRS somewhat lower than the Cognitive Performance and Effort and Discipline composites. The means for the in-unit 2 composites were slightly higher than the means for the in-unit 1 composites, consistent with the additional maturity of the in-unit 2 sample. However, the composite variances for the two in-unit populations were comparable, suggesting both have utility as criterion measures.

Table 5.2. Descriptive Statistics and Reliability Estimates for Composite Performance Rating Scales

Army Wide Performance Rating Scale (AW PRS) Composite	*n*	Min	Max	*M*	*SD*	*α*
In-Unit 1 AW PRS Composites						
Cognitive Performance PRS	914	1.00	7.00	4.89	1.30	.82
Effort and Discipline PRS	914	1.00	7.00	5.20	1.24	.84
Working Effectively with Others PRS	914	1.00	7.00	4.99	1.30	.79
In-Unit 2 AW PRS Composites						
Cognitive Performance PRS	653	1.00	7.00	5.24	1.23	.85
Effort and Discipline PRS	654	1.00	7.00	5.46	1.19	.86
Working Effectively with Others PRS	654	1.00	7.00	5.26	1.22	.79

Note. *α* = coefficient alpha. Max possible for all AW PRS composites = 7.0.

Table 5.3 displays the AW and CDPRS scale-level descriptive statistics for both the in-unit 1 and in-unit 2 PRS by MOS. The in-unit 1 PRS mean scores suggest general elevation in the ratings, with Leadership Potential in the total sample yielding the lowest mean score (*M* = 4.64) and Interactions with Indigenous People and Soldiers yielding the highest mean score (*M* = 5.71). There was enough variance in the in-unit 1 PRS (*SD* = 1.08 – 1.70) for research purposes. The in-unit 2 PRS means suggest a similar pattern. Leadership Potential had the lowest average mean score across MOS (*M* = 4.87), while Interactions with Indigenous People and Soldiers yielded the highest average score (*M* = 5.85). The internal consistency estimates for the MOS-specific and CDPRS composite scores in both samples were comparably high (alpha = .90 – .95).

Table 5.3. Descriptive Statistics for In-Unit 1 and In-Unit 2 Performance Rating Scales (PRS)

	In-Unit 1[a]		In-Unit 2[b]	
Composite/Scale	*M*	*SD*	*M*	*SD*
AW PRS				
Performing Core Warrior Tasks	4.99	1.35	5.27	1.33
Performing MOS-Specific Tasks	4.96	1.36	5.22	1.37
Communicating with Others	4.94	1.45	5.17	1.39
Processing Information	4.95	1.43	5.23	1.38
Solving Problems	4.83	1.38	5.12	1.36
Exhibiting Effort	4.99	1.51	5.17	1.46
Exhibiting Personal Discipline	5.18	1.57	5.35	1.52
Contributing to the Team	5.40	1.44	5.6	1.35
Exhibiting Fitness and Bearing	4.94	1.66	5.14	1.67

Table 5.3. (Continued)

Composite/Scale	In-Unit 1[a]		In-Unit 2[b]	
	M	*SD*	*M*	*SD*
Interactions with Indigenous People and Soldiers	5.71	1.21	5.85	1.19
Following Safety Procedures	5.37	1.23	5.62	1.26
Developing Own Skills	4.93	1.36	5.11	1.38
Managing Personal Matters	5.29	1.60	5.48	1.54
Leadership Potential	4.64	1.70	4.87	1.71
MOS-Specific PRS Composite[c]	5.20	1.08	5.58	1.10
Combat/Deployment PRS[d]				
Field/Combat Judgment			5.36	1.34
Field Readiness			5.76	1.27
Physical Endurance			5.40	1.41
Physical Courage			5.48	1.29
Awareness and Vigilance			5.49	1.29
CDPRS Composite			5.48	1.14

[a] Overall AW PRS *n* = 874-910; Overall MOS-specific PRS Composite *n* = 435. The AW PRS and the MOS-specific PRS Composite range from 1 – 7.

[b] Overall AW PRS *n* = 714-739; Overall MOS-specific PRS Composite *n* = 349. Overall Combat PRS *n* = 319-329. Scores range from 1 – 7.

[c] Coefficient alpha for the total MOS-specific PRS Composite for in-unit 1 is .93, reflecting a sample-weighted average of the estimates for the individual MOS (11B = .91, 19K = .93, 31B = .94, 68W = .93, 88M = .95, 91B = .95). Coefficient alpha for the total MOS-specific PRS Composite for in-unit 2 is .94, reflecting a sample-weighted average of the estimates for the individual MOS (11B = .95, 19K = .95, 31B = .93, 68W = .90, 88M = .94, 91B = .95). Coefficient alpha for the CDPRS Composite is .90 across the entire sample.

[d] Combat/Deployment PRS was not administered to the In-Unit 1 sample Soldiers.

In-Unit Army Life Questionnaire (ALQ)

As with the JKTs, the same ALQ was administered in both the in-unit 1 and in-unit 2 samples. Most of the in-unit ALQ scales were scored by taking the average of various items that range from 1 to 5 on a Likert scale. The exceptions were (a) the *Deployment Tempo* scale, representing Soldiers' self-reported number of months deployed in their current term of service; (b) the *Army Physical Fitness Test (APFT)*, representing Soldiers' self-reported last APFT score; (c) the *Weapons Qualification* score, representing Soldiers' self-reported last weapons qualification score; (d) the *Disciplinary Actions* scale, representing the sum of in-unit ALQ items related to Soldiers' self-reported disciplinary incidents; and (e) the *Qualifications and Awards* scale, representing the sum of in-unit ALQ items related to Soldiers' self-reported career achievements. While the items were administered with both versions of the ALQ, the *Promotion Points* scale, representing Soldiers' self-reported awards that contribute to their enlisted promotion packet score,[5] was only scored in the in-unit 2 sample due to irregularities in the response patterns in the in-unit 1 sample.[6] Descriptive statistics for the in-unit ALQ are reported

[5] See Section 3-43 of Army Regulation 600-8-19 ("Enlisted Promotions and Reductions") for more details.

[6] "Irregularities" in this case refer to instances where the self-reported rate of medal awards in the in-unit 1 sample was much higher than the rate of the awards for the same medal in the Army as a whole. This led us to conclude that some of the Soldiers may not fully understand these medals and hence erroneously indicated they received them.

in Table 5.4. As with the training ALQ, the internal consistency estimates for these scales were generally high (alpha = .71 – .94).

Table 5.4. Descriptive Statistics and Reliability Estimates for In-Unit 1 and In-Unit 2 Army Life Questionnaire (ALQ) Scale Scores

	In-Unit 1				In-Unit 2			
Composite/Scale	*n*	*M*	*SD*	*α*	*n*	*M*	*SD*	*α*
Deployment								
Deployment Tempo[a]	402	8.49	3.85	n/a	780	10.97	2.71	n/a
Deployment Adjustment[a]	405	3.67	0.79	.77	781	3.62	0.79	.77
Performance								
Promotion Points[a]	--	--	--	--	944	24.59	16.64	--
Disciplinary Incidents[b]	1,409	0.57	1.13	.71	--	--	--	--
Quals and Awards[b]	1,399	0.59	0.82	n/a	941	0.98	1.02	n/a
Last APFT Score[b]	1,314	242.64	38.59	n/a	925	249.56	34.11	n/a
Last Weapon Qual. Score[a]	1,401	2.92	0.81	n/a	943	3.20	0.80	n/a
Attitudinal								
Affective Commitment[b]	1,409	3.58	0.85	.90	944	3.28	0.90	.91
Army Fit[b]	1,409	3.88	0.76	.83	944	3.66	0.76	.81
Attrition Cognitions[b]	1,409	1.69	0.79	.79	944	1.83	0.82	.79
Career Intentions[b]	1,409	2.67	1.25	.93	944	2.41	1.24	.93
MOS Fit[b]	1,409	3.28	0.98	.93	944	3.24	0.95	.93
MOS Satisfaction[a]	1,409	3.46	0.98	.94	944	3.36	0.94	.93
Reenlistment Intentions[a]	1,409	3.04	1.20	.81	944	2.75	1.22	.82

α = coefficient alpha. n/a = single-item measure. ALQ scale scores range from 1 – 5 except for the following: (a) Disciplinary Action (0 – 1; not administered in In-Unit 2), (b) Last APFT Score (free response item, Min = 62, Max = 300), (c) Last Weapon Qualification Score (1 – 4), (d) Qualifications and Awards (0 – 3), and (e) Deployment Tempo (free response item, Min = 1, Max = 15), and (f) Promotion Points (0 – 100; administered but not computed in In-Unit 1)

[a] Scales that were added to the ALQ for the in-unit versions.

[b] Scales that were retained from the training ALQ.

Attrition

For the purposes of this research, attrition is a broad category that includes separations because of underage enlistment, conduct, family concerns, sexual orientation, drugs/alcohol, performance, physical standards/weight, mental disorder, or violations of the Uniform Code of Military Justice. Attrition was computed at 3 months (attrition near or after the completion of Basic Combat Training), 4 months (attrition during AIT/OSUT), 6 months (attrition near or after completion of AIT/OSUT), and at regular 3-month intervals thereafter. Data were extracted in the current sample out to 42 months in service, though, due to insufficient time in service, attrition data for the complete sample were only available out to 36 months. As described in Chapter 4, the data used to compute this variable came from the TTAS database. USAR and ARNG Soldiers were excluded from the attrition analysis because reliable data were not available in the TTAS database for those samples. Attrition rates for key populations of interest are reported in Chapter 7

CHAPTER 6: PREDICTING IN-UNIT SOLDIER PERFORMANCE

Joseph P. Caramagno, Matthew T. Allen, and Michael J. Ingerick (HumRRO)

This chapter describes the analyses examining the potential of the experimental predictors to predict Soldiers' in-unit performance beyond the AFQT. We begin with a short summary of relevant findings from previous research followed by a description of the analytic procedures and summary of the results of our analyses.

Background

Army Class builds on a long history of research on ways to enhance new Soldier selection, from Project A (Campbell, McHenry, & Wise, 1990) to the more recent *New Predictors for Selecting and Assigning Future Force Soldiers* (Select21; Knapp et al., 2005). Three previous research efforts examined the same or similar experimental predictor measures as included in Army Class. This previous research includes two concurrent validations with samples of incumbent first-term Soldiers (Ingerick et al., 2009; Knapp & Tremble, 2007) and an analysis of criterion data collected for Army Class while Soldier participants were enrolled in training (Knapp & Heffner, 2009). Results of this research led to the following conclusions:

- The predictive validity of the AFQT for predicting technical or "can-do" performance is high, with uncorrected validity coefficients (*R*) typically greater than .40. However, the AFQT is less predictive of behaviorally-based or will-do performance criteria such as commitment and leadership.

- Several of the experimental measures (especially the RBI, TAPAS, AIM, and WPA) demonstrated potential to predict behaviorally-based or "will-do" performance criteria, with incremental validity estimates (*R*) typically ranging between .05 and .25 depending on the outcome measure.

The present analyses expand on this prior research by examining the potential of the experimental measures to predict Soldier performance in the Army Class longitudinal sample after they have joined their units. Data were collected at two points in time, the first when the Soldiers had an average of 20 months TIS (in-unit 1) and the second when the Soldiers had an average of 35 months TIS (in-unit 2).

Incremental Validity Analysis

Approach

Criterion Measures

The incremental validity analyses were conducted on seven individual criterion scores and four composite scores (described in greater detail in Chapter 5). These criteria were selected because (a) as a group, they provide comprehensive coverage of the performance domain and (b) sufficient data were available on them across both in-unit samples. The criterion measures represent two higher-order dimensions of performance: can-do and will-do (Campbell, Hanson, & Oppler, 2001; Campbell, McHenry, & Wise, 1990). These dimensions can be further

delineated into the lower-order performance constructs, summarized below, along with their constituent measures.

Can-Do Performance Dimensions

1. *Core Technical Proficiency* – Core Technical Proficiency represents the extent to which Soldiers perform the tasks that are essential to their MOS. This dimension was assessed using (a) the MOS-specific JKT and (b) the Army-Wide (AW) Performing MOS-Specific Tasks PRS.

2. *General Soldiering Proficiency* – This dimension represents the extent to which Soldiers effectively perform tasks that are important to all Soldiers. This dimension was assessed using (a) the WTBD JKT and (b) the Cognitive Performance AW PRS.

Will-Do Performance Dimensions

3. *Achievement and Leadership* – This dimension reflects the extent to which the Soldier perseveres in the face of adversity and supports other Soldiers. Achievement and Leadership was measured using (a) the Working Effectively with Others AW PRS and (b) one self-reported ALQ measure, the quantity and type of qualifications and awards the Soldier received.[7]

4. *Effort and Personal Discipline* – Effort and Personal Discipline reflects the extent to which Soldiers demonstrate commitment and discipline. This dimension was assessed using (a) the Effort and Discipline AW PRS and (b) the number of disciplinary incidents the Soldier had during IMT and in-unit, as self-reported on the ALQ.[8]

5. *Physical Fitness and Military Bearing* – This dimension represents the extent to which a Soldier maintains an appropriate Army appearance and good physical condition. It was measured using (a) the Physical Fitness AW PRS and (b) the Soldiers' most recent APFT score, as self-reported on the ALQ.

We also examined a sixth performance dimension called Deployment Adjustment and Performance. This dimension was assessed using (a) the Combat/Deployment Performance Rating Scales (CDPRS; see Chapter 2 and Appendix A for further information on this measure) and (b) a self-report assessment of Deployment Adjustment, administered as part of the ALQ. We did not attempt to integrate this dimension into the can-do and will-do components primarily because CDPRS data were only collected from a small proportion of the in-unit 2 sample. Therefore, the validation analyses associated with this performance dimension were treated separately from the rest.

[7] Qualifications and Awards (ALQ) was assessed for the in-unit 2 sample only.
[8] Disciplinary Incidents (ALQ) was assessed for the in-unit 1 sample only.

Procedure

To identify the measures with the greatest potential to supplement the AFQT in predicting Soldier performance for each of the above criteria, we estimated the incremental validity of the experimental predictor measures over AFQT.[9] In brief, this approach involved testing a series of hierarchical regression models to estimate the observed (uncorrected) multiple correlation (*R*) for the full battery of predictors (i.e., AFQT and the experimental measures), regressing each criterion measure onto Soldiers' AFQT scores in the first step, followed by their scale-level scores for each experimental predictor in the second step. The resulting increment in the multiple correlation (ΔR) when the predictor scale scores were added to the baseline regression models served as our index of incremental validity.

The full set of scale scores for the given experimental predictor measures were used when estimating each of these models. For example, the 14 scales that comprise the RBI were included as separate scores in all models that feature the RBI. We used all of the available scales for each measure to determine the predictive *potential* of each measure as a whole. None of the experimental predictor scores consisted of composite scores that had been optimally weighted or empirically keyed to a criterion.

Two issues should be noted that carry implications for interpreting the results of these analyses. First, the power to detect a significant effect was low for some predictor-criterion combinations due to small sample sizes and a relatively large number of component scales for many of the predictor measures (e.g., RBI, TAPAS, and WPA facets). Second, the results may be attenuated due to range restriction in the in-unit criterion measures. Soldiers in the in-unit sample have necessarily performed well enough at earlier phases of their career (e.g., during IMT) to remain in service, while low-performing Soldiers are more likely to have attrited from the Army. A more detailed examination of attrition over time is reported in Chapter 7.

Finally, sample-specific error could potentially inflate the estimates of *R* for predictor measures with small sample sizes and many scales. As a result, variations in sample sizes and the number of scales constituting each predictor measure make cross-measure comparisons difficult. To address this issue, we adjusted the observed incremental validity estimates using Burket's (1964) formula for shrinkage (cf. Formula 8; Schmitt & Ployhart, 1999). Calculating the corrected incremental validity estimates involved two additional steps:

1. Using the observed (uncorrected) correlations among the new predictor, AFQT, and the selected criterion previously estimated, adjust the correlations between the predictors and the performance-related criteria for sample size and number of predictors using Burket's (1964) formula for shrinkage:

$$\rho_c = (NR^2 - k)/[R(N - k)] \qquad (1)$$

where *k* equals the total number of predictor scale scores in the model.

[9] Readers that are interested in the scale-level correlations should refer to Appendix C.

2. Calculate the corrected incremental validity estimates for the experimental predictors by subtracting the shrunken *R* (the corrected *R* from the equation above) associated with an AFQT-only model from the shrunken *R* obtained from the full model (i.e., AFQT + Experimental Predictor model).

As an aside, instances where there are dramatic differences between the uncorrected and corrected regression coefficients beg the question of which to attend to in interpreting the results. To the extent that the corrected incremental validity estimates are similar to the uncorrected, the more confidence we have in the uncorrected estimates. However, in instances where the incremental validity estimates reduce to nearly zero or negative, this suggests that we cannot rule out measurement error as an explanation for the uncorrected coefficients. This is not to say that the experimental measure does not have any utility for predicting that criterion, but additional caution is necessary in interpreting the results. In most cases, we focus our interpretation of the magnitude and statistical significance of the uncorrected coefficients, but note the uncertainty suggested in the corrected coefficients.

Findings

Results of the incremental validity analyses are summarized in Tables 6.1 through 6.5. For all analyses, we first discuss the uncorrected estimates followed by the results based on the corrected (or shrinkage adjusted) estimates.

Can-Do Performance-Related Criteria

In-Unit 1. As expected, the AFQT performed quite well for predicting knowledge-based outcomes like the MOS-specific JKT. AFQT showed strong potential for predicting MOS-specific and WTBD job knowledge-based performance criteria (average R = .37 and .51, respectively) and moderate to low potential for predicting a composite rating of Soldiers' information processing and problem solving abilities (Cognitive Performance AW PRS) (average R = .15) and MOS-specific task performance (average R = .11). Several of the experimental predictors (i.e., RBI, AKA, AO, PSJT, and WPA [dimensions and facets]) exhibited significant incremental validity over the AFQT in predicting at least one can-do performance criterion. Among the experimental predictors, AO showed significant, although small, incremental validity across all can-do performance-related criteria (ΔRs = .01-.04). Like AO, the strengths of the validity coefficients associated with the other experimental predictors were relatively small, with none greater than .10. In terms of magnitude, the RBI and TAPAS yielded the largest validity coefficients, with the average ΔR around .06 for each.

The greatest number of experimental measures showed incremental validity in predicting WTBD JKT scores. WPA (ΔR = .04-.05), RBI (ΔR = .03), PSJT (ΔR = .02), and AO (ΔR = .01) significantly predicted this criterion over AFQT. Overall, the fewest number of experimental measures showed significant incremental validity in predicting ratings of Soldiers' performance on MOS-specific tasks (only AO showed significant incremental validity, ΔR = .04).

Table 6.1. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 1 Can-Do Performance

		Uncorrected			Corrected		
Predictor/Scale	*N*	AFQT Only	AFQT + Predictor	Δ*R*	AFQT Only	AFQT + Predictor	Δ*R*
MOS-Specific Job Knowledge Test (JKT)							
AO [1]	583	**.36**	**.39**	**.03**	.35	.38	.03
AIM [6]	220	**.42**	**.43**	.01	.41	.36	-.05
TAPAS [12]	222	**.40**	**.44**	.05	.39	.33	-.06
PSJT [1]	373	**.33**	**.36**	**.03**	.32	.34	.02
RBI [14]	520	**.35**	**.40**	.05	.34	.33	-.01
AKA [6]	587	**.36**	**.40**	**.04**	.36	.37	.01
WPA Dimensions [6]	594	**.36**	**.38**	.02	.36	.36	.00
WPA Facets [14]	594	**.36**	**.39**	.03	.36	.34	-.02
Warrior Tasks and Battle Drills (WTBD) JKT							
AO [1]	1,269	**.52**	**.53**	**.01**	.51	.52	.01
AIM [6]	528	**.55**	**.56**	.01	.55	.54	-.01
TAPAS [12]	526	**.55**	**.57**	.02	.55	.54	-.01
PSJT [1]	702	**.46**	**.48**	**.02**	.46	.48	.02
RBI [14]	1,106	**.50**	**.53**	**.03**	.50	.51	.01
AKA [6]	1,270	**.51**	**.52**	.01	.51	.51	.00
WPA Dimensions [6]	1,269	**.51**	**.55**	**.04**	.51	.54	.03
WPA Facets [14]	1,268	**.51**	**.56**	**.05**	.51	.54	.03
Performing MOS-Specific Tasks (AW PRS)							
AO [1]	823	**.12**	**.16**	**.04**	.11	.15	.03
AIM [6]	371	.10	.17	.07	.07	.06	-.02
TAPAS [12]	378	**.10**	.17	.06	.08	.00	-.08
PSJT [1]	427	**.11**	**.13**	.02	.09	.10	.00
RBI [14]	703	**.08**	.18	.10	.07	.07	.00
AKA [6]	818	**.11**	**.16**	.04	.10	.10	.00
WPA Dimensions [6]	821	**.12**	**.16**	.04	.11	.10	-.01
WPA Facets [14]	820	**.12**	**.18**	.06	.11	.08	-.03
Cognitive Performance (AW PRS)							
AO [1]	842	**.17**	**.20**	**.03**	.16	.19	.03
AIM [6]	380	**.16**	**.22**	.06	.14	.15	.00
TAPAS [12]	388	**.13**	.23	.09	.11	.08	-.03
PSJT [1]	437	**.12**	**.18**	**.06**	.10	.16	.05
RBI [14]	722	**.13**	**.22**	.09	.11	.13	.01
AKA [6]	837	**.17**	**.19**	.03	.16	.15	-.01
WPA Dimensions [6]	841	**.17**	**.19**	.02	.16	.14	-.02
WPA Facets [14]	840	**.17**	**.20**	.03	.16	.11	-.05

Note. AFQT = Armed Forces Qualification Test. AFQT + Predictor = Multiple correlation (*R*) between AFQT and selected predictor measure with the criterion. Δ*R* = Increment in *R* over AFQT from adding the selected predictor measure to the regression model ((AFQT + Predictor) – (AFQT Only)). Estimates in bold are statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Listwise deletion was used to account for missing data. Estimates in the "Corrected" columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$, while estimates in the "Uncorrected" columns were not adjusted. Negative corrected coefficients for AFQT Only and AFQT + Predictor were set to .00; however, the corrected Δ*R*s were allowed to reduce to less than .00.

Adjustments made for sample size and number of predictors reduced the magnitude of the observed multiple correlation estimates considerably for all predictor measures. Estimates for the three predictors exhibiting the largest average gains in incremental validity (i.e., RBI, WPA facets, and AIM) dropped to negative or near-zero values.[10] Consequently, the utility of these measures to supplement the AFQT in predicting Soldier can-do performance was no longer evident. Predictor measures with fewer scale scores (i.e., AO and PSJT) were affected the least by the formula-based adjustments and continued to exhibit small gains in prediction over AFQT (AO average ΔR = .03; ΔR = .01-.04; PSJT average ΔR = .03; ΔR = .02-.06). The corrected estimates indicate that the experimental predictors have limited utility for incrementing the prediction of can-do performance-related criteria over AFQT for Soldiers with time in service between 12 and 24 months. While several measures initially appeared to enhance the predictive utility of AFQT, their estimates were nearly zero after adjusting for shrinkage, suggesting that measurement error cannot be ruled out as an explanation for the uncorrected coefficients.

In-Unit 2. Nearly identical analyses were performed on criterion and predictor data collected from Soldiers that had been in the Army for an average of about 3 years (see Table 6.2). Sample sizes decreased between time 1 and time 2 by an average of 26%. The largest proportionate decrease was found for the analyses involving PSJT where sample size decreased by 44% for the analysis involving the MOS-specific JKT and 40% for analyses involving the WTBD JKT. Incremental validity estimates also decreased, though the general pattern of results was relatively stable. AFQT remained a strong predictor of core technical performance on tests of MOS-specific and WTBD job knowledge and composite ratings of Soldiers' cognitive performance. More pronounced change occurred for the correlation between AFQT and Performing MOS-Specific Tasks (composite) where AFQT no longer demonstrated a statistically significant relationship with the criterion.

As expected, the experimental predictors' contribution to predictive validity was limited (average ΔR = .05, ΔR = .00-.15). The PSJT significantly enhanced the prediction of MOS-specific JKT scores (ΔR = .08) and AO, PSJT, RBI, and WPA provided small gains in predictive validity for WTBD JKT scores (average ΔR = .04, ΔR = .00-.05). The largest increase in incremental validity was found for the composite variable Performing MOS-Specific Tasks AW PRS (average ΔR = .07, ΔR = .00-.15), however none of the estimates of the change in *R* were significant. Across the in-unit 2 criteria, the RBI (average ΔR = .07) and TAPAS (average ΔR = .08) again yielded the largest average incremental validity coefficients. Correcting for shrinkage all but eliminated the experimental measures' potential contributions to predicting core technical and general soldiering proficiency of Soldiers with about 3 years in service. Many of the incremental validity estimates dropped to zero or near-zero levels, and the others became negative. Thus, consistent with theoretical and empirical findings, the AFQT remains the strongest predictor of can-do performance throughout a Soldier's first term of service.

[10] Negative values indicate that the shrinkage-adjusted full model regression coefficient (i.e., AFQT + Predictor) is smaller in magnitude than the shrinkage-adjusted AFQT Only model regression coefficient. This can happen with Burket's (1964) formula when the sample size is sufficiently small and the number of scales contributing to the model is sufficiently large.

Table 6.2. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 2 Can-Do Performance

		Uncorrected			Corrected		
Predictor/Scale	*n*	AFQT Only	AFQT + Predictor	Δ*R*	AFQT Only	AFQT + Predictor	Δ*R*
MOS-Specific Job Knowledge Test (JKT)							
AO [1]	402	**.33**	**.33**	.00	.32	.31	-.01
AIM [6]	187	**.25**	**.30**	.05	.23	.19	-.04
TAPAS [12]	199	**.29**	**.41**	.12	.27	.27	.00
PSJT [1]	208	**.39**	**.47**	**.08**	.38	.46	.07
RBI [14]	347	**.31**	**.40**	.08	.31	.30	.00
AKA [6]	404	**.31**	**.33**	.02	.30	.28	-.02
WPA Dimensions [6]	408	**.33**	**.37**	.04	.33	.33	.00
WPA Facets [14]	408	**.33**	**.40**	.06	.33	.31	-.01
Warrior Tasks and Battle Drills (WTBD) JKT							
AO [1]	858	**.42**	**.43**	**.00**	.42	.42	.00
AIM [6]	408	**.46**	**.47**	.01	.45	.44	-.01
TAPAS [12]	416	**.49**	**.51**	.03	.48	.47	-.02
PSJT [1]	423	**.42**	**.44**	**.03**	.41	.43	.02
RBI [14]	728	**.44**	**.48**	**.04**	.44	.44	.01
AKA [6]	849	**.42**	**.43**	.01	.41	.41	.00
WPA Dimensions [6]	856	**.42**	**.46**	**.04**	.42	.45	.03
WPA Facets [14]	856	**.42**	**.48**	**.05**	.42	.45	.03
Performing MOS-Specific Tasks (AW PRS)							
AO [1]	664	**.08**	.08	.00	.06	.04	-.01
AIM [6]	310	.06	.17	.11	.00	.04	.04
TAPAS [12]	319	.08	.21	.13	.04	.01	-.03
PSJT [1]	331	.05	.06	.01	.00	.00	.00
RBI [14]	550	.06	.20	.15	.02	.07	.05
AKA [6]	655	.07	.10	.03	.04	.00	-.04
WPA Dimensions [6]	657	.07	.11	.04	.05	.01	-.04
WPA Facets [14]	657	.07	.18	.11	.05	.05	.00
Cognitive Performance (AW PRS)							
AO [1]	683	**.10**	**.12**	.02	.09	.10	.01
AIM [6]	324	.08	.16	.08	.04	.02	-.02
TAPAS [12]	332	**.14**	.24	.10	.11	.08	-.03
PSJT [1]	338	.08	.10	.02	.05	.05	.00
RBI [14]	567	**.08**	.18	.09	.06	.03	-.03
AKA [6]	678	**.11**	.13	.01	.10	.04	-.06
WPA Dimensions [6]	679	**.12**	**.15**	.03	.10	.08	-.02
WPA Facets [14]	679	**.12**	**.20**	.08	.10	.09	-.02

Note. AFQT = Armed Forces Qualification Test. AFQT + Predictor = Multiple correlation (*R*) between AFQT and selected predictor measure with the criterion. Δ*R* = Increment in *R* over AFQT from adding the selected predictor measure to the regression model ((AFQT + Predictor) – (AFQT Only)). Estimates in bold are statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Listwise deletion was used to account for missing data. Estimates in the "Corrected" columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$, while estimates in the "Uncorrected" columns were not adjusted. Negative corrected coefficients for AFQT Only and AFQT + Predictor were set to .00; however, the corrected Δ*R*s were allowed to reduce to less than .00.

Will-Do Performance-Related Criteria

In-Unit 1. Consistent with previous research (e.g., Select21; Knapp & Tremble, 2007), AFQT did not predict will-do performance criteria as well as it predicted can-do performance criteria (average *R* = .07; Table 6.3). Accordingly, the experimental predictor measures consistently evidenced incremental validity in predicting the will-do criteria. The three temperament measures—RBI (Δ*R* = .10-33), TAPAS (Δ*R* = .10-.26), and AIM (Δ*R* = .08-.25)—demonstrated the largest estimates of incremental validity over AFQT. Due to small sample sizes however, the coefficients for the AIM and TAPAS often failed to achieve statistical significance. Among these three measures, RBI demonstrated the most potential for predicting in-unit will-do performance beyond AFQT, with the largest observed incremental validity estimates among the predictors for three out of four criteria. As a group, the measures best predicted Soldiers' self-reported APFT scores, as RBI, TAPAS, AIM, and WPA (facets) each exhibited uncorrected validity coefficients ranging from .18 to .33.

Although the estimates for the behaviorally-based performance-related criteria were generally larger than those associated with the knowledge-based criteria, they should be interpreted with caution. Adjusting for sample size and number of predictors decreased the observed estimates considerably (AFQT average corrected *R* dropped from .07 to .05; AFQT + experimental predictor average corrected *R* dropped from .16 to .09). This was particularly true for the TAPAS and WPA, where the incremental validity results for several criteria became negative, suggesting measurement error may partially explain the uncorrected coefficients.[11]

The corrected estimates for the RBI (average corrected Δ*R* across in-unit 1 will-do criteria = .14; Δ*R* range = .01-.32) demonstrated the greatest potential to increment prediction of two of the will-do performance-related criteria (i.e., APFT score and number of disciplinary incidents) compared to the other predictor measures. After correction, however, the RBI no longer consistently exhibited the highest increment in predictive validity over AFQT. AO emerged as a stronger candidate for predicting Soldiers' ability to work with others (*R* = .17) and composite ratings for Effort and Discipline AW PRS (*R* = .14). AO (average corrected Δ*R* = .04; corrected Δ*R* = .02-.06) and PSJT (average corrected Δ*R* = .07; corrected Δ*R* = .00-.16) continued to exhibit limited incremental validity over AFQT for predicting Soldiers' physical fitness test scores and number of disciplinary incidents. Validity coefficients associated with AO and PSJT were minimally affected by correcting for shrinkage because the measures consist of a single scale score. The AKA failed to demonstrate appreciable gains in prediction over AFQT for the will-do performance criteria after correcting for shrinkage.

[11] Negative corrected coefficients for AFQT Only and AFQT + Predictor were set to .00; however, the corrected Δ*R*s were allowed to reduce to less than .00.

Table 6.3. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 1 Will-Do Performance

		Uncorrected			Corrected		
Predictor/Scale	*n*	AFQT Only	AFQT + Predictor	Δ*R*	AFQT Only	AFQT + Predictor	Δ*R*
Effort and Discipline (Army-Wide [AW] Performance Rating Scales [PRS])							
AO [1]	842	**.09**	**.16**	**.06**	.08	.14	.06
AIM [6]	380	.08	.18	.10	.04	.08	.04
TAPAS [12]	388	.06	.18	.11	.02	.00	-.02
PSJT [1]	437	.06	**.14**	**.07**	.03	.11	.08
RBI [14]	722	.04	**.20**	**.15**	.01	.09	.08
AKA [6]	837	**.10**	**.14**	.04	.09	.08	.00
WPA Dimensions [6]	841	**.10**	.11	.02	.08	.04	-.05
WPA Facets [14]	840	**.10**	.14	.05	.09	.02	-.07
Working Effectively with Others (AW PRS)							
AO [1]	842	**.14**	**.19**	**.05**	.13	.17	.04
AIM [6]	380	.09	.17	.08	.06	.06	.00
TAPAS [12]	388	.09	.19	.10	.07	.02	-.04
PSJT [1]	437	**.14**	**.19**	**.05**	.13	.16	.04
RBI [14]	722	**.10**	**.20**	.10	.09	.10	.01
AKA [6]	837	**.14**	**.18**	.04	.13	.13	.00
WPA Dimensions [6]	841	**.14**	**.15**	.01	.13	.10	-.03
WPA Facets [14]	840	**.14**	**.18**	.04	.13	.08	-.05
Last Army Physical Fitness Test (APFT) Score (ALQ)							
AO [1]	1,217	.03	.06	.03	.01	.04	.03
AIM [6]	513	.07	**.32**	**.25**	.04	.28	.24
TAPAS [12]	503	.06	**.32**	**.26**	.03	.24	.21
PSJT [1]	664	.02	.04	.02	.00	.00	.00
RBI [14]	1,067	.02	**.35**	**.33**	.00	.32	.32
AKA [6]	1,216	.04	.09	.04	.02	.02	.00
WPA Dimensions [6]	1,214	.05	**.12**	**.08**	.03	.08	.05
WPA Facets [14]	1,214	.05	**.22**	**.18**	.03	.17	.14
Disciplinary Incidents (ALQ)							
AO [1]	1,302	.04	.07	.03	.02	.04	.02
AIM [6]	558	.06	**.16**	.10	.04	.08	.05
TAPAS [12]	550	.06	.19	.13	.03	.06	.04
PSJT [1]	703	.03	.06	.03	.00	.02	.02
RBI [14]	1,137	.05	**.23**	**.18**	.03	.17	.14
AKA [6]	1,303	.05	.10	.05	.03	.04	.01
WPA Dimensions [6]	1,300	.05	.07	.03	.03	.00	-.03
WPA Facets [14]	1,299	.05	.09	.05	.03	.00	-.03

Note. AFQT = Armed Forces Qualification Test. AFQT + Predictor = Multiple correlation (*R*) between AFQT and selected predictor measure with the criterion. Δ*R* = Increment in *R* over AFQT from adding the selected predictor measure to the regression model ((AFQT + Predictor) – (AFQT Only)). Estimates in bold are statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Listwise deletion was used to account for missing data. Estimates in the "Corrected" columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$, while estimates in the "Uncorrected" columns were not adjusted. Negative corrected coefficients for AFQT Only and AFQT + Predictor were set to .00; however, the corrected Δ*R*s were allowed to reduce to less than .00.

In-Unit 2. Table 6.4 displays incremental validity results for will-do performance-related criteria for the in-unit 2 sample. From in-unit 1 to in-unit 2, sample sizes for the will-do criteria dropped by an average of 22%, with the largest decrease at 37% (a loss of 244 data points) for the multiple correlation estimation between PSJT and Last APFT Score.[12] Despite the decrement in sample sizes, average validity coefficients remained comparable to those found in the in-unit 1 sample, with some exceptions. AFQT demonstrated little to no potential for predicting two of the will-do performance criteria (i.e., Effort and Discipline AW PRS and Qualifications and Awards [ALQ]).[13] AFQT scores continued to significantly predict composite ratings of Soldiers' ability to effectively work with their peers (Average R = .12; R = .09-.16); however, in contrast to the in-unit 1 results, AFQT also significantly predicted Soldiers' most recent APFT scores (Average R = .09; R = .05-.17). In fact, on average, correlations between AFQT and APFT scores more than doubled between time 1 and time 2. One potential explanation for this change is that Soldiers not able to meet the Army's physical fitness demands likely attrited from the Army prior to 3 years in service, while those that remained would have maintained or improved their APFT scores.

Gains in predictive validity with the addition of the experimental predictors were generally comparable to in-unit 1 results; however, fewer estimates of the change in R were statistically significant. AIM, TAPAS, RBI, and WPA generally demonstrated greater predictive utility than other predictors. For example, AIM, TAPAS, RBI, and WPA (dimensions and facets) provided modest increment in predictive validity over AFQT for Soldiers' most recent APFT scores (ΔR = .10-.27). The RBI also significantly predicted the number of awards Soldiers received (ΔR = .19) but failed to significantly predict composite ratings of Soldiers' effort and discipline or ability to work with others. Curiously, none of the individual RBI scales correlated significantly with these criteria for the in-unit 2 sample (see Appendix C, Table D4). WPA (facets) provided a small but significant boost to the prediction of a composite rating of Soldiers' effort and discipline (ΔR = .14). Although AO, TAPAS, and WPA significantly correlated with scores on the Working Effectively with Others AW PRS composite, none of the estimates of the increment in R were significant.

Consistent with the in-unit 1 results, most of the multiple correlation estimates decreased sharply after adjusting for shrinkage, with nearly all of the change in multiple R values dropping to near zero or becoming negative. AIM, TAPAS, RBI, and WPA (dimensions and facets) continued to provide small to modest increment in predictive validity over AFQT for Soldiers' most recent APFT scores (average ΔR = .18; ΔR = .07-.27). The RBI continued to enhance the prediction of Qualifications and Awards (ALQ) (ΔR = .11) and WPA (Facets) incremented AFQT in the prediction of Effort and Discipline AW PRS (ΔR = .06) but at a much lower rate.

[12] It should be noted that while larger decreases in sample sizes between in-unit 1 and in-unit 2 were evident for other predictor by criterion combinations (e.g., AKA by APFT score dropped by 369 cases), the proportionate (%) loss of cases were smaller.

[13] Note that Disciplinary Incidents (ALQ) was not assessed at time 2. Instead, Qualifications and Awards (ALQ) was added to the set of will-do criteria because Soldiers with more time in service will have had more opportunity to earn accolades for their performance.

Table 6.4. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit 2 Will-Do Performance

		Uncorrected			Corrected		
Predictor/Scale	*n*	AFQT Only	AFQT + Predictor	Δ*R*	AFQT Only	AFQT + Predictor	Δ*R*
Effort and Discipline (Army-Wide [AW] Performance Rating Scales [PRS])							
AO [1]	684	.07	.08	.01	.05	.04	-.01
AIM [6]	325	.03	.20	.17	.00	.09	.09
TAPAS [12]	333	.09	.24	.15	.06	.08	.03
PSJT [1]	338	.08	.11	.03	.05	.06	.01
RBI [14]	568	.06	.19	.13	.03	.05	.02
AKA [6]	678	**.08**	.11	.03	.06	.01	-.05
WPA Dimensions [6]	679	**.09**	.13	.05	.07	.06	-.01
WPA Facets [14]	679	**.09**	**.22**	**.14**	.07	.13	.06
Working Effectively with Others (AW PRS)							
AO [1]	684	**.11**	**.13**	.02	.10	.11	.01
AIM [6]	325	**.11**	.19	.08	.09	.09	.00
TAPAS [12]	333	**.16**	**.26**	.10	.14	.11	-.02
PSJT [1]	338	.09	.11	.02	.06	.06	.00
RBI [14]	568	**.11**	.20	.09	.09	.07	-.02
AKA [6]	678	**.12**	.14	.02	.11	.07	-.04
WPA Dimensions [6]	679	**.13**	**.16**	.03	.12	.10	-.02
WPA Facets [14]	679	**.13**	**.22**	.09	.12	.12	.00
Last Army Physical Fitness Test (APFT) Score (ALQ)							
AO [1]	855	**.09**	**.09**	.00	.08	.06	-.01
AIM [6]	409	.05	**.33**	**.27**	.01	.28	.27
TAPAS [12]	413	.05	**.32**	**.27**	.01	.23	.22
PSJT [1]	420	**.17**	**.17**	.00	.15	.14	-.01
RBI [14]	724	**.08**	**.35**	**.27**	.06	.30	.23
AKA [6]	847	**.10**	**.14**	.05	.08	.09	.00
WPA Dimensions [6]	850	**.10**	**.20**	**.10**	.09	.17	.07
WPA Facets [14]	850	**.10**	**.26**	**.16**	.09	.20	.11
Qualifications and Awards (ALQ)							
AO [1]	871	.01	.02	.00	.00	.00	.00
AIM [6]	417	.03	.12	.09	.00	.00	.00
TAPAS [12]	422	.06	.19	.13	.03	.04	.01
PSJT [1]	426	**.11**	.11	.00	.09	.07	-.02
RBI [14]	737	.01	**.20**	**.19**	.00	.11	.11
AKA [6]	862	.03	.06	.03	.00	.00	.00
WPA Dimensions [6]	866	.04	.12	.09	.01	.06	.05
WPA Facets [14]	866	.04	.17	.13	.01	.07	.06

Note. AFQT = Armed Forces Qualification Test. AFQT + Predictor = Multiple correlation (*R*) between AFQT and selected predictor measure with the criterion. Δ*R* = Increment in *R* over AFQT from adding the selected predictor measure to the regression model ((AFQT + Predictor) – (AFQT Only)). Estimates in bold are statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Listwise deletion was used to account for missing data. Estimates in the "Corrected" columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$, while estimates in the "Uncorrected" columns were not adjusted. Negative corrected coefficients for AFQT Only and AFQT + Predictor were set to .00; however, the corrected Δ*R*s were allowed to reduce to less than .00.

Deployment-Related Criteria

Ratings on the CDPRS scales were combined into a single composite score. While most Soldiers in the in-unit 2 sample had been deployed at least once, ratings data from the CDPRS were only obtained when the rater and ratee had been jointly deployed. This resulted in very small sample sizes for a major element of this performance dimension. CDPRS ratings were not collected on in-unit 1 Soldiers. Soldiers' ability to adjust to the rigors of deployment was assessed in both the in-unit 1 and in-unit 2 samples with a single ALQ scale (Deployment Adjustment) that measured Soldiers' adjustment to deployment schedule.

Table 6.5 displays incremental validity estimates of the experimental predictors' potential contribution to enhancing AFQT in the prediction of Deployment Adjustment and the CDPRS composite.[14] Sample sizes were considerably smaller for these analyses (n = 141-723, average n = 367) which negatively impacts the power to detect statistically significant results. In addition, lack of variance in ratings of Soldiers' performance suppresses the relationships among the variables, further limiting the likelihood that significant correlations will emerge. Appendix C displays correlations between the CDPRS and the predictors. As shown in Table C.5, a limited number (16%) of the correlations were statistically significant and most were below |.10|.

Given that proficiency in combat involves behavioral, motivational, and physical attributes in addition to technical know-how, it is not surprising that the correlations between AFQT and composite ratings of combat performance were small (average R = .04, R = .00-.06), leaving ample room for the experimental predictors to provide incremental predictive validity. Though many of the predictor measures initially demonstrated small to moderate incremental validity over AFQT (i.e., AIM, TAPAS, RBI, AKA, WPA [facets]), none of these estimates were statistically significant (ΔR = .01-.28). Furthermore, the application of Burket's (1964) formula decreased most of the observed estimates to zero. Only corrected incremental validity estimates associated with the addition of TAPAS, RBI, and WPA (facets) to the model continued to increment AFQT. However, validity coefficients for these predictors' decreased by more than half (ΔR = .09-.10). Though not statistically significant, the uncorrected estimates (and to some extent the corrected estimates) suggest that at least a few of the experimental predictor measures (i.e., TAPAS, RBI, and WPA) show promise for predicting deployment performance over and above AFQT. However, no definitive conclusions can be drawn based on these data.

In contrast to the results associated with the combat performance scales, several of the experimental predictor measures demonstrated statistically significant incremental validity over AFQT in the prediction of Soldiers' self-reported adjustment to deployment. At in-unit 1, AIM (ΔR = .27), RBI (ΔR = .19), and AKA (ΔR = .11) contributed small to moderate incremental validity even after correcting for shrinkage (ΔRs = .21, .08, .06, respectively). At in-unit 2, the RBI (ΔR = .14) and AKA (ΔR = .09) continued to exhibit a small but significant potential for predicting this criterion but the increment in R associated with AIM dropped to a non-significant value (ΔR = .09, *ns*). Correcting for shrinkage cut these estimates by roughly half. TAPAS initially contributed a non-significant boost to AFQT in the prediction of Deployment

[14] Incremental validity estimates for the individual CDPRS are not reported because preliminary analyses suggested that the pattern of results for the individual scales was similar to the pattern for the composite. Correlations between the individual CDPRS and the predictor scales are displayed in Table C.5.

Adjustment at both time points (time 1 ΔR = .17, *ns*; time 2 ΔR = .16, *ns*), however correcting for shrinkage revealed these results to be unstable as well.

Table 6.5. Incremental Validity Estimates for Experimental Predictors over the AFQT for Predicting In-Unit Deployment Adjustment and Performance

		Uncorrected			Corrected		
Predictor/Scale	*n*	AFQT Only	AFQT + Predictor	ΔR	AFQT Only	AFQT + Predictor	ΔR
Combat/Deployment Performance Ratings Scales (CDPRS) Composite [a]							
AO [1]	298	.04	.04	.01	.00	.00	.00
AIM [6]	142	.10	.23	.14	.00	.02	.02
TAPAS [12]	147	.06	.34	.28	.00	.09	.09
PSJT [1]	157	.05	.10	.05	.00	.00	.00
RBI [14]	250	.05	.30	.24	.00	.10	.10
AKA [6]	301	.00	.14	.14	.00	.00	.00
WPA Dimensions [6]	294	.01	.08	.08	.00	.00	.00
WPA Facets [14]	294	.01	.28	.27	.00	.10	.10
In-Unit 1 Deployment Adjustment (ALQ)							
AO [1]	371	**.11**	**.13**	.01	.09	.09	.00
AIM [6]	141	.13	**.39**	**.27**	.07	.28	.21
TAPAS [12]	149	**.17**	.34	.17	.13	.09	-.04
PSJT [1]	234	.11	.11	.00	.07	.02	-.04
RBI [14]	340	**.11**	**.30**	**.19**	.08	.17	.08
AKA [6]	376	.10	**.21**	**.11**	.07	.13	.06
WPA Dimensions [6]	383	**.10**	.16	.06	.08	.05	-.03
WPA Facets [14]	383	**.10**	.20	.09	.08	.00	-.08
In-Unit 2 Deployment Adjustment (ALQ)							
AO [1]	723	.07	.08	.00	.06	.04	-.02
AIM [6]	361	**.10**	**.20**	.09	.08	.10	.03
TAPAS [12]	355	.08	.24	.16	.05	.09	.04
PSJT [1]	342	.05	**.19**	**.14**	.00	.16	.16
RBI [14]	611	.08	**.22**	**.15**	.06	.12	.06
AKA [6]	717	.07	**.16**	**.09**	.04	.10	.05
WPA Dimensions [6]	716	.07	.10	.03	.05	.00	-.06
WPA Facets [14]	716	.07	.14	.07	.05	.00	-.05

Note. AFQT = Armed Forces Qualification Test. AFQT + Predictor = Multiple correlation (R) between AFQT and selected predictor measure with the criterion. ΔR = Increment in R over AFQT from adding the selected predictor measure to the regression model ((AFQT + Predictor) – (AFQT Only)). Estimates in bold are statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Listwise deletion was used to account for missing data. Estimates in the "Corrected" columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$, while estimates in the "Uncorrected" columns were not adjusted. Negative corrected coefficients for AFQT Only and AFQT + Predictor were set to .00; however, the corrected ΔRs were allowed to reduce to less than .00.

[a] The CDPRS were not administered in in-unit 1.

Summary

Throughout Soldiers' first term of service, AFQT scores consistently predict can-do performance such as scores on a job knowledge test. Consequently, the experimental measures evaluated in this research contributed little incrementally to the prediction of can-do performance beyond AFQT. However, the experimental measures did predict behaviorally-based (will-do) criteria and, to a lesser extent, combat-related aspects of Soldiers' in-unit job performance over AFQT at two time points.

For the criteria representing more technical job performance during in-unit 1, the experimental predictors yielded uncorrected incremental validity estimates that ranged from .01 to .10 (average ΔR = .04), an average gain of roughly 14% over AFQT. At in-unit 2, validity estimates ranged from .00 to .15 (average ΔR = .05), an average gain of about 23% over AFQT. In contrast, estimates between .01 and .33 (average ΔR = .09) were observed for the more behaviorally-based will-do criteria during in-unit 1, an average increase that more than doubled the prediction potential of AFQT alone. At in-unit 2, comparable increases were found with estimates between .00 and .27 (average ΔR = .09) that contributed to an average increase in incremental validity of 112% over AFQT. That this pattern of results is consistent with findings from previous research suggests that the results are robust.

Incremental validity estimates for the deployment-related criteria (ranged from .00 to .28 (average ΔR = .12), an average increase over AFQT of about 213%. However, as discussed in the chapter, these results were largely unstable. After correcting for shrinkage, the average ΔR due to the addition of the experimental predictors to the model dropped to .03.

In general, AIM, TAPAS, RBI and WPA (facets) produced the largest average uncorrected estimates over AFQT (see Table 6.6). Although AO and the PSJT more often demonstrated statistically significant incremental validity, the magnitude of these estimates was typically lower than that of RBI, TAPAS, and AIM, which often failed to produce statistically significant results. The smaller sample sizes for several experimental measures likely contributed to these non-significant findings. The RBI demonstrated the most potential for predicting in-unit will-do performance and deployment adjustment beyond AFQT (incremental validity estimates ranged from .15 to .33 across in-unit 1 and in-unit 2). To a lesser degree, WPA (facets) contributed to prediction beyond AFQT for many of the less cognitively-loaded criteria (e.g., physical fitness, effort and discipline).

As described earlier in this chapter, previous research examining the utility of the experimental predictors as selection instruments have found larger incremental validity coefficients than those reported here (e.g., Knapp & Heffner, 2009; Knapp & Tremble, 2007). This and other caveats regarding interpretation of the Army Class validation results are discussed further in Chapter 9.

Table 6.6. Summary of Incremental Validity Estimates for Experimental Predictors over the AFQT by Criterion Domain and Months of Service

	Uncorrected						Corrected					
	In-Unit 1			In-Unit 2			In-Unit 1			In-Unit 2		
Criterion Domain/Predictor	Avg. ΔR	Min ΔR	Max ΔR	Avg. ΔR	Min ΔR	Max ΔR	Avg. ΔR	Min ΔR	Max ΔR	Avg. ΔR	Min ΔR	Max ΔR
Can-Do Performance												
AO [1]	.03	.01	.04	.01	.00	.02	.02	.01	.03	.00	-.01	.01
AIM [6]	.04	.01	.07	.06	.01	.11	-.02	-.05	.00	-.01	-.04	.04
TAPAS [12]	.06	.02	.09	.09	.03	.13	-.05	-.08	-.01	-.02	-.03	.00
PSJT [1]	.03	.02	.06	.03	.01	.08	.02	.00	.05	.02	.00	.07
RBI [14]	.07	.03	.10	.09	.04	.15	.00	-.01	.01	.00	-.03	.05
AKA [6]	.03	.01	.04	.02	.01	.03	.00	-.01	.01	-.03	-.06	.00
WPA Dimensions [6]	.03	.02	.04	.04	.03	.04	.00	-.02	.03	-.01	-.04	.03
WPA Facets [14]	.04	.03	.06	.08	.05	.11	-.02	-.05	.03	.00	-.02	.03
Will-Do Performance												
AO [1]	.04	.03	.06	.01	.00	.02	.04	.02	.06	.00	-.01	.01
AIM [6]	.13	.08	.25	.15	.08	.27	.08	.00	.24	.09	.00	.27
TAPAS [12]	.15	.10	.26	.16	.10	.27	.05	-.04	.21	.06	-.02	.22
PSJT [1]	.04	.02	.07	.01	.00	.03	.03	.00	.08	-.01	-.02	.01
RBI [14]	.19	.10	.33	.17	.09	.27	.14	.01	.32	.08	-.02	.23
AKA [6]	.04	.04	.05	.03	.02	.05	.00	.00	.01	-.02	-.05	.00
WPA Dimensions [6]	.04	.01	.08	.07	.03	.10	-.01	-.05	.05	.02	-.02	.07
WPA Facets [14]	.08	.04	.18	.13	.09	.16	.00	-.07	.14	.06	.00	.11
Deployment Adjustment and Combat Performance												
AO [1]	--	--	--	.01	.00	.01	--	--	--	-.01	-.02	.00
AIM [6]	--	--	--	.12	.09	.14	--	--	--	.03	.02	.03
TAPAS [12]	--	--	--	.22	.16	.28	--	--	--	.07	.04	.09
PSJT [1]	--	--	--	.10	.05	.14	--	--	--	.08	.00	.16
RBI [14]	--	--	--	.20	.15	.24	--	--	--	.08	.06	.10
AKA [6]	--	--	--	.12	.09	.14	--	--	--	.03	.00	.05
WPA Dimensions [6]	--	--	--	.06	.03	.08	--	--	--	-.03	-.06	.00
WPA Facets [14]	--	--	--	.17	.07	.27	--	--	--	.03	-.05	.10

Note. The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Estimates in the "Corrected" columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$, while estimates in the "Uncorrected" columns were not adjusted. Shaded cells identify predictors with higher values for a given statistic with the darker shaded cells containing the highest values. No information is provided for In-Unit 1 Deployment Adjustment and Combat Performance because the Combat/Deployment Performance Rating Scales were not administered in In-Unit 1 and data for Deployment Adjustment (ALQ) are provided in Table 6.3.

CHAPTER 7: PREDICTING IN-UNIT SOLDIER ATTRITION AND CONTINUANCE INTENTIONS OVER TIME

Matthew T. Allen (HumRRO)

Limiting Soldier attrition, or early separation from the enlistment contract term of service, is of key importance to the Army. The cost of attrition to the Army is both monetary and harmful to force readiness. In 2003 the Department of Defense estimated that it cost $15,000 dollars to recruit one enlistee (see Buddin, 2005), and early separation compels the Army to increase recruiting activities. The purpose of this chapter is to evaluate the potential of seven experimental measures to enhance the Army's current procedures for screening out applicants likely to attrit during their first term of service. We begin with a brief review of relevant research on Soldier attrition, followed by a discussion of our analytic approach and results. The results are organized into three parts: (a) predicting cumulative Soldier attrition, (b) predicting Soldier attrition at various points in time, and (c) predicting Soldier retention and continuance intentions.

Background

First-term Soldier attrition is pervasive in the Army. Comprehensive studies of this phenomenon reveal that over one-third of Soldiers that access into the Army eventually separate before the end of their first term of service (Buddin, 2005; Strickland, 2005). For reasons stated earlier, reducing these early separations is an important priority for any new measure used to select Army Soldiers. Given the importance of minimizing attrition, it is not surprising that its associated risk factors have been well researched by academics, practitioners, and Army personnel (for a review, see Knapik et al., 2004). The majority of research into attrition risk factors focuses on three types of variables: (a) demographics (e.g., gender, age), (b) accession policies (e.g., waiver policy, Delayed Entry Program [DEP]), and (c) physical factors (e.g., physical fitness, previous injuries).

Across these studies, one consistent finding is that Soldiers without a high school diploma are about twice as likely to attrit in their first-term of enlistment as those with a high school diploma (Knapik et al., 2004). Recognizing this, the Department of Defense restricts the percentage of enlisted Soldiers without a high school diploma or equivalent to 10% per year. [15] Enlistees with a high school diploma or equivalent are classified as Education Tier 1, while those without a high school diploma are classified as Education Tier 2. However, many Tier 2 recruits do not attrit and go on to become highly successful Soldiers. The ARI developed a Tier Two Attrition Screen (TTAS program) to identify Tier 2 applicants who are at reduced risk for attrition (White et al., 2004). The TTAS combines scores from the AIM, a gender-normed Body Mass Index (BMI), and several ASVAB subtests (i.e. Assembling Objects, Math Knowledge, and Mechanical Comprehension) to forecast the likelihood of a Tier 2 applicant completing his or her first term of enlistment (White et al., 2004; White, Hunter, & Young, 2008).

[15] Individuals with 15 units of college credit, in concert with a GED, may also be considered Tier 1. Home study programs in some states may also be considered Tier 1. Tier 2 equivalents include alternative high school credentials and vocational certificates (see C.L Gilroy Memorandum for the Deputy Chief of Staff for G1 [Subject: Education Credentials – Definitions, Tier Placement, and Enlistment Prioritization], September 21, 2004 for more details).

In 2005-2009, the Army implemented a Tier 2 market expansion program. Under this pilot program, the Army was permitted to enlist additional Tier 2 recruits (beyond the 10% cap) who scored higher on TTAS and met other qualifications (i.e., AFQT CAT I-IIIA) and were projected to have lower attrition rates, similar to Tier 1 recruits. Over 28,000 Regular Army Soldiers have accessed under the TTAS program since its inception. These recruits helped the Army to meet yearly accession goals during a very difficult recruiting period. Results from the 48-month evaluation confirmed that attrition rates of Tier 2 recruits who passed TTAS were significantly lower than the rates for those who failed TTAS and were closer to the rates of Soldiers in Tier 1 (White, Jose, & LaPort, 2011). The TTAS research demonstrates the promise of using both cognitive and non-cognitive measures as a supplement to the Tier system for managing attrition.

The present research examines whether the experimental predictor measures can predict Soldier attrition beyond the Army's primary method for managing attrition—the Education Tier system. Though Education Tier 2 applicants face additional screening requirements compared to Education Tier 1, we believe it is the most appropriate baseline variable for our analysis because attrition is still 57% to 79% higher for Education Tier 2 Soldiers in the present sample. It should also be noted that Tier 2 Soldiers account for roughly 25% of the analysis sample, due in part to the fact that the data collections specifically targeted Tier 2 Soldiers and because the Army was allowed to enlist additional Tier 2 recruits under the TTAS program to meet its yearly accession goals.

A second, though less common, theme of previous attrition research has examined Soldiers' reasons for attrition and how the reasons change throughout their first term of service (General Accounting Office, 2000; Lytell & Drasgow, 2009; Strickland, 2005). This research demonstrates that during IMT (around the first 6 months of service), the primary reasons for Soldier attrition are for performance and medical-related issues, while attrition for moral character-related reasons increase once Soldiers join their units (Strickland, 2005). Different types of predictors, then, are more predictive of certain types of attrition than others. For example, in a study of enlisted Airmen, Hooper, Paullin, Putka, and Strickland (2008) found that Airmen with lower AFQT scores were more likely to attrit for performance reasons. AFQT also predicted other types of attrition. However, the magnitude of the effect was strongest for performance-related attrition. The present research expands on this work by examining whether (a) the experimental measures predict different types of attrition (i.e., moral, medical, and performance) as well as overall attrition, and (b) the experimental measures can predict patterns of attrition over time.

In addition to attrition during IMT, we were also interested in whether the experimental measures can predict whether Soldiers will (a) attrit at some later point in their first term of service, (b) re-enlist after their current term of service, or (c) make the Army a career. Not all of the Soldiers in the present sample had reached the end of their first term of service at the time of these analyses. As a result, we were unable to examine these continuance behaviors directly. To address this, we used self-reported attitudes and behavioral intentions as a proxy for Soldiers' actual continuance behavior. Previous research has shown that specific attitudinal antecedents, such as self-reported affective commitment to the Army and thoughts about attriting, are strong predictors of post-IMT separation behavior (Lytell & Drasgow, 2009; Strickland, 2005).

Approach

As described in Chapter 5, we obtained Soldiers' attrition status from their administrative records at key points during their first term of service—at 3 months (attrition near or after the completion of Basic Combat Training [BCT]), 4 months (attrition during AIT/OSUT), 6 months (attrition near or after completion of AIT/OSUT), and at regular quarterly intervals thereafter. A Soldier's attrition status is cumulative, reflecting whether a Soldier attrited at *any point* prior to the target month in service. Accordingly, overall Soldier attrition never declines from one point in time to the next, as Table 7.1 demonstrates. Consistent with previous research (Knapik et al., 2004), Table 7.1 also demonstrates that Education Tier 1 Soldiers in this sample were notably less likely to attrit at any point in their first term than Tier 2 Soldiers. For this reason, rather than using AFQT (which previous research has shown to only have a modest correlation with attrition, e.g., Knapp & Heffner, 2009) as the baseline predictor, Education Tier was used to evaluate the potential of the experimental measures to reduce Soldier attrition beyond the Army's current policy.

Table 7.1. Cumulative Attrition Rates over Time by Education Tier

	Months in Service							
Group	3	4	6	9	12	15	18	21
% Attrition								
Ed Tier 1	5.0	6.8	9.8	12.0	14.1	16.1	17.7	19.2
Ed Tier 2	8.5	12.1	17.4	21.5	24.5	27.8	30.4	32.4
Total	6.0	8.3	12.0	14.7	17.1	19.5	21.4	23.0
n								
Ed Tier 1	3,690	3,690	3,689	3,688	3,688	3,687	3,687	3,680
Ed Tier 2	1,481	1,481	1,481	1,481	1,481	1,481	1,481	1,481
Total	5,171	5,171	5,170	5,169	5,169	5,168	5,168	5,161

Group	24	27	30	33	36	39	42
% Attrition							
Ed Tier 1	21.2	22.2	23.9	25.2	27.0	28.4	30.3
Ed Tier 2	34.2	36.3	38.4	40.2	42.5	44.8	48.2
Total	24.9	26.3	28.1	29.6	31.5	33.1	35.1
n							
Ed Tier 1	3,680	3,675	3,637	3,618	3,614	3,180	2,593
Ed Tier 2	1,481	1,481	1,480	1,479	1,478	1,278	951
Total	5,161	5,156	5,117	5,097	5,092	4,458	3,544

Note. % Attrition = Percentage in each group that separated through that month of serve out of the total number in that population.

Similar to the method applied in Chapter 6, this evaluation was accomplished by examining the predictive efficacy of a baseline model (Education Tier only) with a model that includes the experimental predictors (Education Tier + scores from one of the experimental predictors). A key difference between the analyses presented in Chapter 6 and the present ones is the use of logistic regression. Logistic regression is more appropriate than Ordinary Least Squares (OLS) regression for binary criterion data like attrition.

Specifically, we tested whether the experimental measures could contribute to the prediction of cumulative attrition at multiple points in time using the following steps:

1. Model attrition using hierarchical logistic regression, treating cumulative attrition (3-month attrition, 6-month attrition, etc.) as the criterion. Education Tier was entered as the sole predictor in the first step of the model (i.e., establishing the baseline model), followed by scores from a given experimental measure as predictors in the second step of the model. We fit a separate hierarchal model for each of the eight experimental predictor measures (i.e., AO, AIM, TAPAS, PSJT, RBI, AKA, WPA dimensions, and WPA facets).

2. Compute the statistical significance of the difference in model fit between the two models using deviance statistics.[16] Deviance statistics have a chi-square distribution. Accordingly, statistical significance was determined by subtracting the deviance statistic for the second step of the model (Education Tier + experimental measure) from the deviance statistic from the first step of the model (Education Tier only), and then determining whether the difference is statistically significant at $p < .05$ (with *df* equal to the number of scales on the given experimental measure). This provided a test of whether adding the given experimental measure to the model containing only Education Tier significantly improved model fit.

3. Lastly, to provide an index of the gain in prediction achieved by adding the given experimental measure into a model that only included Education Tier, we examined the difference between point-biserial correlations that reflected (a) the correlation between the predicted probability of attrition computed based on the first step of the model (Education Tier) and actual attrition behavior, and (b) the correlation between the predicted probability of attrition computed based on the second step of the model (Education Tier + experimental measure) and actual attrition behavior.

One limitation of using cumulative overall attrition as the criterion is that it treats many types of early separation from the Army, regardless of reason, the same. For example, a Soldier that separates due to a major injury is treated the same as an individual that separates for character reasons (e.g., breaking the law). Presumably, the Army would be more interested in being able to predict the latter types of attrition than the former, as the former is more likely to be related to circumstances beyond the Soldier's control. As described earlier, previous studies have identified multiple "types" of attrition using administrative records of the reasons for separation. Though there are a number of problems with these archived records, such as deliberate falsification (General Accounting Office, 1998), previous studies have successfully used these records to gain a better understanding of the complex reasons for Soldier attrition. Following the procedures used in previous research (e.g., Hooper et al., 2008, Putka, Noble, Becker, & Ramsberger, 2004; Strickland, 2005), the attrition "types" were created with the following steps:

[16] The deviance statistic (–2log likelihood) can be used to assess model fit in logistic regression (Singer & Willett, 1993). Deviance statistics capture the difference between the current model and the best possible (i.e., saturated) model.

1. Soldiers' reasons for separation were identified in their administrative records using Separation Program Designators (SPDs). These SPDs were converted to Interservice Separation Codes (ISCs) so that the resulting categorization scheme would be consistent with previous research (Strickland, 2005). Soldiers with SPDs that could not be converted to an ISC were not considered for further analysis.[17]

2. Following Strickland (2005), the ISCs were placed into five attrition categories: (a) medical, (b) family, (c) moral character, (d) performance, and (e) other. A summary of this categorization process can be found in Table 7.2.

3. Soldiers that attrited before reaching 36 months in service and had an ISC code from one of the five aforementioned categories were considered that "type" of attrit. A separate attrition variable was created for each type. We excluded those that attrited for a reason other than the target type from this analysis.

Table 7.2. Treatment of Select Interservice Separation Codes (ISC) for Different Types of Attrition Analyses

Interservice Separation Codes (ISC)	Attrition Type
10: Condition existing prior to service	Medical
14: Disability, no condition existing prior to service, no severance pay	Medical
16: Unqualified for active duty, other	Medical
17: Failure to meet weight or body fat standards	Medical
22: Dependency or hardship	Family
60: Character or behavior disorder	Other
64: Alcoholism	Moral
65: Discreditable incidents, civilian or military	Moral
67: Drugs	Moral
71: Civil court conviction	Moral
73: Court-martial	Moral
74: Fraudulent entry	Moral
75: AWOL or desertion	Moral
76: Homosexuality	Other
77: Sexual perversion	Moral
78: Good of the service (discharge in lieu of court-martial)	Moral
79: Juvenile offender	Moral
80: Misconduct, reason unknown	Moral
83: Pattern of minor disciplinary infractions	Moral
84: Commission of a serious offense	Moral
86: Unsatisfactory performance (former Expeditious Discharge Program)	Performance
87: Entry level perform and conduct (former Trainee Discharge Program)	Performance
90: Secretarial authority	Other
91: Erroneous enlistment or induction	Other
92: Sole surviving family member	Other

[17] Some Army administrative data sources use SPDs, while others use ISCs. The TTAS database (from where the attrition data was drawn) uses SPDs. A total of 70 (4.7%) SPDs could not be converted into a valid ISC.

Table 7.2. (Continued)

Interservice Separation Codes (ISC)	Attrition Type
94: Pregnancy	Family
95: Minority (underage)	Other
96: Conscientious objector	Other
97: Parenthood	Family
98: Breach of contract	Other
Total	

Note. For the purposes of this analysis, all Separation Program Designators (SPDs) in the TTAS database were converted into ISCs. This table only reflects Soldiers with SPD codes that (a) could be converted to ISCs and (b) could be categorized into one of the targeted attrition "types." Previous research determined which codes to include in each attrition type (e.g., Hooper et al., 2008).

We chose 36 months as the time period for analysis because nearly the entire sample (98.4%) had the opportunity to reach this point in their first enlistment term. As Table 7.3 shows, only three out of the five types of attrition had base rates large enough for analysis (performance, moral character, and medical). The remaining two types of attrition (other and family) were not considered for any further analysis. Consistent with the findings for overall attrition, Tier 1 Soldiers were less likely to attrit for moral character (10.7% versus 24.6%), performance (4.5% versus 6.6%), and medical (12.7% versus 20.9%) reasons than Tier 2 Soldiers. This suggests that Education Tier is an appropriate baseline predictor for analyses involving these three types of attrition as well as overall attrition. We used the same logistic regression procedure employed for overall attrition to analyze the results for the three specific types of attrition.

Table 7.3. Type of 36-Month Cumulative Attrition by Education Tier

	Moral Attrition			Performance Attrition			Medical Attrition		
Group	*n*	*n Attrit*	*% Attrit*	*n*	*n Attrit*	*% Attrit*	*n*	*n Attrit*	*% Attrit*
Education									
Tier 1	2,953	316	10.7	2,762	125	4.5	3,020	383	12.7
Tier 2	1,127	277	24.6	910	60	6.6	1,075	225	20.9
Total[a]	4,080	593	14.5	3,672	185	5.0	4,095	608	14.8
	Family Attrition			Other Attrition			Overall Attrition		
Group	*n*	*n Attrit*	*% Attrit*	*n*	*n Attrit*	*% Attrit*	*n*	*n Attrit*	*% Attrit*
Education									
Tier 1	2,725	88	3.2	2,667	30	1.1	3,614	977	27.0
Tier 2	877	27	3.1	870	20	2.3	1,478	628	42.5
Total[a]	3,602	115	3.2	3,537	50	1.4	5,092	1,605	31.5

Note. Tier 1 = High School Diploma or Equivalent, Tier 2 = Non-High School Diploma. All types of attrition are cumulative through 36 months in service. Results are limited to Regular Army, non-prior service Soldiers. Other types of attrition aside from the target type were set to system missing.
[a]Total sample sizes differ across types of attrition because of attrition of a non-target type was treated as missing data.

Findings

Results of analyses examining the prediction of cumulative overall attrition are reported in Table 7.4. With the exception of the PSJT, all of the experimental measures predicted attrition at a significantly higher rate than Education Tier alone. Across all of the time periods, three measures consistently predicted attrition at a higher rate than the other ones, the RBI (Δr_{pb} = .07 to .14, average Δr_{pb} = .10), AIM (Δr_{pb} = .05 to .10, average Δr_{pb} = .08), and TAPAS (Δr_{pb} = .05 to .09, average Δr_{pb} = .07). Though the changes are not large, in general, the rates of prediction for the RBI are higher at earlier months than later months, while the rates for the AIM and TAPAS are fairly steady across all time periods. The WPA (at both the facet and dimension level) predicted attrition beyond Education Tier at the next highest rate, with Δr_{pb} ranging from .03 to .06 (average Δr_{pb} = .04 for the dimension level and .05 for the facet level). Finally, AO (Δr_{pb} = .01 to .03, average Δr_{pb} = .02) and AKA (Δr_{pb} = .01 to .04, average Δr_{pb} = .02) also predicted attrition at a significantly higher rate than Education Tier only. However, they did so at a lower rate, on average, than the other experimental measures.

Table 7.4. Incremental Validity for Experimental Predictors over Education Tier for Predicting Cumulative Attrition through 36 Months of Service

Predictor	Ed Tier Only	Ed Tier + Predictor	Δ r_{pb}	Ed Tier Only	Ed Tier + Predictor	Δ r_{pb}	Ed Tier Only	Ed Tier + Predictor	Δ r_{pb}
	3 Months			*4 Months*			*6 Months*		
AO [1]	.07	.08	**.01**	.09	.10	**.01**	.11	.13	**.02**
AIM [6]	.05	.12	**.07**	.08	.13	**.05**	.09	.17	**.08**
TAPAS [12]	.09	.14	**.06**	.10	.18	**.08**	.11	.19	**.08**
PSJT [1]	.07	.07	.00	.09	.09	.00	.11	.11	.00
RBI [14]	.06	.19	**.14**	.07	.20	**.13**	.10	.21	**.11**
AKA [6]	.07	.10	**.04**	.09	.12	**.03**	.10	.12	**.02**
WPA Dimensions [6]	.07	.11	**.04**	.09	.14	**.04**	.11	.15	**.04**
WPA Facets [14]	.07	.13	**.05**	.09	.15	**.06**	.11	.17	**.06**
	9 Months			*12 Months*			*15 Months*		
AO [1]	.12	.13	**.02**	.12	.14	**.02**	.13	.15	**.02**
AIM [6]	.08	.18	**.10**	.08	.17	**.09**	.09	.18	**.09**
TAPAS [12]	.11	.20	**.09**	.12	.19	**.07**	.13	.21	**.08**
PSJT [1]	.16	.16	.00	.17	.17	.00	.18	.18	.00
RBI [14]	.11	.23	**.12**	.11	.22	**.11**	.12	.23	**.10**
AKA [6]	.12	.14	**.02**	.12	.14	**.02**	.13	.15	**.02**
WPA Dimensions [6]	.12	.16	**.04**	.13	.16	**.04**	.14	.17	**.04**
WPA Facets [14]	.12	.18	**.06**	.13	.17	**.05**	.14	.18	**.04**
	18 Months			*21Months*			*24 Months*		
AO [1]	.14	.15	**.02**	.14	.16	**.02**	.13	.15	**.02**
AIM [6]	.10	.19	**.10**	.11	.19	**.08**	.10	.19	**.09**
TAPAS [12]	.14	.21	**.07**	.14	.21	**.07**	.13	.21	**.07**
PSJT [1]	.18	.18	.00	.18	.18	.00	.18	.18	.00
RBI [14]	.13	.22	**.09**	.13	.22	**.08**	.13	.22	**.09**
AKA [6]	.14	.15	**.01**	.14	.16	**.02**	.13	.15	**.02**
WPA Dimensions [6]	.14	.18	**.04**	.15	.18	**.04**	.14	.17	**.04**
WPA Facets [14]	.14	.19	**.05**	.15	.19	**.04**	.14	.18	**.05**

Table 7.4. (Continued)

Predictor	Ed Tier Only	Ed Tier + Predictor	Δ r_{pb}	Ed Tier Only	Ed Tier + Predictor	Δ r_{pb}	Ed Tier Only	Ed Tier + Predictor	Δ r_{pb}
	27Months			*30Months*					
AO [1]	.14	.16	**.02**	.14	.17	**.03**			
AIM [6]	.11	.19	**.08**	.12	.19	**.08**			
TAPAS [12]	.14	.21	**.07**	.15	.21	**.06**			
PSJT [1]	.18	.18	.00	.18	.18	.00			
RBI [14]	.14	.22	**.08**	.14	.22	**.08**			
AKA [6]	.14	.16	**.01**	.15	.16	**.01**			
WPA Dimensions [6]	.15	.19	**.04**	.15	.18	**.03**			
WPA Facets [14]	.15	.20	**.05**	.15	.19	**.04**			
	33 Months			*36 Months*					
AO [1]	.15	.17	**.02**	.15	.17	**.02**			
AIM [6]	.12	.20	**.08**	.13	.19	**.07**			
TAPAS [12]	.15	.22	**.06**	.16	.21	**.05**			
PSJT [1]	.18	.18	.00	.18	.18	.00			
RBI [14]	.15	.22	**.07**	.15	.22	**.07**			
AKA [6]	.15	.16	**.02**	.15	.16	**.01**			
WPA Dimensions [6]	.15	.19	**.03**	.15	.18	**.03**			
WPA Facets [14]	.15	.20	**.04**	.15	.19	**.04**			

Note. r_{pb} = Point-biserial correlation between Soldiers' predicted probability of attriting with their actual attrition behavior. Bolded values indicate that adding the experimental measure(s) to a model containing only education tier resulted in significantly better model fit (based on change in -2 log likelihood statistics discussed in the text, $p < .05$).The numbers in brackets indicate the number of scale scores that the measure contributed to the regression model. Soldiers that attrited for reasons other than the target type were coded as system missing for the purpose of this analysis. The WPA consists of six dimensions and 14 facets embedded within those dimensions. Results are limited to Regular Army, non-prior service Soldiers. AO *n* = 4,762-4,840; AIM *n* = 2,290-2,338; TAPAS *n* = 2,272-2,316; PSJT *n* = 2,262-2,287; RBI *n* = 4,022-4,086; AKA *n* = 4,715-4,791; WPA *n* = 4,669-4,746.

Results examining the baseline and experimental models predicting the three types of attrition at 36 months are presented in Table 7.5. Consistent with the findings for overall attrition, these results suggest that multiple experimental Army Class measures predict Soldier attrition for moral character, medical, and performance reasons beyond Education Tier only. Three of these measures predicted these types of attrition at a higher rate than the other experimental measures: RBI (Δr_{pb} = .06 to .11, average Δr_{pb} = .09), AIM (Δr_{pb} = .05 to .11, average Δr_{pb} = .08), and TAPAS (Δr_{pb} = .05 to .09, average Δr_{pb} = .08). These three measures predicted performance and medical attrition particularly well. Next, the WPA dimensions and facets also predicted all three types of attrition at a higher rate than models that included Education Tier only, albeit not as consistently as the three temperament measures. The Δr_{pb} for the WPA ranged from .01 to 07, with the highest rates of prediction emerging for medical attrition. Finally, AO incrementally predicted all three types of attrition at a significantly higher rate than the baseline model. However, the magnitude of the increment was generally lower than that observed for the other predictor measures (Δr_{pb} = .01 to .04, average Δr_{pb} = .02). The highest coefficient for AO emerged for performance-related attrition. Finally, the AKA predicted moral character attrition at a higher rate than Education Tier (Δr_{pb} = .02), but did not predict either performance or medical attrition. Consistent with the cumulative attrition results, the PSJT did not predict attrition beyond Education Tier.

Table 7.5. Incremental Validity of Experimental Predictors over Education Tier for Type of Cumulative Attrition through 36 Months of Service

Predictor	*n*	Ed Tier Only	Ed Tier + Predictor	Δr_{pb}
Moral Character				
AO [1]	3,815	.17	.19	**.01**
AIM [6]	1,780	.18	.23	**.05**
TAPAS [12]	1,795	.19	.24	**.05**
PSJT [1]	1,874	.18	.18	.00
RBI [14]	3,195	.18	.23	**.06**
AKA [6]	3,776	.17	.18	**.01**
WPA Dimensions [6]	3,732	.18	.19	.01
WPA Facets [14]	3,729	.18	.21	**.03**
Performance				
AO [1]	3,426	.03	.07	**.04**
AIM [6]	1,624	.02	.11	**.09**
TAPAS [12]	1,632	.04	.13	**.09**
PSJT [1]	1,666	.08	.09	.01
RBI [14]	2,851	.04	.15	**.11**
AKA [6]	3,400	.03	.06	.02
WPA Dimensions [6]	3,380	.04	.09	**.05**
WPA Facets [14]	3,377	.04	.10	.06
Medical				
AO [1]	3,813	.10	.12	**.02**
AIM [6]	1,807	.08	.19	**.11**
TAPAS [12]	1,819	.11	.19	**.09**
PSJT [1]	1,863	.11	.11	.00
RBI [14]	3,211	.10	.19	**.09**
AKA [6]	3,798	.10	.11	.01
WPA Dimensions [6]	3,765	.10	.15	**.05**
WPA Facets [14]	3,762	.10	.17	**.07**

Note. r_{pb} = Point-biserial correlation between Soldiers' predicted probability of attriting with their actual attrition behavior. Bolded Δr_{pb} indicate that adding the experimental measure(s) to a model containing only education tier resulted in significantly better model fit (based on change in -2 log likelihood statistics discussed in the text, $p < .05$).The numbers in brackets indicate the number of scale scores that the measure contributed to the regression model. Soldiers that attrited for reasons other than the target type were coded as system missing for the purpose of this analysis. The WPA consists of six dimensions and 14 facets embedded within those dimensions. Results are limited to Regular Army, non-prior service Soldiers.

In summary, these results suggest that the RBI, AIM, and TAPAS are the best predictors of overall cumulative attrition, followed by the WPA. AO and AKA are also non-trivial predictors of attrition, though not at the same magnitude. In predicting the three "types" of attrition (moral, performance, and medical), the AIM, TAPAS, and RBI predicted all three types beyond Education Tier only, while WPA predicted medical attrition and AO predicted performance attrition.

Predicting Soldier Attrition Over Time

Approach

While the previous section described the experimental measures' potential to predict whether a Soldier would attrit (or not), it did not address whether the experimental measures can be used to predict *when* attrition would occur. Predicting when attrition occurs is potentially important to Army decision-makers because certain experimental measures may be able to predict attrition at multiple points in a Soldier's first term, while others may be less able to do so. For example, if the Army were able to significantly reduce attrition during BCT through policy changes (e.g., through more rigorous medical screening), the primary time period of interest would be a later point in the Soldier's first term, such as the end of AIT/OSUT or during their first unit assignment. Knowing which experimental measure(s) best predict attrition post-BCT then would be important when evaluating which measure(s) to consider for operational use.

Examining the experimental measures' potential to predict when attrition occurs requires a different analytic approach than the one used to model cumulative attrition in the previous section. Examination of cumulative attrition via logistic regression is limited because of redundancy between early attrition (e.g., through 3 months) and later cumulative attrition criteria (e.g., through 12 months). For example, the experimental measures that predict early attrition will also predict later attrition by virtue of the fact that these later months also capture attrition through the early months. One option to avoid this would be to form separate attrition variables that are non-cumulative (i.e., those that attrit at earlier months would be treated as system-missing). However, this approach is not optimal either because (a) it does not make full use of the data (i.e., it lowers overall sample size for the analysis), and (b) it does not allow one to systematically evaluate differences in a predictor's relationship to attrition over time.

To address these limitations, we employed Event History Analysis (EHA; also referred to as Discrete-Time Hazard Models) to analyze the data. EHA has been used successfully in a number of studies examining attrition in the armed services. Because the process for using EHA to study attrition has been outlined in great detail in previous research (e.g., Hooper et al., 2008; Strickland, 2005; for a more complete description, see Singer & Willett, 2003), we will only briefly describe the steps taken in these analyses.

Step 1: Convert the original analysis dataset to a person-period dataset.

In a traditional analysis file, there is one record for each Soldier in the database. In a person-period dataset, there is one record for every Soldier by time period under investigation. A Soldier that has 6 months in service would have two records, one at 3 months and one at 6 months. In our dataset, all time periods where the Soldier did not attrit were coded as "0." If an attrition event occurred for a Soldier at a particular 3-month time period, that instance was coded as "1." To make the time periods equivalent, we dropped the 4-month attrition variable so that each record represented 3 months in service. For each Soldier, we had records for up to 14 time periods (every 3 months from 3 to 42), or less depending on whether separation occurred. If separation did occur during those 10 time periods, the Soldier did not have a record for the remaining time periods. For example, if a Soldier separated after 6 months, s/he would have a

record at 3 months (coded as "0") and 6 months (coded as "1"), and no time periods after that. In the previous analyses, we only analyzed the data for Soldiers who could have been in service for 36 months (based on their accession dates). However, for the purposes of EHA, we could use data on all Soldiers, regardless of how long they could have been in service. Cases of Soldiers that had not been in service for 42 months were treated as censored observations up to their number of months in service, meaning their attrition record included a "0" for each time period.[18]

Step 2: Compute a time parameter for each attrition variable.

Most OLS and logistic regression models include an intercept, or starting point, for the model. In these static models, the intercepts are a constant. In EHA models, however, an intercept represents a model for time, with one intercept for each "bend" in a plot of conditional attrition probability over time. The most general specification has one parameter for each time period. In our case, there were 14 parameters total, one for each 3-month interval over 42 months. However, more parsimonious time parameters can be specified for models that have less erratic differences from one time period to the next. Because of the relatively small number of time periods, and the relatively large sample size, the EHA models in this analysis used a general (14-parameter) specification of time. For the three types of attrition (moral character, medical, and performance), all non-target types of attrition were censored up until the point of the event (i.e., the records were "0" until the attrition event).

Step 3: Use logistic regression to test nested experimental predictor models.

To test whether the experimental Army Class measures contribute uniquely to predicting attrition over time, we applied logistic regression to the person-period datasets created in Step 2, treating attrition (or type of attrition) as the dependent variable. The hierarchical EHA models included the following variables in the following steps:

1. The time parameter variables
2. The time parameter variables + Education Tier
3. The time parameter variables + Education Tier + the target experimental predictor scales

We then subtracted the deviance statistics for the higher order models from the deviance statistics obtained for the lower order models. This process was repeated for each experimental measure to identify the most promising ones for comparative evaluation. Note that these analyses test the "nested" effect of the experimental measures. In other words, this analysis focuses on whether each experimental measure predicted attrition over time *beyond the baseline models* (1 and 2 above), rather than comparing the experimental measures to one another. In fact, comparing the deviance statistics across predictors would be inappropriate, as the analyses must be limited to the same sample (Singer & Willett, 2003).

[18] "Censoring" is a term used to describe data where an individual does not experience the target event (in this case, attrition). Censored data are problematic for traditional statistics because they only inform event/attrition non-occurrence. EHA accounts for censored data by analyzing the "hazards"—the proportion of Soldiers in the beginning of a 3-month time period that attrit during that time period—at each unit of analysis.

Step 4: Use logistic regression to test non-nested predictor models.

One issue with the current data is that it is difficult to compare results across experimental measures because some of the measures (e.g., the PSJT and AIM) were not administered to the same sample. To address this issue, the analyses described in Step 3 were repeated, but the sample was limited to Soldiers with complete data for the entire set of "best bet" measures identified (n = 1,406). These analyses were "non-nested" because the purpose was to make comparisons *across the models* rather than within. Once these analyses were completed, we used the deviance statistics to compute two model fit statistics: (a) the Akaike Information Criterion (AIC), and (b) the Bayesian Information Criterion (BIC). These indices were computed to account for spuriously large effects that can result when multiple parameters are included in a model.

The difference between the deviance and AIC/BIC statistics is that the latter penalizes less parsimonious models more heavily. Both the AIC and BIC were computed with the following formula (Singer & Willett, 2003):

> Deviance + 2 * (scale factor) * (number of parameters in the model) (1)
>
> where the scale factor = 1 for the AIC and half the log of the number of events (i.e., number attriting) for the BIC.

For all three fit indices (deviance, AIC, BIC), the smaller the value, the better the model fit.

Findings

The results of the nested EHA models are shown in Table 7.6. Overall, the results suggest that multiple experimental predictor measures uniquely explain attrition over time beyond the time parameter and Education Tier. For overall attrition, regardless of reason, all of the experimental measures except for the PSJT had models that fit the data significantly better than the model with Education Tier only. For moral character attrition, all of the models with the experimental measures fit the data significantly better than the baseline models. For performance attrition, five measures had significantly better fitting models than the baseline model when all of the component scales were included: AO, AIM, TAPAS, RBI, and WPA dimension scores. These same five measures also predicted medical attrition, with the WPA yielding a significantly better fitting model at both the dimension and facet level.

As mentioned above, the deviance statistics for the models across predictors are not directly comparable due to dependency on the sample. However, the differences in deviance statistics are somewhat interpretable across models, as they represent the better fit of one model compared to the next. Consistent with the cumulative attrition results reported in Tables 7.4 and 7.5, the experimental measures with the highest incremental model fit tended to be (in order of magnitude): (a) RBI, (b) TAPAS and AIM , (c) WPA, and (d) AO. These five measures constitute the "best bet" predictors that were included in the nested analyses. Consistent with the cumulative attrition results previously reported, the AKA predicted some significant variance in attrition but at a consistently lower rate than the aforementioned measures.

Table 7.6. Event History Analysis Assessing the Goodness-of-Fit of Nested Experimental Predictor Models Through 42 Months of Service

	Deviance Statistics (-2LL)				
Predictor	Time Parameter (Step 1)	Time + Ed Tier (Step 2)	Time + Ed Tier + Predictor (Step 3)	Step 1 v. Step 2	Step 2 v. Step 3
Overall Attrition					
AO [1]	14,174.15	14,051.54	14,010.32	**122.61**	**41.22**
AIM [6]	7,099.52	7,058.56	7,004.43	**40.96**	**54.13**
TAPAS [12]	6,859.89	6,795.68	6,742.51	**64.22**	**53.17**
PSJT [1]	6,278.31	6,194.58	6,193.80	**83.72**	0.78
RBI [14]	12,214.85	12,088.29	11,975.53	**126.56**	**112.77**
AKA [6]	14,020.88	13,897.05	13,883.70	**123.83**	**13.35**
WPA Dimensions [6]	13,802.67	13,677.98	13,630.08	**124.69**	**47.90**
WPA Facets [14]	13,791.81	13,667.97	13,603.69	**123.84**	**64.28**
Moral Character Attrition					
AO [1]	6,700.26	6,574.64	6,558.37	**125.62**	**16.27**
AIM [6]	3,238.76	3,181.02	3,148.85	**57.74**	**32.17**
TAPAS [12]	3,154.11	3,083.97	3,038.03	**70.14**	**45.94**
PSJT [1]	3,008.64	2,945.25	2,940.97	**63.39**	**4.28**
RBI [14]	5,768.40	5,665.60	5,584.90	**102.80**	**80.70**
AKA [6]	6,502.45	6,382.72	6,368.49	**119.73**	**14.22**
WPA Dimensions [6]	6,258.31	6,134.37	6,120.91	**123.94**	**13.47**
WPA Facets [14]	6,248.24	6,125.56	6,091.48	**122.68**	**34.09**
Performance Attrition					
AO [1]	1,974.48	1,973.77	1,962.60	0.72	**11.17**
AIM [6]	1,189.11	1,188.99	1,174.06	0.12	**14.93**
TAPAS [12]	1,068.82	1,068.20	1,042.66	0.63	**25.54**
PSJT [1]	598.94	593.50	592.02	**5.44**	1.48
RBI [14]	1,679.14	1,677.52	1,640.32	1.62	**37.20**
AKA [6]	1,935.64	1,934.82	1,929.11	0.83	5.71
WPA Dimensions [6]	1,978.84	1,976.85	1,960.71	2.00	**16.13**
WPA Facets [14]	1,978.52	1,976.52	1,954.33	2.00	22.19
Medical Attrition					
AO [1]	7,685.50	7,632.76	7,617.79	**52.74**	**14.97**
AIM [6]	3,994.80	3,978.13	3,944.35	**16.68**	**33.78**
TAPAS [12]	3,854.63	3,825.94	3,790.29	**28.69**	**35.65**
PSJT [1]	3,412.82	3,387.73	3,387.22	**25.08**	0.51
RBI [14]	6,798.66	6,757.42	6,700.79	**41.23**	**56.63**
AKA [6]	7,759.80	7,709.07	7,707.23	**50.73**	1.84
WPA Dimensions [6]	7,583.06	7,535.72	7,510.68	**47.34**	**25.04**
WPA Facets [14]	7,573.78	7,527.19	7,489.77	**46.59**	**37.43**

Note. Deviance differences in bold are statistically significant, $p < .05$, using a chi-square distribution. Model comparisons were computed by subtracting the lower step (e.g., Step 1) from the higher step (e.g., Step 2) so that large positive numbers in the last two columns always reflect lower deviance. For the Step 1 v. 2 comparison, the degrees of freedom (*df*) is always 1; for the Step 2 v. 3 comparison, the *df* is equal to the number of scales the experimental predictor measure contributes to the model. Soldier that attrited for reasons other than the target type were censored for the purpose of this analysis. LL = Log Likelihood. Results are limited to Regular Army, non-prior service Soldiers.

The results of the non-nested EHA models comparing the model fit of the "best bet" experimental predictors to each other are shown in Table 7.7. We can interpret the relative fit of each predictor using the three indices reported in the table. The deviance statistics results, which do not make any adjustments for the number of parameters in the model, suggest that the type of attrition changes the predictors that provide the best incremental fit to the data beyond the time parameters and Education Tier. For overall attrition, the strongest predictor was the RBI. For moral character and performance attrition, the TAPAS emerged as the strongest predictor. Finally, for medical attrition, the strongest predictor was AO.

However, when examining the AIC and BIC, which does penalize predictors that contribute more parameters to the model, the picture of what experimental measures contribute most to the models changes. In interpreting the BIC, one rule of thumb is a difference of 0 to 2 is considered "weak," a difference of 2 to 6 is considered "positive," and a difference of 6 to 10 is considered "strong" (cf. Singer & Willett, 2003). The AIC is often interpreted similarly. When examining the results for overall attrition, the AO, AIM, and RBI emerge as the strongest predictors according to the AIC. Using the BIC, the AO emerges as the strongest predictor, followed by the AIM. Recall that the RBI emerged as the strongest predictor according to the deviance statistics.

For moral character attrition, the TAPAS emerged as the strongest predictor according to the deviance statistic. When examining the AIC, the AIM and AO emerged as comparably strong predictors to the TAPAS. When examining the BIC, the AO again emerged as the strongest predictor, followed by the AIM. For performance attrition, the TAPAS emerged as the only statistically significant predictor. When taking into account the number of predictors in the model, the AO again emerges as the strongest predictor, followed by the AIM. The WPA dimensions and TAPAS were comparable to one another in their AIC estimates, but the WPA dimensions had a lower BIC estimate.

Finally, a different pattern of results emerges for medical attrition. Across all three metrics (deviance, AIC, BIC), the AO subtest clearly emerges as the strongest predictor of attrition over time. For previous analyses, it mostly emerged for the BIC, which penalizes the experimental measures heavily for the number of predictors in the model. The next best predictor according to the AIC was the RBI. The next best predictors based on the BIC were the AIM and WPA dimensions.

In summary, these results are mostly consistent with the results found for cumulative attrition, with a few notable exceptions. First, when taking into account the number of parameters contributing to the model using the BIC, the AO provides the best fit to the data, followed by the AIM. The fact that AO consistently emerges as the best predictor, regardless of attrition type, suggests that the BIC may be overcorrecting for the number of parameters in the model. However, these results demonstrate that cognitive ability, in the form of AO, predicts attrition over time beyond Education Tier only. Using less stringent criteria (-2LL and AIC), the three temperament measures (RBI, TAPAS, and AIM) emerged as the strongest predictors of attrition over time relative to the WPA. In particular, the TAPAS emerged as a strong predictor of moral character and performance attrition, while the RBI emerged as a strong predictor of overall attrition. However, the WPA also emerged as a relatively strong predictor of medical attrition.

Table 7.7. Event History Analysis Assessing the Goodness-of-Fit of Non-Nested Experimental Predictor Models Through 42 Months of Service

Predictor	Deviance (-2LL)	*k* parameters	*AIC*	*BIC*
Overall Attrition				
AO [1]	**4,330.30**	17	4,364.30	4,436.91
AIM [6]	**4,320.82**	22	4,364.82	4,458.78
TAPAS [12]	**4,318.48**	28	4,374.48	4,494.07
RBI [14]	**4,306.22**	30	4,366.22	4,494.35
WPA Dimensions [6]	**4,330.31**	22	4,374.31	4,468.27
WPA Facets [14]	**4,320.30**	30	4,380.30	4,508.43
Moral Character Attrition				
AO [1]	**1,917.32**	17	1,951.32	2,006.25
AIM [6]	**1,902.67**	22	1,946.67	2,017.75
TAPAS [12]	**1,893.00**	28	1,949.00	2,039.47
RBI [14]	1,903.58	30	1,963.58	2,060.51
WPA Dimensions [6]	1,918.42	22	1,962.42	2,033.50
WPA Facets [14]	1,907.42	30	1,967.42	2,064.35
Performance Attrition				
AO [1]	762.50	17	796.50	835.90
AIM [6]	765.20	22	809.20	860.19
TAPAS [12]	**755.47**	28	811.47	876.36
RBI [14]	766.20	30	826.20	895.72
WPA Dimensions [6]	766.72	22	810.72	861.70
WPA Facets [14]	755.06	30	815.06	884.59
Medical Attrition				
AO [1]	**2,425.91**	17	2,459.91	2,521.82
AIM [6]	**2,523.00**	22	2,567.00	2,647.12
TAPAS [12]	**2,512.69**	28	2,568.69	2,670.66
RBI [14]	**2,497.88**	30	2,557.88	2,667.14
WPA Dimensions [6]	**2,520.08**	22	2,564.08	2,644.20
WPA Facets [14]	**2,505.58**	30	2,565.58	2,674.84

Note. As described in the text, results are limited to the "best bet" experimental measures for predicting overall attrition. AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, LL = Log Likelihood. Soldiers that attrited for reasons other than the target type were censored for the purpose of this analysis. Results are limited to Regular Army, non-prior service Soldiers with complete data for all six predictors. The deviance statistics in bold are incrementally statistically significant (beyond the time parameters and Education Tier). "*k* parameters" includes 14 time parameters, Education Tier, the number of scales constituting the experimental measure, and an error term.

Predicting Soldier Continuance

Approach

The final set of analyses examined whether the experimental Army Class measures could predict key antecedents of future continuance, namely Army-related attitudes and intentions to remain in the Army, beyond what is afforded by Education Tier. These retention-related criteria were chosen based on previous research showing the Soldier attitudes and intentions that were most predictive of attrition and first-term re-enlistment behavior (e.g., Lytell & Drasgow, 2009;

Strickland, 2005) and were measured by scales administered in the in-unit 1 and in-unit 2 ALQs (see Chapter 2). The attitudes selected were as follows:

1. Affective Commitment
2. Career Intentions
3. Attrition Cognitions
4. Reenlistment Intentions
5. Perceived Army Fit

With the exception of the reenlistment intentions scale, all of these measures have been used previously as retention-related criteria in the Army Class research program (Knapp & Heffner, 2009; 2010). To examine the experimental measures' predictive potential using these criteria, we computed a hierarchical OLS regression where the criterion of interest was regressed on (a) Education Tier in Step 1 (Education Tier Only) and (b) Education Tier and the scores for the predictor measure in Step 2 (Education Tier + Predictor). The difference in multiple correlations (ΔR) between the two steps was used to evaluate the incremental validity of the experimental measure. These analyses were repeated while adjusting the multiple correlations for shrinkage. See Chapter 6 for a more detailed description of this analytic approach.

Findings

Results of the incremental validity analyses for the in-unit 1 retention-related criteria are reported in Table 7.8. In examining the uncorrected predictive validity estimates, we found that Education Tier generally does not predict any of the self-report criteria analyzed here. Only one of the 40 coefficients (2.55%) was statistically significant, which is lower than the number we would expect by chance with a *p*-value of .05. In general, the experimental measures predicted these retention-related criteria well. Only three measures did not predict significant variance in one or more of these criteria: (a) AO, which only predicted attrition cognitions; (b) PSJT, which did not add to the prediction of career intentions; and (c) TAPAS, which yielded promising results for both attrition cognitions and reenlistment intentions (ΔR = .17 and ΔR = .13), but the sample sizes were much smaller and the validity estimates were non-significant. Across these retention-related criteria, the best predictors were the RBI (ΔR = .19 to .26, average ΔR = .22), the WPA (ΔR = .09 to .21, average ΔR at dimension level = .15, at facet level = .18), the TAPAS (ΔR = .13 to .24, average ΔR = .17), the AKA (ΔR = .08 to .17, average ΔR = .13), and the AIM (ΔR = .13 to .19, average ΔR = .16). The PSJT also predicted non-trivial variance in these criteria. The pattern of findings does not change much when these estimates are adjusted for shrinkage. The best predictors overall were still the RBI (ΔR = .16 to .23, average ΔR = .19), AIM (ΔR = .10 to .18, average ΔR = .13), WPA (ΔR = .07 to .18, average ΔR = .14 for both dimension and facet levels), and AKA (ΔR = .05 to .16, average ΔR = .11). The TAPAS (ΔR = .03 to .19, average ΔR = .08) experiences the largest decrease as a result of the shrinkage adjustment due to the relatively large number of scales that contribute to the model (12) and the small sample size (n = 551). The pattern of uncorrected results for AO and PSJT holds for the adjusted versions as well.

Table 7.8. Incremental Validity Estimates for Experimental Predictors over the Education Tier for Predicting In-Unit 1 Retention-Related Criteria

		Uncorrected			Corrected		
Predictor	*N*	Education Tier Only	Education Tier + Predictor	Δ*R*	Education Tier Only	Education Tier + Predictor	Δ*R*
Affective Commitment (ALQ)							
AO [1]	1,302	.01	.01	.00	.00	.00	.00
AIM [6]	562	.04	**.21**	**.17**	.00	.16	.16
TAPAS [12]	551	.06	**.21**	**.16**	.02	.11	.08
PSJT [1]	707	.03	**.12**	**.09**	.00	.09	.09
RBI [14]	1,145	.02	**.27**	**.26**	.00	.23	.23
AKA [6]	1,311	.02	**.19**	**.17**	.00	.16	.16
WPA Dimensions [6]	1,308	.01	**.20**	**.19**	.00	.18	.18
WPA Facets [14]	1,307	.01	**.23**	**.21**	.00	.18	.18
Career Intentions (ALQ)							
AO [1]	1,302	.04	.04	.01	.01	.01	-.01
AIM [6]	562	**.10**	**.24**	**.13**	.09	.18	.10
TAPAS [12]	551	.08	**.21**	**.13**	.06	.10	.04
PSJT [1]	707	.03	.06	.03	.00	.01	.01
RBI [14]	1,145	.05	**.24**	**.20**	.03	.19	.16
AKA [6]	1,310	.04	**.15**	**.12**	.02	.12	.10
WPA Dimensions [6]	1,308	.04	**.21**	**.18**	.02	.19	.18
WPA Facets [14]	1,307	.04	**.23**	**.19**	.02	.18	.16
Attrition Cognitions (ALQ)							
AO [1]	1,302	.03	**.11**	**.08**	.00	.10	.09
AIM [6]	562	.04	**.19**	**.15**	.00	.12	.12
TAPAS [12]	551	.01	.18	.17	.00	.05	.05
PSJT [1]	707	.04	**.10**	**.06**	.00	.07	.07
RBI [14]	1,145	.04	**.23**	**.20**	.01	.18	.17
AKA [6]	1,311	.02	**.10**	**.08**	.00	.05	.05
WPA Dimensions [6]	1,308	.03	**.11**	**.09**	.00	.07	.07
WPA Facets [14]	1,307	.03	**.16**	**.13**	.00	.09	.09
Reenlistment Intentions (ALQ)							
AO [1]	1,302	.01	.02	.01	.00	.00	.00
AIM [6]	562	.08	**.22**	**.14**	.06	.16	.10
TAPAS [12]	551	.06	.19	.13	.03	.07	.03
PSJT [1]	707	.05	**.11**	**.06**	.03	.08	.06
RBI [14]	1,145	.02	**.22**	**.19**	.00	.16	.16
AKA [6]	1,311	.01	**.14**	**.13**	.00	.10	.10
WPA Dimensions [6]	1,308	.01	**.17**	**.15**	.00	.13	.13
WPA Facets [14]	1,307	.01	**.19**	**.18**	.00	.13	.13
Army Fit (ALQ)							
AO [1]	1,302	.05	.06	.01	.03	.03	.00
AIM [6]	562	.04	**.24**	**.19**	.00	.18	.18
TAPAS [12]	551	.04	**.28**	**.24**	.00	.19	.19
PSJT [1]	707	.06	**.14**	**.09**	.03	.12	.09
RBI [14]	1,145	.04	**.27**	**.23**	.02	.23	.21
AKA [6]	1,311	.04	**.19**	**.15**	.03	.16	.14
WPA Dimensions [6]	1,308	.04	**.19**	**.14**	.03	.16	.13
WPA Facets [14]	1,307	.04	**.22**	**.17**	.03	.17	.14

Note. Ed Tier = Education Tier. Ed Tier + Predictor = Multiple correlation (*R*) between Education Tier and selected predictor measure with the criterion. Δ*R* = Increment in *R* over Education Tier from adding the selected predictor measure to the regression model ((Ed Tier + Predictor) – (Ed Tier Only)). Estimates in bold are statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Listwise deletion was used to account for missing data. Numbers in the Adjusted columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$. Negative estimates were set to .00.

Results of the incremental validity analyses for in-unit 2 retention-related criteria are reported in Table 7.9. As with the in-unit 1 results, Education Tier did not significantly predict variance in the self-report criteria. By contrast, all of the experimental measures demonstrated significant incremental validity in predicting these criteria. Overall, the strongest predictors were the RBI (ΔR = .20 to .26, average ΔR = .23) and the TAPAS (ΔR = .18 to .23, average ΔR = .21), followed by the WPA (ΔR = .07 to .21, average ΔR at dimension level = .15, average ΔR at facet level = .17), AKA (ΔR = .08 to .20, average ΔR = .14), and AIM (ΔR = .08 to .16, average ΔR = .14). Both the PSJT (ΔR = .02 to .13, average ΔR = .08) and AO (ΔR = .02 to .07, average ΔR = .05) also emerged as significant predictors across most criteria in this sample. After adjusting for shrinkage, the RBI (ΔR = .12 to .20, average ΔR = .17) still emerged as the strongest predictor of the self-reported criteria, followed by the WPA (ΔR = .00 to .17, average ΔR = .10-.11), AKA (ΔR = .02 to .17, average ΔR = .10), and TAPAS (ΔR = .05 to .12, average ΔR = .09). The AIM, PSJT, and AO also predicted non-trivial variance in the criteria after adjusting for shrinkage.

In summary, these results suggest that the experimental measures generally predict key self-reported continuance criteria extremely well. Affective commitment to the Army was predicted quite well by the RBI, WPA, and AKA in both the in-unit 1 and in-unit 2 samples. The career intentions scale was also predicted well by those three measures, as well as by the AIM and TAPAS. The RBI, WPA, and AIM also predicted reenlistment intentions, while the TAPAS predicted reenlistment intentions at time 2 and AKA at time 1. Army fit was strongly predicted by all experimental measures save AO. Finally, multiple experimental measures predicted attrition cognitions at time 1, but only four held at time 2 (TAPAS, PSJT, RBI, AKA). The PSJT and AO tended to predict these criteria at a higher rate for the in-unit 2 sample than the in-unit 1 sample, though the magnitude of the effects remained generally lower than the attitudinal and person-environment fit measures. The differences between the in-unit 1 and in-unit 2 results could be attributable to either (a) sample-specific attitudinal differences or (b) maturation in the overall sample.

Summary

These results suggest that all of the experimental measures meaningfully predict Soldier attrition through their first 3 years in service and in-unit retention intentions beyond Education Tier. However, which experimental measures evidenced the most predictive potential varied by time period, type of attrition, and criterion measure. Overall, the measures that emerged as the strongest predictors most consistently across all of the analyses were the three temperament measures (RBI, TAPAS, AIM). The RBI, in particular, also emerged as a strong predictor of self-reported continuance criteria. The WPA demonstrated significant incremental validity over and above Education Tier, albeit at a lower magnitude than the temperament measures. The AO emerged as a strong predictor of attrition over time. This was particularly true for medical attrition. The AKA also contributed significantly to many of the models, particularly those that included moral character attrition as the criterion. Finally, while the PSJT generally did not predict actual attrition, it did predict non-trivial variance in many of the self-reported continuance criteria, most notably Army affective commitment and perceived Army fit.

Table 7.9. Incremental Validity Estimates for Experimental Predictors over the Education Tier for Predicting In-Unit 2 Retention-Related Criteria

		Uncorrected			Corrected		
Predictor	*N*	Education Tier Only	Education Tier + Predictor	Δ*R*	Education Tier Only	Education Tier + Predictor	Δ*R*
Affective Commitment (ALQ)							
AO [1]	873	.00	.07	**.07**	.00	.04	.04
AIM [6]	423	.07	.18	.11	.03	.09	.06
TAPAS [12]	426	.00	.20	.20	.00	.05	.05
PSJT [1]	428	.03	**.15**	**.12**	.00	.12	.12
RBI [14]	744	.01	**.25**	**.24**	.00	.17	.17
AKA [6]	871	.01	**.21**	**.20**	.00	.17	.17
WPA Dimensions [6]	876	.00	**.20**	**.20**	.00	.17	.17
WPA Facets [14]	876	.00	.22	.21	.00	.14	.14
Career Intentions (ALQ)							
AO [1]	873	.01	.08	**.07**	.00	.05	.05
AIM [6]	423	.04	**.19**	**.15**	.00	.11	.11
TAPAS [12]	426	.01	**.23**	**.22**	.00	.10	.10
PSJT [1]	428	.04	.08	.03	.00	.01	.01
RBI [14]	744	.01	**.24**	**.23**	.00	.16	.16
AKA [6]	871	.01	**.13**	**.12**	.00	.07	.07
WPA Dimensions [6]	876	.01	**.19**	**.17**	.00	.14	.14
WPA Facets [14]	876	.01	**.21**	**.20**	.00	.13	.13
Attrition Cognitions (ALQ)							
AO [1]	873	.03	.06	.03	.00	.02	.02
AIM [6]	423	.07	.15	.08	.03	.04	.00
TAPAS [12]	426	.05	**.23**	**.18**	.01	.10	.09
PSJT [1]	428	.01	.11	**.10**	.00	.07	.07
RBI [14]	744	.02	**.26**	**.24**	.00	.19	.19
AKA [6]	871	.03	**.15**	**.12**	.00	.10	.10
WPA Dimensions [6]	876	.02	.09	.07	.00	.00	.00
WPA Facets [14]	876	.02	.14	.12	.00	.02	.02
Reenlistment Intentions (ALQ)							
AO [1]	873	.02	.08	**.05**	.00	.05	.05
AIM [6]	423	.05	.18	**.13**	.01	.09	.08
TAPAS [12]	426	.01	**.24**	**.23**	.00	.12	.12
PSJT [1]	428	.08	.10	.02	.04	.05	.00
RBI [14]	744	.01	**.21**	**.20**	.00	.12	.12
AKA [6]	871	.02	.10	.08	.00	.02	.02
WPA Dimensions [6]	876	.03	**.14**	**.12**	.00	.09	.09
WPA Facets [14]	876	.03	**.18**	**.16**	.00	.09	.09
Army Fit (ALQ)							
AO [1]	873	.02	.04	.02	.00	.00	.00
AIM [6]	423	.04	**.20**	**.16**	.00	.12	.12
TAPAS [12]	426	.02	**.24**	**.22**	.00	.11	.11
PSJT [1]	428	.03	**.16**	**.13**	.00	.13	.13
RBI [14]	744	.01	**.27**	**.26**	.00	.20	.20
AKA [6]	871	.02	**.21**	**.18**	.00	.17	.17
WPA Dimensions [6]	876	.03	**.20**	**.17**	.00	.16	.17
WPA Facets [14]	876	.03	**.21**	**.18**	.00	.13	.14

Note. Ed Tier = Education Tier. Ed Tier + Predictor = Multiple correlation (*R*) between Education Tier and selected predictor measure with the criterion. Δ*R* = Increment in *R* over Education Tier from adding the selected predictor measure to the regression model ((Ed Tier + Predictor) – (Ed Tier Only)). Estimates in bold are statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Listwise deletion was used to account for missing data. Numbers in the Adjusted columns were adjusted for shrinkage using Burket's (1964) formula $\rho_c = (NR^2 - k)/[R(N - k)]$. Negative estimates were set to .00.

CHAPTER 8: EVALUATING CLASSIFICATION POTENTIAL

Matthew Trippe, Michael Ingerick, and Ted Diaz (HumRRO)

Overview and Background

In addition to examining the experimental predictor measures' potential for screening Army applicants, we evaluated their potential for improving new Soldier classification into entry-level MOS. Previous research suggests that several of these experimental predictor measures could significantly enhance new Soldier classification beyond the existing ASVAB, particularly if the Army's goal is to maximize first-term Soldier retention (Ingerick et al., 2009).

Approach to Estimating the Classification Potential of the Experimental Predictors

Similar to previous research (Ingerick et al., 2009), we evaluated the classification potential of the experimental predictor measures using (a) Horst's (1954, 1955) index of differential validity (H_d) and (b) mean predicted criterion score (MPCS). Conceptually, H_d provides an index of the predictor measure(s)' ability to differentiate among the predicted criterion scores for a sample of jobs. The greater the H_d value, the larger the cross-job differences in the predicted criterion scores. Analytically, H_d represents the average standardized mean difference between all possible pairs of predicted criterion scores for a sample of jobs. Conversely, the mean predicted criterion score (MPCS) reflects the average predicted criterion score for Soldiers classified into a sample of jobs using the predictor measure(s). The greater the MPCS, the higher Soldiers are predicted to perform or be satisfied, on average, when classified into a sample of jobs using the selected predictor measure(s). Although the two indices are related (i.e., larger H_d values tend to be associated with higher MPCS values), each captures unique information about the classification potential of the predictor measure(s). Whereas H_d provides information on cross-job differences (or variability) in Soldiers' predicted criterion scores resulting from the use of the predictor measure(s) to classify Soldiers into a sample of jobs, the MPCS supplies information on the average level at which Soldiers are predicted to score on the targeted criterion (e.g., performance, retention). For example, a measure that predicts criterion scores equally well for two jobs can have a high overall MPCS but a low H_d because prediction does not vary between jobs. Similarly, a measure that predicts very well for one job but not as well for another may have a high H_d value because of the variability but a lower overall MPCS because of the poorer prediction in the second job.

Comparable to the incremental predictive validity analyses, we estimated the increment in H_d and MPCS resulting from using the experimental predictor measures over ASVAB to classify new Soldiers to a selected sample of MOS. We investigated the measures' potential for enhancing both performance and retention-related criteria at two different points in time (at the end of training and in-unit). Unlike the selection-oriented results, classification potential using training criterion data have not been previously reported. In-unit 2 were not analyzed because the MOS-specific sample sizes were insufficient.

Table 8.1 summarizes the criterion measures used in these analyses, organized by type and time period. For a selected subset of these criterion measures, we had sufficient data to analyze the predictor measures' potential for classifying new Soldiers to an expanded sample of MOS in addition to the six target MOS (11B, 19K, 31B, 68W, 88M, and 91B). Absent from Table 8.1 are any criterion variables collected during the in-unit 2 phase of the project. The only MOS with reasonable sample sizes (11B, 19K, and 31B) are relatively homogeneous with regard to occupational requirements. Since analyzing a set of two to three somewhat similar occupations would not produce meaningful results, we did not perform classification analyses using in-unit 2 criterion data.

Table 8.1. Criterion Measures Used in Classification Potential Analyses

Criterion Type	When Collected	Criterion Measure
Performance	Training	• MOS-Specific JKT • MOS-Specific PRS Composite • Number of Times Restarted Through AIT/OSUT [a]
	In-Unit	• MOS-Specific JKT • MOS-Specific PRS Composite • Army-Wide PRS – Performing MOS-Specific Tasks [a]
Retention	Training	• Perceived MOS Fit (ALQ)
	In-Unit	• Perceived MOS Fit (ALQ)[a] • MOS Satisfaction (ALQ)[a]

[a]Denotes those criterion measures for which there were sufficient data to analyze the predictor measures' classification potential for an expanded sample of MOS.

Our analysis approach consisted of the following steps:

1. Estimate the observed (uncorrected) covariance matrix for each MOS.

2. Apply a ridge adjustment (a small constant multiplied by the diagonal and then added to the matrix) to the matrices estimated in Step 1 (as appropriate) to ensure that the matrices are positive definite.

3. Correct the predictor-criterion covariances and predictor covariances from Steps 1 and 2 for multivariate range restriction on the ASVAB (Lawley, 1943). Data on FY 2004 Army accessions were used as the reference population when making these corrections.

4. Using the corrected covariance matrices from Step 3, compute two indices of classification potential: (a) (H_d) and (b) MPCS (DeCorte, 2000).

Observed covariance matrices estimated in Step 1 were computed using pairwise deletion. Using listwise deletion would have resulted in a significant loss of data and severely restricted the analyses to the point of being of little practical value. The primary disadvantage to

this approach is that estimating covariance matrices from pairwise data can result in matrices that are irregular or not positive definite. The correction for multivariate range restriction in Step 3 requires the observed matrices to be positive definite. Accordingly, a ridge adjustment was applied in cases where the observed matrices were found to not be positive definite. This adjustment involved introducing a small constant (.01) that generally retains the properties of the original matrix but produces a matrix that is positive definite (Joreskog & Sorbom, 1996).

Results

Tables 8.2 through 8.9 summarize the results of the experimental predictor measures' classification potential, as measured by H_d and MPCS. Several factors should be kept in mind when interpreting these results. First, our analyses did not model important organizational factors and other operational constraints that contribute to the Soldier-job matching process under the Army's existing classification system (e.g., demand for certain MOS, availability of training seats at the time of accession). As a result, the estimates reported reflect the *potential* of the predictor measures to enhance new Soldier classification and not the *actual* expected gains in classification if the measures were used operationally. Second, the results reported could differ if a different sample of MOS or set of criterion measures were examined. Consistent with previous research, we expected MOS-specific criteria to afford the predictor measures the greatest opportunity to show their classification potential. This is because criteria whose content (or frame-of-reference) more strongly matches an MOS are potentially more sensitive to differential validity than Army-wide criteria. Accordingly, we focused our analyses on a targeted set of MOS-specific criteria. Third, there are presently no standards or conventions for interpreting the magnitude of or gain in H_d relative to some baseline. Consequently, previous research involving the same or comparable experimental predictor measures provides the best basis for making relative comparisons about the magnitude or gain in H_d observed in the current research. With regard to MPCS, there is some evidence that increments in MPCS as low as .10 represent significant and practical gains (Nord & Schmitz, 1991). Past research examining the Project A experimental predictor measures found increments in MPCS ranging from .05 to .10 when the selected experimental predictors were combined with the ASVAB to maximize a performance-based criterion (Rosse, Campbell, & Peterson, 2001; Scholarios, Johnson, & Zeidner, 1994). The concurrent validation phase of the Army Class project found average increments in H_d ranging from .14 to 1.57 and average MPCS increments ranging from .05 to .44 across all criteria (Ingerick et al., 2009).

We first present the results summarizing the predictor measures' potential for maximizing the selected performance-related criteria for the six target MOS at two different points in time (end of training and in-unit) followed by the results for the retention-related criteria. We then present the results for a selected subset of criterion measures for which we had sufficient data from an expanded sample of MOS.

Maximizing Performance-Related Criteria

Tables 8.2 and 8.3 report the experimental predictor measures' potential to enhance new Soldier classification over the existing ASVAB where the goal is to maximize MOS-specific performance criteria. Examination of Tables 8.2 and 8.3 evidences the following:

Table 8.2. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted MOS-Specific Job Knowledge Across and Within MOS

	Overall				MOS											
					11B		19K		31B		68W		88M		91B	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
Training MOS-Specific Job Knowledge Test (JKT)																
ASVAB	0.04	--	0.17	--	-0.02	--	-0.24	--	0.39	--	0.22		--	--	0.91	--
TAPAS [12]	0.12	0.07	0.32	0.15	0.20	0.22	0.35	0.58	0.22	-0.17	0.68	0.45	--	--	0.93	0.02
RBI [14]	0.12	0.08	0.3	0.13	0.03	0.05	0.27	0.51	0.35	-0.04	0.98	0.76	--	--	1.01	0.10
AKA [6]	0.06	0.02	0.23	0.06	0.09	0.11	0.01	0.25	0.28	-0.11	0.44	0.21	--	--	0.88	-0.03
WPA-F [14]	0.13	0.09	0.31	0.14	0.15	0.17	0.07	0.30	0.28	-0.11	0.99	0.77	--	--	1.11	0.21
WPA-D [6]	0.09	0.05	0.24	0.08	0.05	0.07	-0.02	0.22	0.32	-0.07	0.75	0.53	--	--	1.01	0.11
AIM [6][a]	0.08	0.03	0.20	0.04	-0.04	0.04	--	--	0.33	0.03	0.34	0.21	--	--	0.84	0.02
AO [1]	0.05	0.01	0.18	0.01	-0.01	0.01	-0.25	-0.01	0.35	-0.04	0.57	0.35	--	--	0.93	0.02
In-Unit MOS-Specific Job Knowledge Test (JKT)																
ASVAB	0.08	--	0.28	--	0.08	--	-0.04	--	0.46	--	0.54		--	--	0.98	--
TAPAS [12]	0.31	0.23	0.54	0.26	0.29	0.21	1.24	1.27	0.38	-0.08	0.97	0.43	--	--	1.01	0.03
RBI [14]	0.28	0.20	0.51	0.23	0.27	0.19	0.46	0.49	0.52	0.06	1.40	0.85	--	--	1.21	0.22
AKA [6]	0.13	0.05	0.34	0.06	0.13	0.05	-0.01	0.03	0.45	-0.02	0.94	0.39	--	--	1.22	0.24
WPA-F [14]	0.24	0.16	0.43	0.15	0.10	0.02	0.21	0.24	0.58	0.12	1.48	0.94	--	--	1.29	0.31
WPA-D [6]	0.15	0.07	0.35	0.07	0.11	0.03	0.05	0.09	0.46	0.00	1.07	0.52	--	--	1.20	0.22
AIM [6][a]	0.12	0.06	0.32	0.09	0.02	0.06	--	--	0.45	0.08	0.70	0.27	--	--	1.10	0.17
AO [1]	0.09	0.01	0.30	0.02	0.14	0.06	0.01	0.05	0.40	-0.06	0.71	0.17	--	--	0.93	-0.05

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 31B =212, 68W = 39, 88M = 61, 91B = 65.

[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

Table 8.3. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted MOS-Specific Performance Ratings Across and Within MOS

	Overall				MOS											
					11B		19K		31B		68W		88M		91B	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
Training MOS-Specific PRS Composite																
ASVAB	0.03	--	0.18	--	0.06	--	0.22	--	0.12	--	0.78	--	--	--	0.49	--
TAPAS [12]	0.10	0.06	0.30	0.13	0.10	0.05	0.41	0.19	0.32	0.20	1.03	0.25	--	--	0.59	0.10
RBI [14]	0.13	0.09	0.33	0.16	0.16	0.11	0.26	0.04	0.30	0.18	1.19	0.41	--	--	0.83	0.34
AKA [6]	0.04	0.01	0.20	0.02	0.10	0.04	0.21	-0.02	0.13	0.00	0.88	0.10	--	--	0.49	-0.01
WPA-F [14]	0.12	0.08	0.30	0.12	0.10	0.05	0.40	0.18	0.20	0.08	1.17	0.39	--	--	0.86	0.36
WPA-D [6]	0.08	0.05	0.23	0.06	0.09	0.04	0.24	0.02	0.15	0.03	1.10	0.32	--	--	0.65	0.16
AIM [6][a]	0.07	0.04	0.24	0.09	0.07	0.02	--	--	0.28	0.17	0.93	0.17	--	--	0.55	0.08
AO [1]	0.04	0.00	0.18	0.00	0.05	-0.01	0.24	0.02	0.14	0.01	0.75	-0.03	--	--	0.51	0.02
In-Unit MOS-Specific PRS Composite																
ASVAB	0.10	--	0.33	--	0.37	--	0.48	--	0.15	--	--	--	0.56		0.27	--
TAPAS [12]	0.41	0.31	0.70	0.37	0.39	0.02	1.35	0.87	0.67	0.52	--	--	1.07	0.51	0.97	0.70
RBI [14]	0.28	0.18	0.55	0.22	0.42	0.06	0.61	0.14	0.44	0.29	--	--	1.30	0.74	0.66	0.39
AKA [6]	0.17	0.07	0.43	0.10	0.36	-0.01	0.53	0.05	0.30	0.14	--	--	0.74	0.18	0.76	0.49
WPA-F [14]	0.26	0.16	0.53	0.20	0.41	0.04	0.67	0.19	0.34	0.19	--	--	1.13	0.57	0.90	0.63
WPA-D [6]	0.19	0.09	0.46	0.14	0.39	0.02	0.50	0.03	0.33	0.17	--	--	0.96	0.40	0.75	0.48
AIM [6][a]	0.20	0.12	0.40	0.14	0.27	-0.04	--	--	0.37	0.26	--	--	0.83	0.32	0.67	0.43
AO [1]	0.12	0.02	0.37	0.05	0.38	0.02	0.52	0.04	0.19	0.04	--	--	0.54	-0.02	0.55	0.29
In-Unit Army-Wide PRS – Performing MOS-Specific Tasks																
ASVAB	0.11	--	0.35	--	0.25	--	0.44	--	0.34	--	--	--	0.45		0.60	--
TAPAS [12]	0.39	0.28	0.63	0.29	0.26	0.01	1.40	0.96	0.58	0.24	--	--	1.11	0.66	1.04	0.43
RBI [14]	0.31	0.20	0.58	0.23	0.32	0.07	0.80	0.36	0.53	0.19	--	--	1.19	0.74	1.09	0.49
AKA [6]	0.15	0.04	0.39	0.05	0.24	-0.01	0.50	0.07	0.35	0.01	--	--	0.63	0.18	0.85	0.24
WPA-F [14]	0.24	0.13	0.51	0.17	0.31	0.06	0.65	0.21	0.44	0.10	--	--	1.00	0.55	1.06	0.46
WPA-D [6]	0.18	0.07	0.44	0.09	0.29	0.04	0.56	0.13	0.41	0.07	--	--	0.88	0.43	0.67	0.07
AIM [6][a]	0.16	0.06	0.36	0.08	0.14	-0.05	--	--	0.52	0.23	--	--	0.68	0.28	0.56	0.04
AO [1]	0.13	0.02	0.37	0.02	0.25	0.00	0.46	0.03	0.33	-0.01	--	--	0.48	0.03	0.86	0.25

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 31B =212, 68W = 39, 88M = 61,91B = 65.

[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

The experimental non-cognitive predictor measures exhibited non-trivial classification gains, on average, for the target MOS sampled. The average increments in H_d and MPCS for the non-cognitive experimental predictor measures over the ASVAB were .06 and .10, respectively, for the training MOS-specific JKT and .13 and .14 for the in-unit JKT. The average increments in H_d and MPCS for the MOS-specific performance ratings criteria were similarly .06 and .10, respectively, for the training PRS scores. The classification gains were higher, on average, for the in-unit ratings (average ΔH_d = .16, average ΔMPCS = .20 for the in-unit MOS-specific ratings composite; average ΔH_d = .13, average ΔMPCS = .15 for the in-unit AW performing MOS-specific tasks rating scale) than the training ratings. This pattern of results was consistent with findings from previous research that non-cognitive measures demonstrate greater classification potential for maximizing behaviorally-based performance criteria (i.e., ratings of what Soldiers do) than knowledge-based performance criteria (i.e., tests of what Soldiers know) (Ingerick et al., 2009). At the MOS-level, the non-cognitive predictor measures showed the greatest classification gains, on average, for 19K, 68W, and 88M (for the MOS-specific performance ratings). Regarding a cognitive test, AO produced gains in H_d and MPCS that were consistently less than .05. Although this finding could be attributed to the fact that the results for the non-cognitive predictor measures were partly inflated by sampling error (i.e., due to low sample size and the greater number of scores entering into the estimation), AO performed similarly in previous analyses using a simulation-based cross-validation design that accounted for the effects of sampling error on H_d and MPCS (Ingerick et al., 2009).

Among the non-cognitive predictor measures, the TAPAS and RBI emerged as the measures with the greatest potential to supplement the ASVAB, followed by the WPA. On average, the TAPAS and the RBI produced the greatest increments in H_d and MPCS over the ASVAB. For the can-do performance criteria (the MOS-specific JKTs), the gains in H_d associated with the TAPAS ranged from .07 to .23 and .15 to .26 in the MPCS index. Gains in H_d associated with the RBI ranged from .08 to .20 and MPCS gains ranged from .15 to .23. Consistent with the overall pattern of findings from Army Class, the gains in H_d produced by the TAPAS and the RBI over the ASVAB were generally higher, on average, for the will-do performance criteria (the MOS-specific performance ratings) than can-do criteria. For H_d, the increments associated with the TAPAS ranged from .06 to .31 and .13 to .37 in the MPCS index. Gains in H_d associated with the RBI ranged from .09 to .18 and MPCS gains ranged from .16 to .22. Across the different performance criteria, the TAPAS and RBI showed comparable classification gains, with the exception of the in unit MOS-specific ratings composite, with which the TAPAS exhibited greater gains than the RBI. Among the other non-cognitive predictors, the WPA (facets) demonstrated the greatest classification potential after the TAPAS and RBI (ΔH_d = .08-.16, ΔMPCS = .12-.20). Interestingly, the relative rank ordering of the non-cognitive experimental predictors based on their classification potential varied significantly by MOS. This finding suggests that the predictors vary in the extent to which they reflect dimensions of performance relevant for each MOS.

Maximizing Retention-Related Criteria

Tables 8.4 and 8.5 report the experimental predictor measures' potential to enhance new Soldier classification over the existing ASVAB for the purposes of maximizing retention-related criteria. Examination of Tables 8.4 and 8.5 shows the following:

Table 8.4. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted Retention-Related Outcomes Averaged Across and Within MOS

	Overall				MOS											
					11B		19K		31B		68W		88M		91B	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
Training Perceived MOS Fit (ALQ)																
ASVAB	0.03	--	0.16	--	0.07	--	0.17	--	0.08	--	0.61		0.19	--	0.48	--
TAPAS [12]	0.24	0.21	0.44	0.28	0.23	0.16	0.53	0.37	0.17	0.09	0.85	0.24	1.61	1.42	0.81	0.34
RBI [14]	0.09	0.06	0.30	0.15	0.18	0.11	0.25	0.08	0.26	0.18	1.04	0.42	0.42	0.24	0.50	0.02
AKA [6]	0.05	0.02	0.22	0.06	0.10	0.03	0.30	0.14	0.12	0.04	0.72	0.11	0.33	0.14	0.53	0.05
WPA-F [14]	0.10	0.07	0.32	0.17	0.29	0.21	0.18	0.02	0.15	0.06	0.86	0.25	0.61	0.42	0.67	0.20
WPA-D [6]	0.07	0.03	0.25	0.09	0.24	0.17	0.15	-0.02	0.09	0.01	0.78	0.16	0.43	0.24	0.45	-0.02
AIM [6][a]	0.12	0.08	0.28	0.14	0.18	0.12	--	--	0.12	0.05	0.73	0.12	0.94	0.75	0.40	-0.06
AO [1]	0.04	0.00	0.16	0.01	0.07	0.00	0.18	0.02	0.09	0.00	0.62	0.00	0.19	0.00	0.54	0.06
In-Unit Perceived MOS Fit (ALQ)																
ASVAB	0.07	--	0.31	--	0.17	--	0.40	--	0.37	--	0.79	--	0.19	--	0.40	--
TAPAS [12]	0.29	0.21	0.60	0.30	0.36	0.19	0.67	0.27	0.62	0.25	1.39	0.59	0.91	0.72	0.83	0.43
RBI [14]	0.24	0.17	0.51	0.21	0.21	0.04	0.57	0.17	0.52	0.14	1.33	0.54	0.86	0.67	1.10	0.70
AKA [6]	0.12	0.05	0.39	0.09	0.19	0.02	0.54	0.14	0.42	0.05	0.93	0.13	0.52	0.33	0.60	0.20
WPA-F [14]	0.21	0.14	0.51	0.20	0.24	0.07	0.62	0.22	0.53	0.15	1.23	0.44	0.71	0.52	0.90	0.50
WPA-D [6]	0.14	0.07	0.41	0.11	0.20	0.02	0.55	0.15	0.48	0.11	0.98	0.19	0.39	0.20	0.68	0.28
AIM [6][a]	0.17	0.11	0.40	0.13	0.23	0.07	--	--	0.37	0.02	1.08	0.31	0.79	0.60	0.57	0.19
AO [1]	0.08	0.01	0.32	0.01	0.17	-0.01	0.43	0.03	0.38	0.01	0.82	0.02	0.28	0.09	0.40	0.00
In-Unit MOS Satisfaction (ALQ)																
ASVAB	0.08	--	0.30	--	0.27	--	0.12	--	0.20	--	1.11	--	0.42	--	0.45	--
TAPAS [12]	0.35	0.28	0.65	0.35	0.47	0.20	0.82	0.70	0.43	0.23	1.63	0.51	0.94	0.52	1.16	0.71
RBI [14]	0.24	0.16	0.50	0.20	0.34	0.07	0.35	0.22	0.36	0.16	1.39	0.28	1.08	0.66	0.87	0.42
AKA [6]	0.12	0.05	0.37	0.07	0.30	0.03	0.25	0.13	0.25	0.04	1.12	0.01	0.70	0.28	0.57	0.12
WPA-F [14]	0.23	0.16	0.52	0.22	0.30	0.03	0.67	0.55	0.38	0.18	1.49	0.38	0.69	0.27	1.11	0.66
WPA-D [6]	0.18	0.10	0.43	0.13	0.30	0.04	0.40	0.28	0.23	0.03	1.43	0.32	0.56	0.14	1.01	0.56
AIM [6][a]	0.15	0.07	0.38	0.09	0.32	0.09	--	--	0.26	0.08	1.06	-0.02	0.76	0.37	0.31	-0.12
AO [1]	0.09	0.01	0.32	0.02	0.27	0.00	0.16	0.04	0.20	0.00	1.17	0.06	0.56	0.14	0.44	-0.01

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 31B =212, 68W = 39, 88M = 61, 91B = 65.

[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

Table 8.5. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted IMT Performance Outcomes Averaged Across and Within MOS

	Overall				MOS											
					11B		19K		31B		68W		88M		91B	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
	Number of Times Recycled Through AIT/OSUT															
ASVAB	0.05	--	0.18	--	0.06	--	0.45	--	0.05	--	1.00	--	0.26	--	0.25	--
TAPAS [12]	0.10	0.05	0.28	0.10	0.10	0.04	0.66	0.21	0.14	0.10	1.36	0.36	0.31	0.06	0.37	0.13
RBI [14]	0.08	0.03	0.25	0.07	0.07	0.01	0.55	0.11	0.12	0.08	1.20	0.20	0.34	0.08	0.38	0.14
AKA [6]	0.06	0.01	0.21	0.03	0.08	0.02	0.51	0.06	0.06	0.01	1.03	0.03	0.26	0.00	0.33	0.09
WPA-F [14]	0.08	0.03	0.25	0.07	0.07	0.01	0.52	0.07	0.13	0.08	1.21	0.20	0.38	0.12	0.44	0.20
WPA-D [6]	0.06	0.01	0.22	0.04	0.06	0.00	0.50	0.05	0.10	0.05	1.08	0.08	0.36	0.10	0.33	0.08
AIM [6][a]	0.06	0.02	0.21	0.07	0.08	0.02	--	--	0.13	0.08	1.00	0.02	0.45	0.20	0.43	0.19
AO [1]	0.05	0.00	0.19	0.01	0.07	0.01	0.45	0.00	0.05	0.00	1.01	0.01	0.28	0.02	0.24	0.00

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 31B =212, 68W = 39, 88M = 61, 91B = 65.

[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable but not identical to those performed on other predictors.

The classification potential of the experimental predictor measures to maximize retention-related criteria was generally comparable to that of the performance-related criteria. Average gains in H_d associated with the non-cognitive experimental predictor measures ranged from .08 and .14 for both types of criteria. Similarly, the average gains in MPCS across the different retention-related criteria ranged from .15 to .18. Comparable to the results for the performance-related criteria, there were cross-MOS differences, with the non-cognitive predictor measures showing the greatest classification gains, on average, among the non-close combat MOS (31B, 68W, 88M, and 91B).

Among the non-cognitive predictor measures, the TAPAS (ΔH_d = .21-.28, ΔMPCS = .28-.35), followed by the WPA (ΔH_d = .07-.14, ΔMPCS = .17-.22) and the RBI (ΔH_d = .06-.17, ΔMPCS = .15-.21), demonstrated the greatest classification gains over the ASVAB among the target MOS sampled. As with the performance-related criteria, however, sampling error could have inflated these statistics. The available sample sizes prohibited reserving a portion of the sample for cross-validation analyses. Nevertheless, a similar pattern of results was observed in previous analyses of the same or similar non-cognitive predictor measures using a simulation-based cross-validation design (Ingerick et al., 2009). Similar to the performance-related criteria, for the retention-related criteria, the relative rank ordering of the non-cognitive experimental predictors' classification potential varied significantly by MOS, suggesting that coverage of each MOS's performance domain varies by predictor measure.

Classification Potential among an Expanded Sample of MOS

For a selected subset of the criterion measures, sufficient criterion data were available to perform the classification analyses on an expanded sample of MOS. This expanded sample consisted of the six target MOS, plus several additional MOS. These additional MOS were selected because (a) they had sufficient criterion data on the selected measures and (b) they represented career fields or had aptitude requirements different from those covered by the six target MOS. The additional MOS were 25U (Signal Support Systems Specialist), 42A (Human Resources Specialist), 74D (Chemical, Biological, Radiological, and Nuclear [CBRN] Specialist), 92F (Petroleum Supply Specialist), and 92G (Food Service Specialist). Availability of data determined whether some or all of these five additional MOS were included in the analysis. Tables 8.6 through 8.9 report the experimental predictor measures' potential to enhance new Soldier classification based on this expanded sample of MOS. Review of Tables 8.6 through 8.9 shows the following:

The classification gains associated with the experimental predictor measures were somewhat higher, on average, for the expanded sample of MOS than the target MOS. Overall, the estimated gains in H_d and MPCS tended to be somewhat higher, albeit small in magnitude, when based on the expanded sample of MOS than the six target MOS. For example, the average increment in H_d and MPCS for the non-cognitive experimental predictor measures, based on the expanded sample and excluding AO, was .16 and .23, respectively, for the in-unit perceived MOS fit criterion. The average classification gains observed for the same criterion measure based on the six target MOS were somewhat lower at .13 (ΔH_d) and .17 (ΔMPCS). A similar pattern was found for the performance-related criteria. For instance, the average gains in H_d and MPCS based on the expanded sample were .14 and .19, respectively, for a will-do criterion measure (in-unit Performing MOS Specific Tasks rating), compared to .13 and .15 for the six target MOS. Although the cross-sample differences in classification gains were generally small, these findings illustrate the importance of the sample of MOS considered when evaluating classification potential. The findings also suggest that the experimental predictor measures have classification potential beyond the six target MOS.

Table 8.6. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted IMT Performance Outcomes (Number of Times Restarted during IMT) Averaged Across and Within MOS (Expanded Sample)

	Overall				MOS											
					11B		19K		25U		31B		42A		68W	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
ASVAB	0.05	--	0.23	--	0.06	--	0.48	--	0.36	--	0.05	--	0.12	--	1.05	--
TAPAS [12]	0.13	0.08	0.39	0.16	0.10	0.04	0.70	0.22	0.82	0.46	0.15	0.10	0.41	0.30	1.41	0.37
RBI [14]	0.10	0.05	0.35	0.11	0.09	0.03	0.60	0.11	0.59	0.23	0.14	0.09	0.30	0.18	1.25	0.20
AKA [6]	0.06	0.01	0.27	0.04	0.09	0.02	0.55	0.07	0.42	0.06	0.07	0.02	0.22	0.10	1.08	0.03
WPA-F [14]	0.10	0.05	0.35	0.11	0.08	0.02	0.56	0.08	0.60	0.24	0.14	0.09	0.28	0.16	1.25	0.21
WPA-D [6]	0.07	0.02	0.29	0.06	0.07	0.00	0.53	0.05	0.52	0.16	0.11	0.06	0.24	0.12	1.13	0.08
AIM [6][a]	0.08	0.03	0.29	0.10	0.08	0.03	--	--	--	--	0.14	0.09	0.35	0.23	1.04	0.02
AO [1]	0.05	0.00	0.24	0.01	0.07	0.00	0.48	0.00	0.37	0.01	0.06	0.00	0.14	0.02	1.06	0.01

	74D		88M		91B		92F		92G	
	MPCS	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
ASVAB	0.53	--	0.26	--	0.27	--	0.70	--	0.47	
TAPAS [12]	0.95	0.42	0.32	0.05	0.39	0.13	1.08	0.39	0.83	0.36
RBI [14]	0.77	0.24	0.36	0.10	0.41	0.14	0.94	0.25	0.83	0.36
AKA [6]	0.65	0.11	0.27	0.00	0.36	0.09	0.76	0.07	0.50	0.04
WPA-F [14]	0.89	0.36	0.40	0.13	0.46	0.19	0.94	0.25	0.89	0.43
WPA-D [6]	0.79	0.26	0.37	0.10	0.35	0.08	0.74	0.04	0.60	0.13
AIM [6][a]	0.63	0.10	0.47	0.21	0.46	0.20	0.83	0.15	0.67	0.19
AO [1]	0.63	0.10	0.29	0.03	0.27	0.00	0.71	0.02	0.51	0.05

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 25U=53, 31B =212, 42A=60, 68W = 39, 74D=16, 88M = 61, 91B = 65, 92F = 44, 92G= 30.

[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

Table 8.7. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted In-Unit Perceived MOS Fit (ALQ) Averaged Across and Within MOS (Expanded Sample)

	Overall				MOS									
					11B		19K		25U		31B		42A	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
ASVAB	0.10	--	0.38	--	0.17	--	0.40	--	0.72	--	0.38	--	0.60	--
TAPAS [12]	0.33	0.23	0.74	0.36	0.39	0.22	0.72	0.32	1.16	0.45	0.68	0.30	1.13	0.53
RBI [14]	0.29	0.20	0.66	0.28	0.23	0.06	0.60	0.20	1.12	0.40	0.54	0.17	1.06	0.46
AKA [6]	0.16	0.07	0.50	0.12	0.19	0.02	0.54	0.14	1.15	0.44	0.45	0.07	0.72	0.13
WPA-F [14]	0.30	0.20	0.66	0.28	0.26	0.09	0.64	0.24	1.44	0.73	0.53	0.15	1.28	0.68
WPA-D [6]	0.19	0.10	0.53	0.15	0.20	0.03	0.56	0.16	1.01	0.29	0.49	0.11	1.12	0.52
AIM [6][a]	0.26	0.17	0.53	0.19	0.28	0.12	--	--	--	--	0.40	0.04	0.69	0.11
AO [1]	0.10	0.01	0.40	0.02	0.17	0.00	0.43	0.03	0.77	0.05	0.39	0.02	0.59	0.00

	68W		88M		91B		92F	
	MPCS	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
ASVAB	0.82	--	0.26	--	0.40	--	0.92	--
TAPAS [12]	1.46	0.64	1.02	0.77	0.91	0.51	1.31	0.40
RBI [14]	1.40	0.58	1.02	0.76	1.21	0.81	1.35	0.43
AKA [6]	0.97	0.15	0.61	0.36	0.64	0.24	0.92	0.01
WPA-F [14]	1.29	0.47	0.79	0.53	0.96	0.56	1.28	0.36
WPA-D [6]	1.01	0.19	0.42	0.16	0.73	0.33	1.12	0.20
AIM [6][a]	1.15	0.36	0.92	0.67	0.66	0.28	1.51	0.60
AO [1]	0.85	0.03	0.36	0.10	0.41	0.01	0.93	0.01

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 25U=53, 31B =212, 42A=60, 68W = 39, 88M = 61, 91B = 65, 92F = 44.

[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

Table 8.8. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted In-Unit MOS Satisfaction (ALQ) Averaged Across and Within MOS (Expanded Sample)

	Overall				MOS									
					11B		19K		25U		31B		42A	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
ASVAB	0.09	--	0.37	--	0.26	--	0.12	--	0.79	--	0.20	--	0.59	--
TAPAS [12]	0.36	0.27	0.78	0.40	0.52	0.25	0.86	0.74	1.11	0.32	0.47	0.27	1.12	0.54
RBI [14]	0.28	0.18	0.63	0.26	0.36	0.09	0.37	0.25	1.09	0.30	0.40	0.20	1.09	0.50
AKA [6]	0.15	0.06	0.47	0.09	0.29	0.03	0.26	0.14	0.95	0.16	0.24	0.04	0.85	0.26
WPA-F [14]	0.29	0.20	0.67	0.29	0.31	0.04	0.70	0.58	1.35	0.56	0.40	0.20	1.08	0.49
WPA-D [6]	0.20	0.11	0.53	0.15	0.30	0.03	0.40	0.28	1.03	0.24	0.23	0.03	0.96	0.37
AIM [6][a]	0.21	0.12	0.49	0.14	0.35	0.13	--	--	--	--	0.28	0.09	0.53	-0.03
AO [1]	0.11	0.02	0.40	0.02	0.26	0.00	0.16	0.03	0.85	0.06	0.21	0.01	0.60	0.01

	68W		88M		91B		92F	
	MPCS	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
ASVAB	1.15	--	0.48	--	0.39	--	0.86	--
TAPAS [12]	1.70	0.55	1.03	0.55	1.19	0.80	1.22	0.37
RBI [14]	1.45	0.30	1.21	0.73	0.91	0.52	1.20	0.35
AKA [6]	1.15	0.00	0.79	0.31	0.56	0.18	0.96	0.10
WPA-F [14]	1.54	0.39	0.80	0.32	1.13	0.74	1.38	0.52
WPA-D [6]	1.48	0.33	0.61	0.13	1.01	0.62	0.98	0.12
AIM [6][a]	1.11	0.01	0.91	0.46	0.34	-0.06	1.52	0.71
AO [1]	1.20	0.05	0.65	0.16	0.38	-0.01	0.93	0.07

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 25U=53, 31B =212, 42A=60, 68W = 39, 88M = 61, 91B = 65, 92F = 44.

[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

Table 8.9. Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted MOS-Specific Performance Ratings Averaged Across and Within MOS (Expanded Sample)

	Overall				MOS*									
					11B		19K		25U		31B		42A	
	H_d	ΔH_d	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
In-Unit Army-Wide PRS – Performing MOS-Specific Tasks														
ASVAB	0.12	--	0.41	--	0.28	--	0.45	--	0.36	--	0.39	--	0.75	--
TAPAS [12]	0.45	0.33	0.79	0.38	0.31	0.04	1.48	1.02	1.24	0.88	0.65	0.26	1.19	0.44
RBI [14]	0.37	0.25	0.73	0.32	0.36	0.08	0.83	0.38	1.13	0.77	0.57	0.19	1.30	0.55
AKA [6]	0.16	0.04	0.48	0.06	0.28	0.00	0.53	0.08	0.71	0.35	0.40	0.01	0.82	0.08
WPA-F [14]	0.33	0.21	0.68	0.27	0.34	0.06	0.67	0.21	1.33	0.97	0.50	0.12	1.49	0.74
WPA-D [6]	0.22	0.11	0.57	0.16	0.31	0.03	0.58	0.12	1.08	0.72	0.45	0.06	1.23	0.48
AIM [6][a]	0.25	0.13	0.49	0.14	0.19	-0.03	--	--	--	--	0.58	0.25	0.86	0.18
AO [1]	0.14	0.02	0.44	0.02	0.28	0.00	0.48	0.03	0.39	0.03	0.37	-0.02	0.76	0.01

	68W		88M		91B		92F	
	MPCS	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*	*MPCS*	Δ*MPCS*
ASVAB	--	--	0.45	--	0.65	--	0.59	--
TAPAS [12]	--	--	1.15	0.71	1.11	0.46	1.23	0.64
RBI [14]	--	--	1.23	0.78	1.18	0.53	1.27	0.68
AKA [6]	--	--	0.63	0.18	0.88	0.23	0.60	0.01
WPA-F [14]	--	--	1.03	0.58	1.10	0.45	0.94	0.35
WPA-D [6]	--	--	0.89	0.44	0.72	0.07	0.81	0.22
AIM [6][a]	--	--	0.75	0.35	0.63	0.05	1.17	0.65
AO [1]	--	--	0.47	0.03	0.89	0.24	0.68	0.09

Note. MOS whose results are blank "--" had insufficient sample size on the targeted predictor-criterion measure pairing and were excluded from those analyses. Because of pairwise deletion, the sample size associated with the computation of the covariance matrices used as input for calculating Horst's *d* and MPCS varied by predictor-criterion measure pairing. The sample sizes used in the computation of MPCS by MOS were: 11B = 311, 19K = 95, 25U=53, 31B =212, 42A=60, 68W = 39, 88M = 61, 91B = 65, 92F = 44.
[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

The pattern of findings by predictor measure and MOS were generally the same between the expanded and target MOS samples. Overall, the pattern of findings between the expanded and target MOS samples were generally the same in terms of the relative rank ordering of the predictor measures and in cross-MOS differences, even though the average estimated classification gains differed in absolute terms. Consistent with previous results, the TAPAS, WPA, and RBI emerged as the non-cognitive predictor measures evidencing the greatest classification gains over the ASVAB. Similarly, the kinds of MOS most likely to benefit from these measures remained the same, with the non-close combat MOS continuing to demonstrate greater average gains in H_d and MPCS.

Conclusions

Tables 8.10 and 8.11 provide an overall summary of the experimental predictor measures' classification potential for maximizing the performance-related and retention-related criteria, respectively (based on the six target MOS). Overall, the results of the Army Class longitudinal classification-oriented analyses demonstrate the following:

- The experimental non-cognitive predictor measures show promise for enhancing the classification of new Soldiers to entry-level MOS. Consistent with previous research, the non-cognitive measures evidenced non-trivial classification gains, on average, for both the target and expanded MOS samples. The estimated gains over the existing ASVAB were comparable across performance and retention-related criteria.

- Among the non-cognitive predictor measures, the TAPAS, followed by the WPA and the RBI, consistently emerged as the "best bets" for enhancing new Soldier classification. However, the relative rank ordering of these measures varied by MOS suggesting that the measures substantively differ in their coverage of the predictor space. This finding has implications for how Army decision-makers weight the different predictor measures when determining which measures to administer and how two or more of the measures might be combined to make operational classification decisions.

Table 8.10. Summary of the Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted Training Outcomes (Target MOS)

Criterion Domain/ Predictor Measure	H_d				*MPCS*			
	Avg	*Avg* Δ	*Min* Δ	*Max* Δ	*Avg*	*Avg* Δ	*Min* Δ	*Max* Δ
Overall (Training, k = 4)								
ASVAB	0.04				0.17			
TAPAS [12]	0.14	0.10	0.05	0.21	0.34	0.17	0.10	0.28
RBI [14]	0.11	0.07	0.03	0.09	0.30	0.13	0.07	0.16
AKA [6]	0.05	0.02	0.01	0.02	0.22	0.04	0.02	0.06
WPA-F [14]	0.11	0.07	0.03	0.09	0.30	0.13	0.07	0.17
WPA-D [6]	0.08	0.04	0.01	0.05	0.24	0.07	0.04	0.09
AIM [6][a]	0.08	0.04	0.02	0.08	0.23	0.09	0.04	0.14
AO [1]	0.05	0.00	0.00	0.01	0.18	0.01	0.00	0.01
Performance-Related (Training, k = 3)								
ASVAB	0.04				0.18			
TAPAS [12]	0.11	0.06	0.05	0.07	0.30	0.13	0.10	0.15
RBI [14]	0.11	0.07	0.03	0.09	0.29	0.12	0.07	0.16
AKA [6]	0.05	0.01	0.01	0.02	0.21	0.04	0.02	0.06
WPA-F [14]	0.11	0.07	0.03	0.09	0.29	0.11	0.07	0.14
WPA-D [6]	0.08	0.04	0.01	0.05	0.23	0.06	0.04	0.08
AIM [6]a	0.07	0.03	0.02	0.04	0.22	0.07	0.04	0.09
AO [1]	0.05	0.00	0.00	0.01	0.18	0.01	0.00	0.01
Retention-Related (Training, k = 1)								
ASVAB	0.03				0.16			
TAPAS [12]	0.24	0.21	--	--	0.44	0.28	--	--
RBI [14]	0.09	0.06	--	--	0.30	0.15	--	--
AKA [6]	0.05	0.02	--	--	0.22	0.06	--	--
WPA-F [14]	0.10	0.07	--	--	0.32	0.17	--	--
WPA-D [6]	0.07	0.03	--	--	0.25	0.09	--	--
AIM [6]a	0.12	0.08	--	--	0.28	0.14	--	--
AO [1]	0.04	0.00	--	--	0.16	0.01	--	--

Note. k is the number of criterion variables considered in the average.
[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

Table 8.11. Summary of the Classification Potential of the Experimental Predictor Measures Relative to the ASVAB for Maximizing Predicted In-Unit Outcomes (Target MOS)

Criterion Domain/ Predictor Measure	H_d				*MPCS*			
	Avg	*Avg* Δ	*Min* Δ	*Max* Δ	*Avg*	*Avg* Δ	*Min* Δ	*Max* Δ
Overall (In-Unit, k = 5)								
ASVAB	0.09				0.31			
TAPAS [12]	0.35	0.26	0.21	0.31	0.62	0.31	0.26	0.37
RBI [14]	0.27	0.18	0.16	0.20	0.53	0.22	0.20	0.23
AKA [6]	0.14	0.05	0.04	0.07	0.38	0.07	0.05	0.10
WPA-F [14]	0.24	0.15	0.13	0.16	0.50	0.19	0.15	0.22
WPA-D [6]	0.17	0.08	0.07	0.10	0.42	0.11	0.07	0.14
AIM [6][a]	0.16	0.08	0.06	0.12	0.37	0.11	0.08	0.14
AO [1]	0.10	0.01	0.01	0.02	0.34	0.02	0.01	0.05
Performance-Related (In-Unit, k = 3)								
ASVAB	0.10				0.32			
TAPAS [12]	0.37	0.27	0.23	0.31	0.62	0.31	0.26	0.37
RBI [14]	0.29	0.19	0.18	0.20	0.55	0.23	0.22	0.23
AKA [6]	0.15	0.05	0.04	0.07	0.39	0.07	0.05	0.1
WPA-F [14]	0.25	0.15	0.13	0.16	0.49	0.17	0.15	0.2
WPA-D [6]	0.17	0.08	0.07	0.09	0.42	0.10	0.07	0.14
AIM [6]a	0.16	0.08	0.06	0.12	0.36	0.10	0.08	0.14
AO [1]	0.11	0.02	0.01	0.02	0.35	0.03	0.02	0.05
Retention-Related (In-Unit, k = 2)								
ASVAB	0.08				0.31			
TAPAS [12]	0.32	0.25	0.21	0.28	0.63	0.33	0.30	0.35
RBI [14]	0.24	0.17	0.16	0.17	0.51	0.21	0.20	0.21
AKA [6]	0.12	0.05	0.05	0.05	0.38	0.08	0.07	0.09
WPA-F [14]	0.22	0.15	0.14	0.16	0.52	0.21	0.20	0.22
WPA-D [6]	0.16	0.09	0.07	0.10	0.42	0.12	0.11	0.13
AIM [6]a	0.16	0.09	0.07	0.11	0.39	0.11	0.09	0.13
AO [1]	0.09	0.01	0.01	0.01	0.32	0.02	0.01	0.02

Note. k is the number of criterion variables considered in the average.
[a]AIM was analyzed separately because of insufficient data in the 19K MOS. Baseline and incremental calculations are comparable, but not identical to those performed on other predictors.

CHAPTER 9: SUMMARY AND CONCLUSIONS

Matthew T. Allen, Deirdre J. Knapp, (HumRRO), and Kimberly S. Owens (ARI)

The Army Class longitudinal validation research was designed to provide evidence about the usefulness of several measures that could be used to supplement the ASVAB for pre-enlistment screening and classification. This report briefly summarized the activities that took place in the first 3 years of the research program, including (a) development and administration of non-cognitive predictor measures to 11,000 new Soldiers and (b) administration of criterion measures at the end of Initial Military Training (IMT) to over 2,000 Soldiers in six target MOS (Knapp & Heffner, 2009a). The primary purpose of this report has been to describe the work that took place in the last three years of the Army Class research. In Year 4 (2009), data were collected from over 1,500 Soldiers in their first units of assignment, roughly 800 each from the Army-wide and target MOS samples. A second in-unit data collection was conducted in 2010 and early 2011. For the most part, the in-unit data were collected through proctored sessions for Regular Army Soldiers at multiple locations and primarily through unproctored sessions for USAR and ARNG Soldiers. These data were the basis for a set of analyses examining the potential for the experimental measures to (a) select higher-performing Soldiers, (b) improve Soldier retention, and (c) improve classification decisions.

Summary of Main Findings

Predicting In-Unit Soldier Performance

With respect to predicting in-unit Soldier performance, we found the following:

- *Multiple experimental measures predicted can-do in-unit criteria beyond the AFQT.* As expected, AFQT predicted can-do aspects of performance quite well. Even so, there was evidence of incremental validity particularly for the TAPAS, RBI, AIM, and WPA facets; however, coefficients for these measures generally failed to achieve statistical significance. The other three experimental measures (AO, AKA, and PSJT) were more likely to achieve statistical significance, but the magnitude of the effects was generally lower than for the previous four measures before correcting for shrinkage. In the in-unit 1 sample, after correcting for shrinkage, two measures exhibited the highest average incremental validity coefficients—AO and the PSJT. All incremental validity estimates save one (the PSJT prediction of MOS-specific job knowledge) dropped to near zero after shrinkage corrections in the in-unit 2 sample.

- *Multiple experimental measures predicted will-do in-unit criteria, after controlling for the AFQT, and more strongly than they predicted can-do criteria.* Three of the experimental measures had consistently higher incremental validity coefficients than the other measures, the RBI, TAPAS, and AIM. The WPA, AKA, AO, and PSJT also predicted will-do criteria over the AFQT but not as strongly as the three temperament measures. Among the experimental measures, the RBI showed the most promise in predicting will-do criteria after adjusting for shrinkage. On the other hand, the magnitude of the validity coefficients decreased substantially for the AIM, TAPAS, and WPA after adjusting for shrinkage. The pattern of results was similar across the

in-unit 1 and in-unit 2 samples, but the estimates were weaker and less likely to be statistically significant in the in-unit 2 sample.

- *Multiple experimental measures predicted deployment adjustment after controlling for AFQT, but did not predict deployment performance*. The RBI and AKA predicted deployment adjustment beyond AFQT in both the in-unit 1 and in-unit 2 samples. The AIM predicted deployment adjustment in the in-unit 1 sample, while the PSJT predicted deployment adjustment in the in-unit 2 sample. No measure, including the AFQT, predicted ratings of combat/deployment performance.

Predicting Attrition and Retention Intentions

With respect to predicting Soldier attrition and retention intentions, we found the following:

- *Multiple experimental measures predicted cumulative Soldier attrition beyond Education Tier.* The RBI and AIM emerged as the best predictors of overall cumulative attrition, followed by the TAPAS and the WPA. AO and AKA were also non-trivial predictors of attrition, though not at the same magnitude. In predicting the cumulative attrition for moral character, performance, and medical reasons, three experimental measures—AIM, TAPAS, and RBI—had the strongest rates of prediction. The WPA was also a strong predictor of medical attrition. A similar pattern of results emerged for predicting attrition over time, with a few exceptions. First, when adjusting for the number of parameters, AO also emerged as a very strong predictor of attrition over time, along with the AIM and RBI. The TAPAS also emerged as a strong predictor of moral character and performance attrition over time.

- *Multiple experimental measures showed incremental variance in Soldier retention and career intentions beyond Education Tier.* Education tier was generally ineffective for predicting retention and career intentions. The experimental measures, however, showed considerable promise. Affective commitment to the Army was predicted quite well by the RBI, WPA, and AKA in both in-unit samples. The career intentions scale was predicted by these measures as well as the AIM and TAPAS. Perceived Army fit was predicted by all experimental measures except AO. There were minor differences across the two in-unit samples, with the most notable being the lower magnitude of the estimates.

Evaluating Classification Potential

We evaluated the classification potential of the experimental predictors using training and in-unit 1 criterion data. We found the following:

- *In general, the experimental predictors exhibited non-trivial classification gains over the ASVAB for the six target MOS.* This held true for both MOS-specific performance-related criteria, such as job knowledge and ratings of technical performance, and MOS-specific retention-related criteria, such as self-reported MOS fit and MOS satisfaction. Across both sets of criteria, the TAPAS, RBI, and WPA exhibited the greatest classification gains over the ASVAB for the target MOS. That

being said, no single measure exhibited the greatest classification potential across the MOS (i.e., the best measure for an MOS varied by MOS).

- *The classification gains associated with the experimental predictor measures were somewhat higher, on average, for an expanded sample of MOS than the target MOS.* Although the cross-sample differences in classification gains were generally small, these findings illustrate the point that findings of classification potential can change depending on the specific MOS included in the analysis. They also suggest that the experimental predictor measures have classification potential beyond the six target MOS. Also, the pattern of findings by predictor measure was generally the same between the expanded and target MOS samples, with the TAPAS, RBI, and WPA showing the greatest classification gains over ASVAB.

Limitations and Issues

Comparing Results to Previous Army Class Findings

Overall, the results of the in-unit phase of the Army Class longitudinal validation were comparable to those found for the Army Class concurrent validation and the training criterion validation phase of the Army Class longitudinal validation. However, in general, the magnitude of the uncorrected incremental validity estimates were somewhat lower than in the training phase of the longitudinal validation (Knapp & Heffner, 2009) and in the concurrent validation (Ingerick et al., 2009), particularly when considering the in-unit 2 results. A number of factors may be contributing to these differences, such as:

1. *History effects*. The time difference between the training and in-unit longitudinal validation phases likely reduce the magnitude of the effects. For example, there is likely less variance in Soldier performance in their units due to additional training or turnover.

2. *Sample size*. The sample sizes were generally smaller for the in-unit longitudinal validation than for the two previous efforts, particularly for MOS-specific criteria. This decreases power and increases the probability of a Type II error, which makes it less likely to detect statistical significance. Also, given the smaller sample sizes and the variability in the number of component scales contributing to each experimental measure, sample-specific error could artificially inflate some of the estimates. We attempted to account for this inflation using statistical adjustments.

3. *Characteristics of the sample*. Criterion data were collected from Army-wide and MOS-specific samples. In the concurrent validation and the training phase of this longitudinal validation, criterion data were only collected from targeted MOS. As described in Chapter 8, the experimental measures did not predict all criteria equally well for all MOS. Thus, variations in the measures' estimated validities across MOS may have obscured their overall validities.

4. *Unreliability in the supervisory ratings*. Many of the criteria used in this phase of the data collection were based on supervisor ratings. In the training validation, the single-rater reliability (i.e., ICC[C,1]) for many of the rating scales was below .20. In that

phase, there were more raters for each Soldier, which increased the *k*-rater reliability and thus the stability of the estimate. By contrast, in the in-unit data collection, each Soldier was typically rated by only one supervisor. If we assume that the single-rater reliability for this phase (which we cannot estimate) was similar to the single-rater reliability coefficients found in the training validation study, then the effect would certainly have attenuated validity coefficients (see, for example, Guion, 1998, pp. 313-314).

5. *Maturation and distributed administration effects.* With regard to the concurrent versus longitudinal validation in-unit criterion results, there may have been subtle maturation of the Soldiers that was reflected in their responses to the predictor measures which served to increase the observed correlations with the criterion measures in the concurrent design. One would also expect greater common variance among measures administered at the same, as opposed to different, points in time.

For these and other reasons, the results for this phase and other phases of the Army Class research are not directly comparable.

Generalizability of Findings to an Operational Setting

At reception battalions, we were able to collect data from Soldiers at a point in their Army career that was as close to an operational applicant setting as possible. Although the current research is informative, there are substantive differences between the two settings that could limit the generalizability of these findings to an actual applicant context. Chief among these is that respondents in an operational applicant setting are likely to have a greater motivation to fake or otherwise misrepresent themselves on the experimental predictor measures than in the current research. This suggests that the nature of the findings in this report could change when the measures are administered in an operational context, so further work to explore this issue is needed.

Future Research

As described at the beginning of this report, an initial operational test and evaluation (IOT&E) is underway. Based on earlier empirical validation results using training criterion data, as well as consideration of other factors (e.g., anticipated resistance to response distortion in an operational environment and coverage of multiple non-cognitive domains), the TAPAS, and soon the WPA, are being administered to Army applicants as part of this IOT&E. Paralleling the Army Class research design, the IOT&E includes collection and analysis of empirical training and in-unit performance data to evaluate how well the TAPAS and WPA function in an operational environment. As those results continue to become available (Knapp & Heffner, 2011), they will be compared with the Army Class research findings.

REFERENCES

Allen, M.T., Cheng, Y.A., Ingerick, M.J., & Caramagno, J.P. (2009). Measure scoring and psychometric properties. In D.J. Knapp & T.S. Heffner (Eds.) *Validating future force performance measures (Army Class): End of training longitudinal validation* (pp. 24-30) (Technical Report 1257). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Buddin, R.J. (2005). *Success of first-term Soldiers: The effects of recruiting practices and recruit characteristics.* Santa Monica, CA: Rand.

Burket, G.R. (1964). A study of reduced rank models for multiple prediction. *Psychometricka Monograph Supplement*, No. 21.

Campbell, J.P., Hanson, M. A., & Oppler S. H. (2001). Modeling performance in a population of jobs. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification.* Hillsdale, NJ: Erlbaum.

Campbell, J.P., & Knapp, D.J. (Eds.) (2001). *Exploring the limits in personnel selection and classification.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Campbell, J.P., McCloy, R.A., McPhail, S.M., Pearlman, K., Peterson, N.G., Rounds, J., & Ingerick, M. (2007). *U.S. Army Classification Research Panel: Conclusions and recommendations on classification research strategies* (Study Report 2007-05). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43,* 313-333.

Collins, M., Le, H., & Schantz, L. (2005). Job knowledge criterion tests. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 49-58). (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

DeCorte, W. (2000). Estimating the classification efficiency of a test battery. *Educational and Psychological Measurement, 60,* 73-85.

Dover, S.J. (2002). *The characterization and prediction of Soldier performance during routine service and in combat* (Research Note 2002-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Drasgow, F. , Stark, S., & Chernyshenko, O.S. (November, 2006). *Toward the next generation of personality assessment systems to support personnel selection and classification decisions*. Paper presented at the 48th annual conference of the International Military Testing Association, Canada.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-358.

General Accounting Office (1998). *Military attrition: Better data, coupled with policy changes could help the services reduce early separations*. Washington, D.C: General Accounting Office.

General Accounting Office (2000). *Military personnel. Services needed to assess efforts to meet recruiting goals and cut attrition.* (Report No. GAO/NSIAD-00-146). Washington, DC: General Accounting Office.

Guion, R.M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Holland, J.L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources, Inc.

Hooper, A., Paullin, C., Putka, D.J., & Strickland, W.J. (2008). *An empirical analysis of reasons for attrition among first term airmen in the USAF* (FR-08-32). Alexandria, VA: Human Resources Research Organization.

Horst, P. (1954). A technique for the development of a differential predictor battery. *Psychological Monographs: General and Applied, 68,* 1-31.

Horst, P. (1955). A technique for the development of an absolute prediction battery. *Psychological Monographs: General and Applied*, *69,* 1-22.

Ingerick, M., Diaz, T., & Putka, D. (2009). *Investigations into Army enlisted classification systems: Concurrent validation report* (Technical Report 1244). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Joreskog, K. & Sorbom, D. (1996). *LISREL 8: User's Reference Guide*. Lincolnwood, IL: Scientific Software International.

Keene, S.D., & Halpin, S.M. (1993). *How well did the combat training centers prepare units for combat? Questionnaire results from Desert Storm participants* (Technical Report 970). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kilcullen, R.N., Mael, F.A., Goodwin, G.F., & Zazanis, M.M. (1999). Predicting U.S. Army Special Forces field performance. *Human Performance in Extreme Environments, 4(1),* 53-63.

Kilcullen, R.N., Putka, D.J., McCloy, R.A., & Van Iddekinge, C.H. (2005). Development of the Rational Biodata Inventory. In D.J. Knapp, C.E. Sager, & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 105-116) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kilcullen, R.N., White, L.A., Sanders, M. & Hazlett, G. (2003). *The Assessment of Right Conduct (ARC) administrator's manual* (Research Note 2003-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapik, J.J., Jones, B.H., Hauret, K., Darakjy, S., & Piskator, E. (2004). *A review of the literature on attrition from the military services: Risk factors for attrition and strategies to reduce attrition* (Technical Report 12-HF-01Q9A-04). Aberdeen Proving Ground, MD: US Army Center for Health Promotion and Preventive Medicine.

Knapp, D.J., & Campbell, R.C. (Eds). (2006). *Army enlisted personnel competency assessment program: Phase II report* (Technical Report 1174). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Heffner, T.S. (Eds.) (2009). *Predicting Future Force Performance (Army Class): End of Training Longitudinal Validation* (Technical Report 1257). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Heffner, T. S. (Eds.). (2010). *Expanded Enlistment Eligibility Metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Heffner, T.S. (Eds.) (2011). *Tier One Performance Screen Initial operational test and evaluation: 2010 annual report* (Technical Report 1296). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., McCloy, R.A., & Heffner, T.S. (Eds.) (2004). *Validation of measures designed to maximize 21st-century Army NCO performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Sager, C. E., & Tremble, T. R. (Eds.) (2005). *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Tremble, T.R. (Eds) (2007). *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh*, *62*, 28–30.

Lytell, M.C., & Drasgow, F. (2009). "Timely" methods: Examining turnover rates in the U.S. Military. *Military Psychology, 21,* 334-350.

Moriarty, K.O., Campbell, R.C., Heffner, T.S., & Knapp, D.J. (2009). *Validating future force performance measures (Army Class): Reclassification test and criterion development* (Research Product 2009-11). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Nord, R., & Schmitz, E. (1991). Estimating performance ultility effects of alternative selection and classification policies. In J.Zeidner & C.D. Johnson (Eds.). *The economic benefit of perdicting job performance: Vol5. Estimating the gains of alternative policies* (pp.73-131), New York: Praeger.

Peterson, N.G., Russell, T.L., Hallam, G., Hough, L.M., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1992). 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* (Technical Report 952) (pp. 73-199). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Putka, D.J, Noble, C.L., Becker, D.E., & Ramsberger, P.F. (2004). *Evaluating moral character waiver policy against Servicemember attrition and in-service deviance through the first 18 months of service* (FR-03-96). Alexandria, VA: Human Resources Research Organization.

Putka, D.J., & Van Iddekinge, C.H. (2007). Work Preferences Survey. In D.J. Knapp & T.R. Tremble (Eds.), *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rosse, R. L., Campbell, J. P., & Peterson, N. G. (2001). Personnel classification and differential job assignments: Estimating classification gains. In J. P Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 453-506). Mahwah, NJ: Lawrence Erlbaum Associates.

Russell, T.L., Peterson, N.G., Rosse, R.L., Hatten, J.L.T., McHenry, J. J., & Houston, J.S. (2001). The measurement of cognitive, perceptual and psychomotor abilities. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 71-110). Mahwah, NJ: Lawrence Erlbaum Inc.

Russell, T.L., Reynolds, D.H., & Campbell, J.P. (Eds.) (1994). *Building a joint service classification research roadmap: Individual differences measurement* (AL/HR-TP-1994-0009). Brooks AFB TX: Armstrong Laboratory.

Schmitt, N., & Ployhart, R. E. (1999). Estimates of cross-validity for stepwise regression and with predictor selection. *Journal of Applied Psychology, 84*(1), 50-57.

Scholarios, D., Johnson, C. D., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology, 79,* 412-424.

Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.

Smith, P.C., & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47,* I149-155.

Stark, S. (2002). *A new IRT approach to test construction and scoring designed to reduce the effects of faking in personality assessment* [Doctoral Dissertation]. University of Illinois at Urbana-Champaign.

Stark, S., Drasgow, F., & Chernyshenko, O.S. (October, 2008). *Update on Tailored Adaptive Personality Assessment System (TAPAS): The next generation of personality assessment systems to support personnel selection and classification decisions*. Paper presented at the 50th annual conference of the International Military Testing Association, Amsterdam, Netherlands.

Strickland, W.J. (Ed.) (2005). *A longitudinal examination of first term attrition and reenlistment among FY1999 enlisted accessions* (Technical Report 1172). Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Van Iddekinge, C.H., Putka, D.J., & Sager, C.E. (2005). Attitudinal criteria. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 89-104) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Waugh, G.W., & Russell, T.L. (2005). Predictor situational judgment test. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 135-154) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

White, L.A., Hunter, A.W., & Young, M.C. (October, 2008). *Update on the Army's Tier Two Attrition Screen (TTAS).* Paper presented at the annual meeting of the International Military Testing Association, Amsterdam, The Netherlands.

White, L.A., Jose, I.J., & LaPort, K.A. (2011) Army Research Institute analyses for the Eighth TTAS Evaluation Report. In D. Bohn (Ed). Tier Two Attrition Screen Interim Report, January 2011. U.S. Army Accessions Command, Fort Monroe, VA.

White, L.A., Rumsey, M.G., Matyuf, M.M., & Borman, W.C. (1994). *Relations between Soldiers' performance during peacetime and in combat.* Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

White, L.A., & Young, M.C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM).* Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

White, L.A., Young, M.C., Heggestad, E.D., Stark, S., Drasgow, F., & Piskator, G. (2004). Development of a non-high school diploma graduate pre-enlistment screening model to enhance the future force. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

White, L.A., Young, M.C., & Rumsey, M.G. (2001). ABLE implementation issues and related research. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 525-558). Mahwah, New Jersey: Lawrence Erlbaum Associates.

APPENDIX A
DEVELOPMENT OF THE COMBAT/DEPLOYMENT PERFORMANCE RATING SCALES

Laurie Wasko (HumRRO), Kimberly S. Owens (ARI), Roy Campbell, and Teresa Russell (HumRRO)

Although some rating scales developed for prior phases of Army Class reference performance dimensions were relevant for deployment (e.g., Warrior Tasks and Battle Drills [WTBD] Knowledge and Skill), those scales were developed primarily without a deployment focus using samples of in-garrison Soldiers. To assess aspects of Soldiers' performance unique to their time in deployment, we developed the Combat/Deployment Performance Rating Scales (CDPRS).

CDPRS development involved three main stages. First, we reviewed literature on combat performance and previous rating scale development activities to identify preliminary dimensions for the scales. Second, we asked subject matter experts (SMEs) to generate critical incidents. We then used these critical incidents to revise the dimensions and develop behaviorally anchored rating scales (BARS) for each dimension. Last, we revised the draft BARS based on several rounds of SME feedback.

Identification of Preliminary Combat/Deployment Dimensions

To identify potential dimensions for inclusion in the CDPRS, the HumRRO/ARI project team began by reviewing existing combat rating scales. We based a preliminary list of dimensions on an initial content analysis of this available information. We then reviewed and edited the definitions of each dimension.

Review and Analysis of Existing Scales and Information

In order to identify dimensions for potential inclusion in the CPDRS, we first reviewed the all of the source documents we could find, including the Soldier's Combat Evaluation from Dover (2002), the Combat Performance Questionnaire (Operation Desert Shield/Storm) used by White, Rumsey, Matyuf, and Borman (1994), Combat Performance Prediction Scales (Campbell & Knapp, 2001), and survey results regarding preparedness for combat from Keene and Halpin (1993).

Then, using the in-unit Army-wide (AW) rating scale dimensions as a taxonomic structure, we reviewed the individual scales or items in each of the source documents and attempted to categorize each scale/item into the existing AW scales. Scales/items that did not fit in the AW dimensions were listed separately. Based on the review, we created the three lists shown in Table A.1 which identify the AW dimensions that were the most and least common across all of the source documents' rating scales, as well as new dimensions from the source documents that were not already captured with the AW performance dimensions. The results of the sorting exercise helped to identify (a) the dimensions that were most salient in the combat scales, (b) scales that are unique compared to the AW scales, and (c) AW dimensions that could be redefined to have a greater deployment orientation.

Table A.1. Results of the AW and Source Document Dimension Sort

Most Common AW Performance Dimensions	Least Common AW Performance Dimensions
• Contributing to the Team	• Managing Personal Matters
• Warrior Tasks and Battle Drills Knowledge and Skill	• Processing Information
• Effort	• Developing Own Skills
• Solving Problems	• Performing MOS-Specific Tasks
• Exhibiting Personal Discipline	• Following Safety Procedures
• Physical Fitness and Bearing	• Interacting with Indigenous People and Soldiers from Other Countries
	• Communicating with Others
New Dimensions	
• Emotional Stamina	
• Bravery and Courage	
• Vigilance	

In the early stages of CDPRS development, we included relevant AW PRS dimensions to ensure that no unique facets for combat or deployment were overlooked. From this larger list of dimensions, we chose 11 for further consideration for the new deployment performance scales: six AW PRS dimensions that were common across the reviewed source documents (*Contributing to the Team, Warrior Tasks and Battle Drills Knowledge and Skill, Effort, Solving Problems, Exhibiting Personal Discipline, Physical Fitness and Bearing*), two AW PRS dimensions that were not commonly identified (*Interacting with Indigenous People and Soldiers from Other Countries, Managing Personal Matters*), and the three new dimensions (*Emotional Stamina, Bravery and Courage, Vigilance*). The two AW PRS dimensions that were not frequently identified in the existing combat scales were chosen based on the relevance of their content to modern day warfare. Specifically, *Interacting with Indigenous People* is an area that has become more common for Soldiers on deployment than it was in earlier conflicts, and *Managing Personal Matters* is a deployment-oriented category which has to do with managing one's home life while abroad.

Preliminary Dimension Definitions

We edited the dimension definitions to be more combat/deployment-oriented, using content from the existing combat scales and items to inform the new dimension definitions. Based on distinctions between physical and moral courage in the Army values literature, we split *Bravery and Courage* into two separate dimensions: *Physical Courage* and *Moral Courage.*

Critical Incident Review and Analysis

Once the dimensions and their definitions were determined, the next step in the development of the CDPRS was to gather and analyze critical incidents for each dimension.

Critical Incident Workshops

Workshop Overview

We conducted four workshops with NCOs—two at Fort Benning and two at Fort Sill. A total of 30 NCOs participated. Participants were students in either the Advanced Leader Course or Senior Leader Course, with an average of approximately 11 years of experience. All NCOs except for one had served in at least one deployment. Materials were modified in two stages; we revised the dimensions and their definitions based on the results of the workshops held at Fort Benning and then used these revised materials to gather additional data at Fort Sill.

In the dimension review portion of the workshop, we asked NCOs to rate the criticality of the 12 performance dimensions for performance of an entry-level in-unit Soldier with up to 18 months of deployment experience using a 1 to 5 relative criticality scale (1 = *much less critical than other dimensions* to 5 = *much more critical than other dimensions*). The purpose of collecting ratings was to provide a framework for a discussion of the dimensions. After the NCOs made their ratings, we computed the means and standard deviations of the ratings and read them back to the SMEs. Then, we facilitated a discussion about each dimension by asking the NCOs to describe high and low performance for each dimension and to discuss the dimensions in general.

In the second part of the workshop, we collected written critical incidents from the SMEs (Flanagan, 1954). The focus of each critical incident statement is on an individual's behavior in a specific situation, and includes a description of the situation, the actual actions of the individual in response to that situation, and the result of those actions. As NCOs wrote their critical incidents, we circulated through the room, reading incidents and asking for clarification as needed. Periodically, we asked SMEs to read their incidents to the group.

Workshop Results

Across the two groups of NCOs at Fort Benning, the dimensions *Executing Warrior Skills in an Operational Environment, Solving Problems in the Field, Exhibiting Personal Discipline,* and *Vigilance* were rated as the most critical. Participants wrote 43 critical incidents. Based on the discussion with SMEs at Fort Benning, we made a number of changes to the dimensions and their definitions. We dropped *Managing Personal Matters* because its focus was on pre-deployment issues.

The criticality ratings from Fort Sill largely confirmed the criticality ratings obtained at Fort Benning. That is, they indicated that NCOs perceived *Warrior Skills in an Operational Environment, Field/Combat Judgment, Field Readiness, Physical Fitness and Endurance,* and *Vigilance* to be the most critical dimensions. At Fort Sill, the participants wrote 59 more critical incidents. NCO feedback prompted a few minor changes to the titles and definitions of two dimensions.

Further Analysis of Dimensions

To evaluate the dimension structure, three team members conducted a retranslation exercise; specifically, they independently sorted the critical incidents into the 11 remaining dimensions. Table A.2 shows the number of incidents placed into the same category by all three (100%) or two of the three (66%) project staff. We also mapped the commonality between the AW in-unit PRS and the combat/deployment dimensions. The results appear in Table A.3.

Table A.2. Summary of In-House Critical Incident Retranslation Results

Dimension	100%	66%
A. Executing Warrior Skills in an Operational Environment	2	5
B. Field/Combat Judgment	5	7
C. Field Readiness	18	6
D. Contributing to the Team ·	0	3
E. Cultural Awareness	0	0
F. Effort and Initiative	0	4
G. Physical Fitness and Endurance	3	1
H. Emotional Resilience	2	1
I. Physical Courage	1	5
J. Integrity	6	4
K. Awareness and Vigilance	5	8

Table A.3. Mapping of AW and Combat /Deployment Rating Scales

Combat /Deployment Dimension	AW In-Unit Rating Dimension
Executing Warrior Skills in an Operational Environment	Performing Core Warrior Tasks
Field/Combat Judgment	
Field Readiness	
Contributing to the Team	Contributing to the Team
Cultural Awareness	Interacting with Indigenous People and Soldiers from other Countries
Effort and Initiative	Exhibiting Effort
Physical Fitness and Endurance	Exhibiting Fitness and Bearing
Emotional Resilience	
Physical Courage	
Integrity	Exhibiting Personal Discipline
Awareness and Vigilance	

Based on results of the retranslation exercise (Table A.2) and AW in-unit scale mapping (Table A.3), we made the following revisions to the rating dimensions:

- Dropped *Contributing to the Team, Cultural Awareness,* and *Effort and Initiative* because they (a) were not strongly supported by critical incidents and (b) overlapped with the AW scales.
- Combined *Executing Warrior Skills in an Operational Environment* with *Field/Combat Judgment*. Incidents for these dimensions were difficult to distinguish (as identified during the retranslation task). Also, *Executing Warrior Skills in an Operational Environment* overlapped with the AW rating scales.
- Dropped *Integrity* because it was covered by the AW rating scales and the incidents written for it were ones that could easily have been written in a garrison situation. Incidents involved lying to the NCO, using drugs, and other behaviors that were not unique to a deployment environment.
- Retained *Field/Combat Judgment, Field Readiness, Physical Fitness and Endurance, Emotional Resilience, Physical Courage,* and *Awareness and Vigilance.* All of these dimensions (with the exception of *Physical Fitness and Endurance*) were distinct from the AW scales. *Physical Fitness and Endurance* did overlap with the AW scales; however, it appeared to focus more on sustained physical performance than fitness. We renamed the dimension *Physical Endurance.*

Six dimensions emerged for inclusion in the CDPRS. These dimensions and their definitions appear in Table A.4.

Table A.4. Combat/Deployment Dimension Definitions Based on Critical Incident Workshops

A. Field/Combat Judgment
Thinks rationally under pressure. Makes sound on-the-spot decisions in the field based on prior training. Applies correct rules (e.g., rules of engagement [ROE], escalation of force) to the situation. Immediately and correctly performs required warrior tasks and drills.
B. Field Readiness
Keeps self, weapons, and equipment in combat-ready condition. Maintains positive control and accountability of weapons, equipment, tools, and munitions. Follows procedures for handling equipment and weapons safely.
C. Physical Endurance
Is capable of meeting the demands of physical or environmental challenges or stressful situations. Sustains performance as long as the situation requires.
D. Emotional Resilience
Deals effectively with the cumulative effects of stress from work and home. Reacts to the signs of combat and operational stress. Takes positive steps in managing stress reactions.
E. Physical Courage
Overcomes fears of bodily harm. Takes necessary risks in spite of fears. Does not act recklessly or place self or others at unwarranted risk.
F. Awareness and Vigilance
Maintains sense of awareness and alertness to enemy and environment threats. Acts as constant sensor to unusual or threatening persons or conditions. Remains focused and alert despite sleep deprivation, extended missions, and difficult environmental conditions.

Behaviorally Anchored Rating Scale (BARS) Development

After analyzing the results of the critical incident workshops, choosing the appropriate dimensions for the CDPRS, and fine-tuning the dimension definitions, we drafted anchors for the six dimensions listed in Table A.4.

Development of Initial Draft Scales

Initial development of the BARS took place in three stages: content analysis, draft development, and review. First, the project team conducted a content analysis on the critical incidents gathered from Forts Benning and Sill. As part of this process, we generated behavioral summary statements from each of the more detailed critical incidents. We then reviewed the behavioral statements, and identified themes across the statements and used the behavioral summary information to create low, moderate, and high effectiveness anchors for each theme. The result of this effort was a draft comprised of six behaviorally anchored rating scales, with one scale for each dimension.

Internal project team members then reviewed the scales for content and clarity. As a result of this review, we made a few additional edits and dropped the scale associated with the dimension *Emotional Resilience* due to its potentially pejorative tone and sensitive nature.

SME Workshops

Retranslation exercises with NCOs at Fort Gordon evaluated the extent to which each rating scale anchor was representative of both the dimension and level of effectiveness it was written to embody.

Workshop Overview

We held two workshops with a total of 16 NCOs at Fort Gordon. The participants were E-5 and E-6 NCOs with an average of 8.63 years of experience in the Army. All NCOs had served in at least one deployment, with a little more than half having served two or more. We repeated the same exercise in both workshops, using the same raw materials.

Each workshop had two main activities—a retranslation exercise (Smith & Kendall, 1963) and a discussion of results. For the retranslation exercise, the 39 anchors were presented in a random order. NCOs made two judgments about each anchor. One judgment was an effectiveness rating placed on a scale from 1 to 7, where 1 = Low Effectiveness and 7 = High Effectiveness. The second judgment was to categorize each of the anchors into one of the five dimensions (see Table A.4; recall *Emotional Resilience* was eliminated). This exercise determined whether an anchor accurately embodied the performance dimension it was intended to represent.

After the NCOs made their ratings, we calculated the percentage of raters who had categorized the anchor into the intended dimension. We also calculated the mean effectiveness rating of the anchor, categorizing it as either a low (<3), medium ($3 – 4.99$), or high (≥ 5) level of effectiveness.

The second part of the workshop was a discussion of the retranslation exercise results. We showed the SMEs a draft of the rating scales (i.e., a draft version of what the BARS would look like operationally) and reviewed the ratings for each of the anchors. We explored anchors that were categorized below a 62.5% level of agreement (i.e., those where fewer than five out of eight raters categorized the anchor as intended), and those that were thought to be written at a level of effectiveness that was higher or lower than intended. If the anchor appeared to be too low or too high for its position on the scale, we asked the SMEs for input on how the anchor could be edited to better reflect the intended level of performance. Lastly, a final sweep through the BARS was made to identify vocabulary that was either at too high a reading level, or that could be reworded into terms that were more familiar to Army Soldiers and NCOs.

Workshop Results

Overall, there was a substantial amount of agreement in both the dimension categorization and the effectiveness ratings, in both the morning and afternoon sessions. As depicted in Table A.5, there was 100% agreement for *Physical Endurance* across both sessions. *Physical Courage* and *Awareness and Vigilance* had the most disagreement in the categorization of anchors.

Table A.5. Percentage of Anchors Categorized as Intended

	Morning Session			Afternoon Session		
	≥ 62.5%	62.5-50.0%	<50%	≥ 62.5%	62.5-50.0%	<50%
A. Field/Combat Judgment	100.0	0.0	0.0	77.8	11.1	11.1
B. Field Readiness	100.0	0.0	0.0	88.9	11.1	0.0
C. Physical Endurance	100.0	0.0	0.0	100.0	0.0	0.0
D. Physical Courage	66.7	0.0	0.3	50.0	33.3	16.7
E. Awareness and Vigilance	77.8	11.1	11.1	66.7	11.1	22.2

With regard to the effectiveness ratings, a majority of the ratings reflected the level of effectiveness they were written to address. We handled the exceptions by rewriting or editing anchors in accordance with the discussion. As a whole, the types of changes we made to the scales based on NCO feedback were small but impactful. The biggest change was that we dropped the second theme in *Physical Courage* (which was intended to be written around split-second heroics) from the scale. We implemented a majority of the edits to further reduce the reading level of the anchors (e.g., we replaced words such as “cognizant,” “circumvent,” and “oblivious”), or to use terminology that would be more familiar to Soldiers and NCOs (e.g., replaced the phrase “stowed weapons” with “secured weapons”).

Army Test Program Advisory Team Meeting

Army Test Program Advisory Team (ATPAT) members, a group of senior NCOs familiar with the Army Class research program, reviewed the revised CDPRS draft with a careful eye for content (e.g., appropriateness of content, reading level) and clarity. After all members

had sufficient time to review, we facilitated a discussion about each of the dimensions and their associated anchors.

The most substantial revision stemming from the ATPAT meeting was with regard to *Physical Endurance*. There were three major comments about this particular dimension: (a) the medium and high anchors of the first theme were not sufficiently contextually rich, (b) the second theme was redundant with the AW *Exhibiting Fitness and Bearing* scale and should be dropped, and (c) anchors for an additional theme tapping into the mental aspect of physical endurance should be written.

CDPRS Review

The last stage in the development of the CDPRS was an additional SME review and a pilot test of the scales. First, we conducted a review with 10 NCOs enrolled in the Senior Leader Course at Fort Leonard Wood. By and large, the NCOs felt that the rating scales were descriptive of Soldier's deployment performances. A few minor wording edits were made to some anchors.

Lastly, we piloted the CDPRS with eight NCOs at Fort Knox. Here, we asked NCOs to provide input on the rating scales by trying out the scales and then discussing them. NCOs were asked to think of three Soldiers and rate those Soldiers' performance using the rating scales. NCOs were also asked to provide feedback on the rating scales, specifically to include feedback on the content of the scales and comments on the potential for variability in ratings when using the scales. Overall, NCOs felt that the scales were well-defined, easy to understand, and easy to use.

The final content of the CDPRS is shown on the following pages. This was programmed into ARI's web-based survey platform, InterForm, for computer-based administration during the second in-unit criterion validation phase of the Army Class project.

COMBAT/DEPLOYMENT PERFORMANCE RATING SCALES

A. Field / Combat Judgment: Thinks rationally under pressure. Makes sound on-the-spot decisions in the field. Applies correct rules (e.g., ROE, escalation of force) to the situation. Immediately and correctly performs required warrior tasks and drills.

1 2	3 4 5	6 7
– Freezes in pressure situations, failing to accomplish even the more basic warrior tasks and drills.	– Responds quickly and effectively in situations that are similar to those encountered in training or during prior combat experience; sometimes hesitates or requires prompting when faced with unfamiliar or difficult tasks.	– Always responds quickly and effectively to threat situations.
– Makes bad decisions or has to rely on the directions of others.	– Usually makes acceptable and effective decisions under pressure.	– Uses sound judgment to positively impact a negative or potentially dangerous situation; quickly improvises in new and challenging situations.
– Does not follow ROE and escalation of force procedures; actions may result in unnecessary casualties or risks to non-combatants.	– Knows and applies ROE and escalation of force procedures in most situations; requires additional guidance and reinforcement in some situations.	– Consistently, rapidly, and correctly applies ROE and escalation of force procedures; actions stop potentially catastrophic event(s) from occurring.

B. Field Readiness: Keeps self, weapons, and equipment in combat ready condition. Maintains positive control and accountability of weapons, equipment, tools, and munitions. Follows procedures for handling equipment and weapons safely.		
1 2	3 4 5	6 7
– Fails to perform function checks on weapons, munitions, and equipment prior to and during missions; fails to follow instructions or SOP for mission prep; does not use, or incorrectly uses, Personal Protective Equipment (PPE).	– Performs function checks on weapons, munitions, and equipment prior to and during mission; sometimes needs reminders on checking/maintaining additional equipment and in proper use of PPE.	– Is proactive in performing function checks on weapons, munitions, and equipment prior to mission; maintains weapons and equipment in highest state of readiness; properly uses PPE and reinforces PPE use in others.
– Does not follow correct procedures in unloading and clearing weapons, muzzle orientation, or use of clearance barrel; does not follow safety procedures when mounting / dismounting weapons systems; has had an accidental weapons discharge.	– Follows safety procedures when handling weapons; needs occasional reminders or reinforcement.	– Follows correct safety procedures in handling all weapons; is alert to and enforces safety procedures in others.
– Fails to safeguard or account for weapons, munitions, tools or equipment; has lost or misplaced a weapon; fails to properly secure weapons, munitions, tools, or equipment, resulting in loss or damage.	– Is careful about safeguarding and accounting for weapons, munitions, tools, and equipment; needs some reminders and supervision on safeguarding and securing weapons, munitions, tools, and equipment.	– Is accountable for weapons, munitions, tools, and equipment at all times; always properly secures weapons, munitions, tools, and equipment.

C. Physical Endurance: Is capable of meeting the demands of physical or environmental challenges or stressful situations. Sustains performance as long as the situation requires.		
1 2	3 4 5	6 7
– Due to lack of physical endurance, causes other team members to have to compensate by taking over responsibilities when the Soldier is no longer able to perform; often lacks physical ability and endurance to complete the mission.	– Usually meets the demands of physical and environmental challenges that require exertion over extended periods of time.	– Exceeds expectations of physical endurance; is able to compensate for others that are less physically able (e.g., by carrying another Soldier's load, or carrying another Soldier when that individual is no longer able to walk).
– Is not able to mentally push through levels of physical or mental discomfort to meet the demands of a mission; quits during challenging situations.	– Usually sets aside thoughts of physical or mental discomfort; is able to push through mental / physical obstacles most of the time.	– Displays mental conviction; is able to persevere through physical challenges and stressful situations when others are not able.

D. Physical Courage: Overcomes fears of bodily harm. Takes necessary risks in spite of fears. Does not act recklessly or place self or others at unwarranted risk.		
1 2	3 4 5	6 7
– Avoids direct physical threat, fire, or exposure (e.g., by hiding); lets fear threaten mission or expose other Soldiers; fails to perform team tasks or follow directions because of fear.	– Follows leader directions in threat situations; is able to overcome fear in threat or exposure situations.	– Performs critical functions in threat situations without additional directions; takes calculated risks in threat or exposure situations, putting safety of others before threat to self.

E. Awareness and Vigilance: Maintains sense of alertness to enemy and environment threats. Is always aware of unusual or threatening persons or conditions. Remains focused and alert despite sleep deprivation, extended missions, and difficult environmental conditions.		
1 2	3 4 5	6 7
– Unaware of surroundings in situations where alertness is essential; lack of awareness results in increased risk or casualty.	– Maintains acceptable level of awareness of potential threats and surroundings; is able to contribute to group awareness.	– Is highly aware of surroundings; is able to identify threats and avoid potentially hazardous situations.
– Displays a lack of awareness of enemy; is unable to distinguish threats and non-threats; does not improve with experience.	– Is aware of threat and able to distinguish enemy personnel and activities; improves with experience.	– Quickly identifies enemy, suspicious personnel, and activities; displays keen sense of awareness of out-of-place persons or behaviors.
– Falls asleep during times of required vigilance (i.e., guard duty, OPs).	– Functions well in normal vigilance situations; requires reinforcement or back-up in extended or more extreme conditions.	– Stays alert and awake during periods of little sleep or the most difficult conditions.

APPENDIX B
DESCRIPTIVE STATISTICS AND SCORE INTERCORRELATIONS FOR SELECTED PREDICTOR MEASURES

Table B.1. Descriptive Statistics for Education Tier, Armed Services Vocational Aptitude Battery (ASVAB) Subtests, and Armed Forces Qualification Test (AFQT)

Scale	*M*	*SD*
Education Tier	1.25	0.43
ASVAB Subtests		
General Science (GS)	51.34	7.36
Arithmetic Reasoning (AR)	51.82	6.29
Word Knowledge (WK)	49.94	5.97
Paragraph Comprehension (PC)	51.47	5.09
Math Knowledge (MK)	52.17	6.30
Electronics Information (EI)	52.04	7.79
Auto and Shop Information (AS)	50.76	8.56
Mechanical Comprehension (MC)	53.18	7.62
Assembling Objects (AO)	54.88	7.95
AFQT	56.13	19.31

Note. n = 9,467-10,785. Subtests are reported as Sum of Standardized Subtest Scores (SSSS), AFQT is reported as a percentile.

Table B.2. Intercorrelations among Education Tier, ASVAB Subtest, and AFQT Scores

Scale		1	2	3	4	5	6	7	8	9	10
1	Education Tier										
2	General Science (GS)	-.03									
3	Arithmetic Reasoning (AR)	-.05	.39								
4	Word Knowledge (WK)	.04	.61	.25							
5	Paragraph Comprehension (PC)	.03	.43	.28	.43						
6	Math Knowledge (MK)	-.26	.28	.56	.09	.15					
7	Electronics Information (EI)	.03	.57	.36	.43	.32	.16				
8	Auto and Shop Information (AS)	.11	.42	.25	.29	.20	-.03	.58			
9	Mechanical Comprehension (MC)	.03	.52	.45	.36	.30	.24	.58	.57		
10	Assembling Objects (AO)	-.05	.30	.39	.16	.19	.32	.31	.23	.49	
11	AFQT	-.10	.66	.76	.70	.62	.65	.49	.28	.52	.41

Note. n = 9,084 – 10,736. All correlations are statistically significant, $p < .05$ (two-tailed).

Table B.3. Descriptive Statistics and Reliability Estimates for Assessment of Individual Motivation (AIM) Scales

Scale	*M*	*SD*	α
Adjustment	1.26	.29	.74
Agreeableness	1.26	.27	.70
Dependability	1.26	.28	.77
Leadership	1.20	.28	.76
Physical Conditioning	1.19	.34	.78
Work Orientation	1.20	.29	.74
Validity Scale	.15	.16	n/a

Note. n = 4,707 – 4,939. α = coefficient alpha. AIM scales scores range from 0 – 2 except for the Validity scale, which ranges from 0 – 1.

Table B.4. Intercorrelations among AIM Scales

Scale		1	2	3	4	5	6
1	Adjustment						
2	Agreeableness	**.63**					
3	Dependability	**.52**	**.52**				
4	Leadership	**.29**	**.17**	**.37**			
5	Physical Conditioning	**.30**	**.29**	**.31**	**.24**		
6	Work Orientation	**.40**	**.32**	**.34**	**.57**	**.54**	
7	Validity Scale	**.11**	**.09**	**.08**	**.04**	.02	**.13**

Note. n = 4,696 – 4,939. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table B.5. Descriptive Statistics for Tailored Adaptive Personality Assessment System (TAPAS-95s) Scales

Scale	Items	*M*	*SD*
Achievement	16	.17	.64
Curiosity	13	-.08	.79
Non-Delinquency	17	.09	.65
Dominance	17	-.15	.61
Even-Temper	13	-.46	.76
Attention-Seeking	14	-.14	.79
Intellectual Efficiency	14	-.19	.64
Order	13	-.04	.64
Physical Conditioning	17	.12	.71
Tolerance	13	-.43	.67
Cooperation/Trust	17	-.30	.86
Optimism	15	-.07	.59

Note. n = 4,637. Scores have a theoretical distribution of approximately -3 to +3.

Table B.6. Intercorrelations among TAPAS-95s Scales

	Scale	1	2	3	4	5	6	7	8	9	10	11
1	Achievement											
2	Curiosity	**.21**										
3	Non-Delinquency	**.17**	**.12**									
4	Dominance	**.15**	**.14**	.02								
5	Even-Temper	**.06**	**.22**	**.12**	**-.05**							
6	Attention-Seeking	**-.11**	**-.11**	**-.37**	**.13**	**-.12**						
7	Intellectual Efficiency	**.16**	**.34**	**.03**	**.15**	**.14**	**-.06**					
8	Order	**.19**	**.05**	**.15**	**.07**	-.02	**-.07**	**.07**				
9	Physical Conditioning	**.19**	**.04**	**-.09**	**.06**	-.01	**.10**	.02	**.05**			
10	Tolerance	**.06**	**.21**	**.06**	**.10**	**.08**	**-.03**	**.15**	**.06**	.01		
11	Cooperation/Trust	.01	**-.05**	**.19**	**-.13**	**.12**	**-.05**	**-.07**	.02	**-.13**	-.01	
12	Optimism	**.06**	**.12**	.03	**.08**	**.22**	**-.03**	**.17**	.00	**.07**	**.09**	**.09**

Note. n = 4,637. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table B.7. Descriptive Statistics and Reliability Estimates for Rational Biodata Inventory (RBI) Scale Scores

Scale	Items	*M*	*SD*	α
Peer Leadership	6	3.60	.65	.71
Cognitive Flexibility	8	3.47	.64	.76
Achievement	9	3.54	.58	.70
Fitness Motivation	7	3.30	.68	.73
Interpersonal Skills - Diplomacy	5	3.65	.75	.71
Stress Tolerance	11	3.01	.51	.67
Hostility to Authority	7	2.52	.65	.68
Self-Efficacy	6	4.02	.62	.78
Cultural Tolerance	5	3.75	.73	.69
Internal Locus of Control	8	3.55	.57	.67
Army Affective Commitment	7	3.73	.69	.71
Respect for Authority	4	3.51	.69	.65
Narcissism	6	3.61	.57	.55
Gratitude	3	3.95	.72	.43
Lie Scale	7	0.09	.14	.51
Pure Fitness Motivation[a]	5	3.40	.72	.70

Note. *n* = 8,625-8,626. Items = number of items comprising each final scale. α = coefficient alpha. RBI scale scores range from 1 – 5, except for the Lie scale, which ranges from 0 – 1.

[a]An alternative version of the Fitness Motivation scale with the ability items removed.

Table B.8. Intercorrelations among RBI Scale Scores

	Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Peer Leadership															
2	Cognitive Flexibility	**.51**														
3	Achievement	**.55**	**.49**													
4	Fitness Motivation	**.29**	**.16**	**.27**												
5	Interpersonal Skills - Diplomacy	**.49**	**.30**	**.38**	**.22**											
6	Stress Tolerance	**.12**	**.14**	**.06**	**.22**	**.24**										
7	Hostility to Authority	**-.10**	**-.18**	**-.25**	**-.05**	**-.18**	**-.37**									
8	Self-Efficacy	**.57**	**.44**	**.56**	**.38**	**.46**	**.24**	**-.19**								
9	Cultural Tolerance	**.35**	**.42**	**.31**	**.13**	**.42**	**.30**	**-.34**	**.40**							
10	Internal Locus of Control	**.31**	**.28**	**.35**	**.21**	**.37**	**.42**	**-.39**	**.45**	**.38**						
11	Army Affective Commitment	**.31**	**.19**	**.29**	**.30**	**.29**	**.22**	**-.20**	**.44**	**.27**	**.34**					
12	Respect for Authority	**.28**	**.29**	**.49**	**.10**	**.20**	-.01	**-.21**	**.30**	**.19**	**.21**	**.19**				
13	Narcissism	**.37**	**.23**	**.34**	**.18**	**.21**	**-.15**	**.15**	**.39**	**.08**	**.10**	**.18**	**.15**			
14	Gratitude	**.27**	**.24**	**.34**	**.12**	**.33**	**.10**	**-.28**	**.35**	**.30**	**.35**	**.24**	**.32**	**.11**		
15	Lie Scale	**.16**	**.15**	**.17**	**.12**	**.12**	**.24**	**-.20**	**.19**	**.20**	**.17**	**.12**	**.09**	**.02**	.01	
16	Pure Fitness Motivation[a]	**.32**	**.20**	**.33**	**.93**	**.24**	**.19**	**-.08**	**.42**	**.17**	**.23**	**.34**	**.14**	**.19**	**.16**	**.13**

Note. n = 8,624-8,626. Statistically significant correlations are bolded, $p < .05$ (two-tailed).
[a]An alternative version of the Fitness Motivation scale with the ability items removed.

Table B.9. Descriptive Statistics and Reliability Estimates for Army Knowledge Assessment (AKA) Scales

Scale	Items	*M*	*SD*	α
Realistic Interests	5	4.05	.61	.76
Investigative Interests	5	3.39	.74	.82
Artistic Interests	5	2.75	.93	.89
Social Interests	5	3.78	.71	.82
Enterprising Interests	5	3.69	.71	.81
Conventional Interests	5	3.93	.69	.84

Note. n = 10,048-10,075. Items = number of items comprising each final scale. α = coefficient alpha. AKA scale scores range from 1 – 5.

Table B.10. Intercorrelations among AKA Scales

Scale	1	2	3	4	5
1 Realistic Interests					
2 Investigative Interests	.39				
3 Artistic Interests	.14	.50			
4 Social Interests	.39	.38	.30		
5 Enterprising Interests	.40	.38	.25	.48	
6 Conventional Interests	.44	.29	.10	.45	.52

Note. n = 10,044 – 10,074. All correlations are statistically significant, $p < .05$ (two-tailed).

Table B.11. Descriptive Statistics and Reliability Estimates for Work Preferences Assessment (WPA) Dimension and Facet Scores

Scale	Items	*M*	*SD*	α
Realistic Interests (D)	13	3.50	.79	.90
Mechanical (F)	5	3.20	1.05	.90
Physical (F)	7	3.73	.84	.89
Investigative Interests (D)	12	3.28	.65	.85
Critical Thinking (F)	6	3.76	.72	.82
Conduct Research (F)	6	2.79	.77	.76
Artistic Interests (D)	12	2.79	.76	.87
Artistic Activities (F)	8	2.39	.86	.85
Creativity (F)	4	3.59	.86	.82
Social Interests (D)	10	3.60	.65	.83
Work with Others (F)	5	3.81	.71	.77
Help Others (F)	5	3.39	.75	.71
Enterprising Interests (D)	13	3.36	.59	.81
Prestige (F)	5	3.88	.66	.68
Lead Others (F)	4	3.56	.74	.70
High Profile (F)	4	2.52	.88	.72
Conventional Interests (D)	12	3.23	.62	.82
Information Management (F)	6	2.63	.84	.81
Detail Orientation (F)	3	3.88	.78	.73
Clear Procedures (F)	3	3.90	.76	.64

Note. n = 9,924-9,926. D = Dimension. F = Facet. Items = number of items comprising each final scale. α = coefficient alpha. WPA scale scores range from 1 – 5.

Table B.12. Intercorrelations among WPA Dimension and Facet Scores

Scale		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	Realistic Interests (D)																			
2	Mechanical (F)	**.83**																		
3	Physical (F)	**.86**	**.45**																	
4	Investigative Interests (D)	**.16**	**.12**	**.15**																
5	Critical Thinking (F)	**.20**	**.09**	**.25**	**.86**															
6	Conduct Research (F)	**.08**	**.12**	**.02**	**.88**	**.52**														
7	Artistic Interests (D)	**.10**	**.18**	.01	**.42**	**.24**	**.47**													
8	Artistic Activities (F)	**.08**	**.17**	**-.03**	**.31**	**.10**	**.43**	**.94**												
9	Creativity (F)	**.12**	**.13**	**.08**	**.47**	**.43**	**.40**	**.76**	**.50**											
10	Social Interests (D)	**.09**	**-.06**	**.19**	**.54**	**.53**	**.41**	**.29**	**.21**	**.35**										
11	Work with Others (F)	**.20**	.00	**.32**	**.45**	**.51**	**.29**	**.20**	**.11**	**.30**	**.88**									
12	Help Others (F)	**-.03**	**-.10**	**.04**	**.50**	**.43**	**.44**	**.32**	**.26**	**.31**	**.90**	**.58**								
13	Enterprising Interests (D)	**.16**	**.06**	**.19**	**.61**	**.57**	**.50**	**.39**	**.30**	**.42**	**.59**	**.54**	**.51**							
14	Prestige (F)	**.18**	**.05**	**.24**	**.50**	**.55**	**.32**	**.19**	**.08**	**.34**	**.50**	**.50**	**.39**	**.80**						
15	Lead Others (F)	**.21**	**.03**	**.31**	**.48**	**.52**	**.32**	**.24**	**.14**	**.35**	**.59**	**.56**	**.49**	**.81**	**.57**					
16	High Profile (F)	.00	**.07**	**-.07**	**.46**	**.28**	**.51**	**.46**	**.46**	**.30**	**.32**	**.23**	**.34**	**.74**	**.33**	**.37**				
17	Conventional Interests (D)	**.12**	**.12**	**.08**	**.61**	**.53**	**.53**	**.26**	**.23**	**.24**	**.55**	**.48**	**.51**	**.59**	**.47**	**.42**	**.48**			
18	Information Management (F)	**-.05**	**.07**	**-.14**	**.49**	**.31**	**.54**	**.36**	**.37**	**.22**	**.41**	**.28**	**.44**	**.51**	**.28**	**.29**	**.61**	**.85**		
19	Detail Orientation (F)	**.24**	**.13**	**.28**	**.53**	**.61**	**.32**	**.07**	**-.02**	**.23**	**.47**	**.49**	**.36**	**.42**	**.48**	**.39**	**.13**	**.69**	**.30**	
20	Clear Procedures (F)	**.21**	**.11**	**.24**	**.48**	**.54**	**.30**	**.05**	**-.02**	**.18**	**.49**	**.49**	**.39**	**.41**	**.48**	**.37**	**.14**	**.72**	**.33**	**.89**

Note. n = 9,924-9,926. D = Dimension. F = Facet. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

APPENDIX C
SCALE-LEVEL CORRELATIONS BETWEEN SELECTED PREDICTOR AND IN-UNIT CRITERION MEASURES

Table C.1. Correlations between all Experimental Predictors and Select In-Unit 1 Performance-Related Can-Do Criteria

Measure/Scale	Can-Do Criteria							
	MOS-Specific Job Knowledge Test (JKT)		Warrior Tasks and Battle Drills (WTBD) JKT		Performing MOS-Specific Tasks AW PRS		Cognitive Performance AW PRS	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AFQT	626	**.36**	1,365	**.51**	889	**.11**	909	**.16**
Education Tier	628	.01	1,374	.03	894	**-.10**	914	**-.09**
Assembling Objects (AO)	583	**.30**	1,269	**.33**	823	**.15**	842	**.17**
AIM								
Adjustment	227	-.03	549	**.13**	389	-.06	398	-.06
Agreeableness	221	.05	534	**.10**	375	-.02	384	-.04
Dependability	226	.04	546	.08	387	.04	396	.04
Leadership	228	.01	550	**.11**	391	.01	400	.02
Physical Conditioning	224	.04	541	.03	382	-.02	391	.02
Work Orientation	222	-.06	540	.05	379	-.05	388	.02
TAPAS-95S								
Achievement	222	.05	528	.06	379	.04	389	.06
Curiosity	222	**.19**	528	**.17**	379	.03	389	.02
Non-Delinquency	222	.09	528	-.01	379	.08	389	.08
Dominance	222	-.01	528	.05	379	.04	389	.07
Even-Temper	222	.12	528	**.15**	379	-.02	389	-.04
Attention-Seeking	222	-.09	528	-.08	379	.00	389	-.05
Intellectual Efficiency	222	**.17**	528	**.26**	379	.05	389	.08
Order	222	.04	528	-.02	379	.01	389	.02
Physical Condition	222	-.11	528	-.02	379	.08	389	.09
Tolerance	222	.05	528	-.01	379	-.01	389	.00
Cooperation/Trust	222	.08	528	-.05	379	-.04	389	-.01
Optimism	222	.09	528	**.19**	379	.04	389	.10
PSJT	375	**.23**	706	**.27**	429	.09	439	**.16**
RBI								
Peer Leadership	522	**.11**	1,114	.05	708	**.08**	727	.06
Cognitive Flexibility	522	**.14**	1,114	**.19**	708	.05	727	.05
Achievement	522	.03	1,114	.01	708	**.08**	727	.04
Fitness Motivation	522	.03	1,114	.06	708	.02	727	.04
Interpersonal Skills/Diplomacy	522	.04	1,114	.03	708	.01	727	.04
Stress Tolerance	522	.06	1,114	**.13**	708	-.01	727	.05
Hostility to Authority	522	**-.12**	1,114	**-.13**	708	-.05	727	**-.08**
Self-efficacy	522	.08	1,114	**.06**	708	.03	727	.02
Cultural Tolerance	522	-.03	1,114	**.07**	708	-.01	727	.01
Internal Locus of Control	522	**.12**	1,114	**.18**	708	.00	727	.03
Army Affective Commitment	522	**.13**	1,114	**.12**	708	-.01	727	.03
Respect for Authority	522	.04	1,114	.02	708	.05	727	**.10**
Narcissism	522	.05	1,114	-.04	708	.03	727	-.03
Gratitude	522	**.11**	1,114	**.13**	708	**.10**	727	**.13**

Table C.1. (Continued)

Measure/Scale	Can-Do Criteria							
	MOS-Specific Job Knowledge Test (JKT)		Warrior Tasks and Battle Drills (WTBD) JKT		Performing MOS-Specific Tasks AW PRS		Cognitive Performance AW PRS	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AKA								
Realistic	592	.04	1,283	**.08**	827	.04	846	.06
Investigative	592	**-.13**	1,283	**-.13**	827	.00	846	-.02
Artistic	592	**-.10**	1,283	**-.19**	827	-.04	846	**-.08**
Social	592	**.10**	1,283	-.01	827	.01	846	-.03
Enterprising	592	.02	1,283	.02	827	.00	846	.04
Conventional	589	**.14**	1,278	**.10**	823	**.10**	842	**.08**
WPA Dimensions								
Realistic	596	-.03	1,277	.05	826	-.06	846	-.06
Investigative	596	-.03	1,277	.03	826	.03	846	.01
Artistic	596	**-.11**	1,277	**-.11**	826	-.04	846	-.05
Social	596	**-.14**	1,277	**-.14**	826	.03	846	.00
Enterprising	596	-.06	1,277	**-.09**	826	.03	846	.01
Conventional	596	**-.10**	1,277	**-.19**	826	.02	846	-.04
WPA Facets								
Mechanical	596	.00	1,277	.04	826	-.06	846	-.05
Physical	596	-.05	1,277	.05	826	-.05	846	-.03
Critical Thinking	596	.02	1,277	**.12**	825	.06	846	.04
Conduct Research	596	-.07	1,277	**-.07**	826	.00	846	-.02
Artistic Activities	596	**-.12**	1,277	**-.13**	826	**-.07**	846	**-.08**
Creativity	596	-.05	1,277	-.01	826	.03	846	.01
Work with Others	596	**-.13**	1,277	**-.12**	826	.03	846	-.02
Help Others	596	**-.11**	1,277	**-.14**	826	.03	846	.01
Prestige	596	.02	1,277	.01	826	.04	846	.05
Lead Others	596	**-.09**	1,277	-.04	826	.04	846	.03
High Profile	596	**-.08**	1,277	**-.18**	826	-.01	846	-.05
Information Management	596	**-.11**	1,277	**-.24**	826	.00	846	-.05
Detail Orientation	596	-.02	1,277	-.03	826	.03	846	-.01
Clear Procedures	596	-.03	1,277	**-.07**	826	.00	846	.00

Note. Correlations in bold are statistically significant, $p < .05$.

Table C.2. Correlations between all Experimental Predictors and Select In-Unit 1 Performance-Related Will-Do Criteria

Measure/Scale	Will-Do Criteria							
	Effort and Discipline (Army-Wide [AW] Performance Rating Scales [PRS])		Working Effectively with Others (AW PRS)		Last Army Physical Fitness Test (APFT) Score (ALQ)		Disciplinary Incidents (ALQ)	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AFQT	909	**.08**	909	**.12**	1,305	.03	1,400	-.04
Education Tier	914	**-.14**	914	**-.12**	1,314	.03	1,409	**.08**
Assembling Objects (AO)	842	**.15**	842	**.17**	1,217	**.06**	1,302	**-.07**
AIM								
Adjustment	398	-.04	398	-.05	529	.08	578	-.05
Agreeableness	384	-.01	384	-.02	519	**.10**	564	**-.09**
Dependability	396	.08	396	.06	528	.00	575	**-.12**
Leadership	400	.01	400	.01	531	.07	579	.00
Physical Conditioning	391	.05	391	.04	524	**.27**	571	-.07
Work Orientation	388	.00	388	-.01	524	**.21**	570	-.05
TAPAS-95S								
Achievement	389	.01	389	.00	504	.03	551	-.07
Curiosity	389	.00	389	.01	504	-.03	551	-.06
Non-Delinquency	389	**.10**	389	.08	504	-.07	551	**-.13**
Dominance	389	.03	389	.00	504	-.01	551	-.05
Even-Temper	389	.00	389	-.04	504	.01	551	-.02
Attention-Seeking	389	-.06	389	-.06	504	.03	551	**.12**
Intellectual Efficiency	389	.02	389	.03	504	.03	551	-.03
Order	389	-.03	389	.02	504	.05	551	-.04
Physical Condition	389	.07	389	.09	504	**.29**	551	-.03
Tolerance	389	-.03	389	-.02	504	.02	551	.02
Cooperation/Trust	389	-.03	389	-.03	504	**-.11**	551	.00
Optimism	389	.06	389	.05	504	.04	551	-.02
PSJT	439	**.14**	439	**.16**	668	.03	707	-.06
RBI								
Peer Leadership	727	.01	727	**.08**	1,075	**.11**	1,145	.04
Cognitive Flexibility	727	-.03	727	.05	1,075	.03	1,145	.02
Achievement	727	.01	727	.05	1,075	**.07**	1,145	-.03
Fitness Motivation	727	.03	727	.06	1,075	**.35**	1,145	.02
Interpersonal Skills/Diplomacy	727	-.01	727	.04	1,075	**.09**	1,145	.04
Stress Tolerance	727	.00	727	.06	1,075	**.11**	1,145	-.02
Hostility to Authority	727	**-.09**	727	**-.08**	1,075	.02	1,145	**.17**
Self-efficacy	727	.01	727	.05	1,075	**.14**	1,145	.04
Cultural Tolerance	727	-.04	727	.03	1,075	.05	1,145	-.03
Internal Locus of Control	727	.00	727	.03	1,075	**.11**	1,145	-.02
Army Affective Commitment	727	.01	727	.04	1,075	**.08**	1,145	.00
Respect for Authority	727	.06	727	**.08**	1,075	.04	1,145	**-.07**
Narcissism	727	-.05	727	.00	1,075	.04	1,145	**.07**
Gratitude	727	**.11**	727	**.13**	1,075	.04	1,145	**-.12**

Table C.2. (Continued)

Measure/Scale	Will-Do Criteria							
	Effort and Discipline (Army-Wide [AW] Performance Rating Scales [PRS])		Working Effectively with Others (AW PRS)		Last Army Physical Fitness Test (APFT) Score (ALQ)		Disciplinary Incidents (ALQ)	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AKA								
Realistic	846	**.07**	846	**.07**	1,228	.03	1,315	**-.06**
Investigative	846	.00	846	.01	1,228	.01	1,315	.01
Artistic	846	-.03	846	**-.07**	1,228	-.03	1,315	.01
Social	846	-.02	846	.00	1,228	-.01	1,315	.01
Enterprising	846	.03	846	.02	1,228	.04	1,315	-.03
Conventional	842	**.08**	842	**.10**	1,224	-.02	1,311	-.05
WPA Dimensions								
Realistic	846	-.01	846	.00	1,222	**.09**	1,308	.04
Investigative	846	-.01	846	.03	1,222	.02	1,308	.00
Artistic	846	-.04	846	-.04	1,222	-.02	1,308	-.01
Social	846	.00	846	.01	1,222	.02	1,308	-.02
Enterprising	846	.01	846	.03	1,222	.01	1,308	-.01
Conventional	846	-.03	846	-.01	1,222	-.03	1,308	.02
WPA Facets								
Mechanical	846	-.01	846	-.01	1,222	.00	1,308	.03
Physical	846	.00	846	.00	1,222	**.17**	1,308	.04
Critical Thinking	845	.02	845	.06	1,222	.05	1,307	-.01
Conduct Research	846	-.03	846	.00	1,222	-.02	1,308	.01
Artistic Activities	846	**-.07**	846	**-.07**	1,222	-.04	1,308	.00
Creativity	846	.03	846	.04	1,222	.02	1,308	-.02
Work with Others	846	.00	846	.01	1,222	.02	1,308	-.01
Help Others	846	.00	846	.02	1,222	.01	1,308	-.04
Prestige	846	.04	846	.06	1,222	.00	1,308	-.02
Lead Others	846	.05	846	.05	1,222	.03	1,308	-.02
High Profile	846	-.05	846	-.04	1,222	-.01	1,308	.00
Information Management	846	-.05	846	-.05	1,222	**-.06**	1,308	.02
Detail Orientation	846	.01	846	.05	1,222	.05	1,308	-.01
Clear Procedures	846	.01	846	.05	1,222	.03	1,308	-.01

Note. Correlations in bold are statistically significant, $p < .05$.

Table C.3. Correlations between all Experimental Predictors and Select In-Unit 2 Performance-Related Can-Do Criteria

Measure/Scale	Can-Do Criteria							
	MOS-Specific JKT		WTBD JKT		Performing MOS-Specific Tasks AW PRS		Cognitive Performance AW PRS	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AFQT	429	**.32**	921	**.42**	712	**.08**	735	**.10**
Education Tier	430	-.04	928	-.05	716	-.02	740	-.03
Assembling Objects (AO)	402	**.11**	858	**.23**	664	.05	683	**.10**
AIM								
Adjustment	195	.08	427	**.10**	328	.01	343	.00
Agreeableness	190	.09	417	.10	315	.05	330	.08
Dependability	192	.11	422	.06	321	.02	336	.05
Leadership	197	.08	429	.07	327	.02	342	.01
Physical Conditioning	193	.01	423	.03	321	**.15**	336	.09
Work Orientation	192	-.08	421	-.03	320	.07	335	.03
TAPAS-95S								
Achievement	200	.03	419	.06	322	.07	335	.09
Curiosity	200	.05	419	.07	322	-.01	335	.02
Non-Delinquency	200	**.21**	419	.06	322	.08	335	.10
Dominance	200	-.01	419	.01	322	-.05	335	-.04
Even-Temper	200	**.21**	419	.05	322	.05	335	-.01
Attention-Seeking	200	**-.17**	419	-.09	322	-.06	335	-.09
Intellectual Efficiency	200	**.20**	419	**.21**	322	-.05	335	-.02
Order	200	-.07	419	-.05	322	.04	335	.01
Physical Condition	200	-.12	419	.02	322	.08	335	.08
Tolerance	200	.05	419	-.02	322	.01	335	-.03
Cooperation/Trust	200	.01	419	-.05	322	-.01	335	.02
Optimism	200	.12	419	**.12**	322	-.07	335	-.07
PSJT	208	**.35**	424	**.25**	331	.05	338	.08
RBI								
Peer Leadership	348	.02	735	.03	554	.00	572	-.03
Cognitive Flexibility	348	**.14**	735	**.13**	554	-.03	572	-.08
Achievement	348	-.01	735	.00	554	.01	572	-.01
Fitness Motivation	348	.07	735	**.08**	554	.06	572	.02
Interpersonal Skills/Diplomacy	348	.00	735	-.03	554	-.03	572	-.02
Stress Tolerance	348	.07	735	**.10**	554	.01	572	.01
Hostility to Authority	348	-.10	735	**-.12**	554	.01	572	-.01
Self-efficacy	348	.05	735	**.08**	554	-.06	572	-.04
Cultural Tolerance	348	.07	735	.07	554	-.07	572	-.06
Internal Locus of Control	348	.03	735	**.07**	554	.01	572	-.03
Army Affective Commitment	348	**.14**	735	**.11**	554	.02	572	-.01
Respect for Authority	348	.09	734	**.08**	553	.05	571	.01
Narcissism	348	.06	735	-.01	554	-.07	572	-.07
Gratitude	348	.08	735	**.10**	554	-.02	572	-.01

Table C.3. (Continued)

Measure/Scale	Can-Do Criteria							
	MOS-Specific JKT		WTBD JKT		Performing MOS-Specific Tasks AW PRS		Cognitive Performance AW PRS	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AKA								
Realistic	408	.06	861	.02	664	-.02	688	.00
Investigative	408	-.03	861	**-.11**	664	-.04	688	-.05
Artistic	407	**-.11**	860	**-.21**	663	-.05	687	-.06
Social	408	.08	861	-.01	664	-.03	688	-.01
Enterprising	408	.06	861	-.02	664	-.04	688	-.02
Conventional	406	.07	857	**.08**	660	.02	684	.00
WPA Dimensions								
Realistic	409	.05	863	**.09**	661	.01	684	.02
Investigative	409	.00	863	-.04	661	-.03	684	-.06
Artistic	409	-.03	863	**-.13**	661	-.07	684	-.06
Social	409	-.02	863	**-.14**	661	-.02	684	-.02
Enterprising	409	-.08	863	**-.08**	661	-.06	684	-.01
Conventional	409	**-.11**	863	**-.18**	661	-.04	684	-.04
WPA Facets								
Mechanical	409	.07	863	.05	661	.01	684	.03
Physical	409	-.01	863	**.09**	661	.00	684	.00
Critical Thinking	409	.08	863	.04	661	-.02	684	-.05
Conduct Research	409	-.08	863	**-.12**	661	-.03	684	-.05
Artistic Activities	409	-.07	863	**-.15**	661	-.06	684	-.05
Creativity	409	.05	863	-.04	661	-.06	684	-.05
Work with Others	409	-.02	863	**-.11**	661	-.04	684	-.05
Help Others	409	-.01	863	**-.13**	661	.01	684	.01
Prestige	409	-.02	863	.01	661	-.05	684	-.02
Lead Others	409	-.02	863	-.02	661	-.06	684	.01
High Profile	409	**-.14**	863	**-.16**	661	-.03	684	.00
Information Management	409	**-.13**	863	**-.24**	661	-.05	684	-.04
Detail Orientation	409	-.01	863	-.05	661	-.03	684	-.05
Clear Procedures	409	-.03	863	-.05	661	.02	684	-.02

Note. Correlations in bold are statistically significant, $p < .05$.

Table C.4. Correlations between all Experimental Predictors and Select In-Unit 2 Performance-Related Will-Do Criteria

Measure/Scale	Will-Do Criteria							
	Effort and Discipline (Army-Wide [AW] Performance Rating Scales [PRS])		Working Effectively with Others (AW PRS)		Last Army Physical Fitness Test (APFT) Score (ALQ)		Qualifications and Awards (ALQ)	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AFQT	736	**.08**	736	**.12**	918	**-.09**	934	-.02
Education Tier	741	-.03	741	-.04	925	.02	941	-.01
Assembling Objects (AO)	684	.06	684	**.11**	855	-.04	871	-.01
AIM								
Adjustment	344	.05	344	.05	427	**.14**	434	.00
Agreeableness	331	**.11**	331	.09	417	.03	425	-.02
Dependability	337	**.11**	337	.10	423	.07	430	-.01
Leadership	343	-.03	343	.01	429	**.16**	436	.09
Physical Conditioning	337	**.14**	337	**.12**	423	**.29**	430	.06
Work Orientation	336	.04	336	.06	421	**.21**	428	.06
TAPAS-95S								
Achievement	336	.04	336	.07	416	**.11**	425	-.03
Curiosity	336	.01	336	.03	416	.08	425	**.13**
Non-Delinquency	336	**.11**	336	.11	416	.02	425	.00
Dominance	336	-.06	336	.00	416	-.01	425	.02
Even-Temper	336	.02	336	.02	416	-.05	425	-.01
Attention-Seeking	336	**-.16**	336	-.09	416	.02	425	-.01
Intellectual Efficiency	336	-.05	336	.00	416	-.05	425	.04
Order	336	.03	336	.08	416	-.01	425	.02
Physical Condition	336	.08	336	**.12**	416	**.29**	425	.06
Tolerance	336	-.01	336	-.02	416	.02	425	-.02
Cooperation/Trust	336	.00	336	.01	416	-.06	425	-.02
Optimism	336	-.07	336	-.05	416	-.07	425	-.07
PSJT	338	.09	338	.08	421	-.05	427	-.04
RBI								
Peer Leadership	573	-.02	573	-.01	731	**.09**	744	.06
Cognitive Flexibility	573	-.04	573	-.04	731	.02	744	-.02
Achievement	573	.03	573	.01	731	**.09**	744	.06
Fitness Motivation	573	.07	573	.05	731	**.31**	744	.05
Interpersonal Skills/Diplomacy	573	-.02	573	-.01	731	.05	744	-.02
Stress Tolerance	573	.02	573	.00	731	.02	744	-.01
Hostility to Authority	573	-.07	573	-.03	731	.07	744	.04
Self-efficacy	573	-.05	573	-.02	731	**.10**	744	-.03
Cultural Tolerance	573	-.01	573	-.06	731	-.04	744	-.03
Internal Locus of Control	573	.02	573	.01	731	.06	744	.00
Army Affective Commitment	573	.03	573	.01	731	.05	744	.01
Respect for Authority	572	.06	572	.05	730	.03	743	**.12**
Narcissism	573	-.04	573	-.03	731	**.08**	744	.04
Gratitude	573	.03	573	.02	731	-.05	744	.04

Table C.4. (Continued)

Measure/Scale	Will-Do Criteria							
	Effort and Discipline (Army-Wide [AW] Performance Rating Scales [PRS])		Working Effectively with Others (AW PRS)		Last Army Physical Fitness Test (APFT) Score (ALQ)		Qualifications and Awards (ALQ)	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AKA								
Realistic	688	.03	688	.02	858	.01	873	-.02
Investigative	688	-.03	688	-.03	858	.07	873	.02
Artistic	687	-.02	687	-.04	858	.01	873	.02
Social	688	.00	688	-.03	858	-.06	873	.00
Enterprising	688	-.03	688	-.03	858	-.02	873	-.01
Conventional	684	-.01	684	-.01	854	-.03	869	.02
WPA Dimensions								
Realistic	684	.03	684	.01	857	**.09**	873	.06
Investigative	684	-.05	684	-.06	857	.04	873	-.01
Artistic	684	-.04	684	-.05	857	.02	873	-.02
Social	684	.01	684	-.01	857	.03	873	.01
Enterprising	684	.01	684	-.01	857	**.12**	873	.06
Conventional	684	-.03	684	-.06	857	.00	873	.00
WPA Facets								
Mechanical	684	.03	684	.03	857	.01	873	.05
Physical	684	.01	684	-.01	857	.16	873	.04
Critical Thinking	684	-.05	684	-.05	857	.04	873	-.02
Conduct Research	684	-.03	684	-.05	857	.04	873	.00
Artistic Activities	684	-.02	684	-.03	857	.01	873	-.02
Creativity	684	-.05	684	-.06	857	.02	873	-.02
Work with Others	684	-.02	684	-.04	857	**.08**	873	.03
Help Others	684	.04	684	.02	857	-.03	873	-.01
Prestige	684	-.01	684	-.01	857	**.08**	873	.01
Lead Others	684	-.02	684	-.01	857	**.11**	873	.06
High Profile	684	.03	684	.00	857	**.09**	873	**.07**
Information Management	684	-.02	684	-.03	857	-.03	873	-.01
Detail Orientation	684	-.06	684	-.07	857	.03	873	.00
Clear Procedures	684	.00	684	-.03	857	.04	873	-.01

Note. Correlations in bold are statistically significant, $p < .05$.

Table C.5. Correlations between all Experimental Predictors and In-Unit Combat Performance and Deployment Adjustment Criteria

Measure/Scale	Combat Performance and Deployment Adjustment Criteria													
	Field/Combat Judgment (Combat/ Deployment Performance Rating Scales [CDPRS])		Field Readiness (CDPRS)		Physical Courage (CDPRS)		Awareness and Vigilance (CDPRS)		Combat Performance Ratings Composite (CDPRS)		In-Unit 1 Deployment Adjustment (ALQ)		In-Unit 2 Deployment Adjustment (ALQ)	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*
AFQT	315	.03	318	-.03	309	.00	315	-.03	319	-.01	405	**.10**	777	**.08**
Education Tier	317	.10	320	-.02	311	.03	317	-.04	321	.01	405	-.08	781	-.01
Assembling Objects (AO)	294	-.01	297	.02	288	-.04	294	-.10	298	-.04	371	**.11**	723	.02
AIM														
Adjustment	145	-.09	147	-.08	144	.05	146	-.08	147	-.07	144	**.20**	372	-.02
Agreeableness	141	.01	143	.06	140	.08	142	.09	143	.06	141	**.27**	364	-.07
Dependability	143	-.01	145	-.04	142	.06	144	.09	145	.01	144	**.20**	368	-.07
Leadership	146	.04	148	.05	145	-.03	147	.00	148	-.02	146	**.22**	375	.04
Physical Conditioning	142	.13	144	.11	141	.14	143	.09	144	.13	144	.00	368	.07
Work Orientation	142	.04	144	.02	141	.05	143	-.03	144	.00	143	**.25**	367	**.11**
TAPAS-95S														
Achievement	147	.12	148	.06	146	.03	147	.01	148	.06	149	.09	357	.06
Curiosity	147	-.04	148	-.04	146	.01	147	-.06	148	-.04	149	.10	357	-.06
Non-Delinquency	147	-.03	148	-.02	146	-.02	147	.03	148	.00	149	.09	357	-.05
Dominance	147	-.01	148	.02	146	-.03	147	-.04	148	-.03	149	.06	357	.01
Even-Temper	147	-.01	148	.02	146	.02	147	.03	148	.02	149	-.08	357	**-.11**
Attention-Seeking	147	**-.19**	148	-.10	146	**-.24**	147	**-.18**	148	**-.22**	149	-.06	357	-.02
Intellectual Efficiency	147	-.09	148	-.05	146	-.15	147	-.12	148	-.11	149	.09	357	-.04
Order	147	.01	148	-.01	146	-.01	147	-.03	148	-.02	149	.00	357	-.05
Physical Condition	147	.10	148	.12	146	.05	147	.02	148	.07	149	.02	357	**.13**
Tolerance	147	-.06	148	-.06	146	.03	147	-.03	148	-.04	149	-.01	357	-.06
Cooperation/Trust	147	-.14	148	-.10	146	-.15	147	-.08	148	-.12	149	**-.18**	357	-.10
Optimism	147	-.13	148	-.08	146	-.10	147	**-.20**	148	-.15	149	.09	357	-.05
PSJT	155	.05	156	.13	150	.00	154	.14	157	.10	234	.04	343	**.19**

Table C.5. (Continued)

Measure/Scale	Combat Performance and Deployment Adjustment Criteria													
	Field/Combat Judgment (Combat/ Deployment Performance Rating Scales [CDPRS])		Field Readiness (CDPRS)		Physical Courage (CDPRS)		Awareness and Vigilance (CDPRS)		Combat Performance Ratings Composite (CDPRS)		In-Unit 1 Deployment Adjustment (ALQ)		In-Unit 2 Deployment Adjustment (ALQ)	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*
RBI														
Peer Leadership	248	.09	252	.00	245	-.07	248	.02	252	.02	340	**.11**	615	**.10**
Cognitive Flexibility	248	-.06	252	-.09	245	-.04	248	-.04	252	-.08	340	**.14**	615	.07
Achievement	248	.07	252	.00	245	.04	248	.06	252	.04	340	**.17**	615	.05
Fitness Motivation	248	.12	252	.06	245	.09	248	.10	252	.11	340	.04	615	.06
Interpersonal Skills/Diplomacy	248	.01	252	.03	245	.03	248	-.01	252	.01	340	**.21**	615	-.02
Stress Tolerance	248	-.02	252	-.05	245	-.04	248	-.05	252	-.06	340	.10	615	.02
Hostility to Authority	248	-.03	252	-.04	245	-.05	248	-.07	252	-.04	340	**-.13**	615	.00
Self-efficacy	248	.06	252	-.01	245	-.02	248	-.02	252	.00	340	**.13**	615	.08
Cultural Tolerance	248	-.12	252	**-.15**	245	-.09	248	-.07	252	-.15	340	**.15**	615	.04
Internal Locus of Control	248	-.03	252	-.03	245	.06	248	.03	252	-.01	340	**.14**	615	**.09**
Army Affective Commitment	248	.02	252	-.01	245	.00	248	.05	252	.00	340	**.13**	615	.05
Respect for Authority	247	-.03	251	-.07	244	-.04	247	-.02	251	-.05	340	**.16**	614	**.14**
Narcissism	248	-.01	252	-.05	245	-.03	248	.00	252	-.04	340	-.05	615	.06
Gratitude	248	-.10	252	-.11	245	-.07	248	-.04	252	-.11	340	.12	615	.01
AKA														
Realistic	299	.07	302	-.04	294	.02	299	.09	303	.05	379	**.10**	724	**.09**
Investigative	299	-.03	302	-.10	294	-.04	299	.03	303	-.04	379	.06	724	.04
Artistic	299	-.01	302	-.08	294	-.06	299	.03	303	-.04	379	**.10**	724	.03
Social	299	-.06	302	-.11	294	-.10	299	-.09	303	-.09	379	**.13**	724	**.11**
Enterprising	299	-.03	302	**-.12**	294	-.04	299	-.06	303	-.08	379	**.12**	724	**.08**
Conventional	299	-.02	302	-.05	294	-.05	299	-.05	303	-.03	376	**.11**	721	**.13**
WPA Dimensions														
Realistic	292	.11	296	.00	287	.04	292	.05	296	.05	383	.02	720	.02
Investigative	292	.02	296	.02	287	.08	292	.01	296	.03	383	**.12**	720	.03
Artistic	292	.00	296	.02	287	.04	292	.07	296	.02	383	.04	720	-.03

Table C.5. (Continued)

Measure/Scale	Combat Performance and Deployment Adjustment Criteria													
	Field/Combat Judgment (Combat/ Deployment Performance Rating Scales [CDPRS])		Field Readiness (CDPRS)		Physical Courage (CDPRS)		Awareness and Vigilance (CDPRS)		Combat Performance Ratings Composite (CDPRS)		In-Unit 1 Deployment Adjustment (ALQ)		In-Unit 2 Deployment Adjustment (ALQ)	
	n	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*
Social	292	-.02	296	.02	287	.07	292	.07	296	.04	383	.04	720	.01
Enterprising	292	.06	296	.05	287	.09	292	.08	296	.07	383	.07	720	.03
Conventional	292	-.01	296	.02	287	.01	292	.00	296	.01	383	.06	720	.03
WPA Facets														
Mechanical	292	**.12**	296	.05	287	.05	292	.06	296	.06	383	.02	720	.01
Physical	292	.08	296	-.03	287	.04	292	.03	296	.03	383	.00	720	.02
Critical Thinking	292	-.05	296	-.04	287	.01	292	-.05	296	-.04	383	**.12**	720	.03
Conduct Research	292	.07	296	.07	287	**.12**	292	.08	296	.09	383	**.10**	720	.01
Artistic Activities	292	.00	296	.00	287	.05	292	.08	296	.02	383	.02	720	-.02
Creativity	292	.00	296	.04	287	.02	292	.04	296	.00	383	.08	720	-.03
Work with Others	292	-.08	296	-.03	287	.02	292	.02	296	-.02	383	.06	720	-.01
Help Others	292	.04	296	.06	287	.11	292	.11	296	.10	383	.01	720	.02
Prestige	292	.04	296	-.04	287	.06	292	.00	296	.00	383	.09	720	.01
Lead Others	292	.04	296	.06	287	.07	292	.04	296	.05	383	.04	720	.01
High Profile	292	.07	296	.10	287	.10	292	**.13**	296	.10	383	.03	720	.06
Information Management	292	.02	296	.07	287	.06	292	.07	296	.07	383	.04	720	.01
Detail Orientation	292	-.07	296	-.04	287	-.10	292	-.11	296	-.10	383	.04	720	.03
Clear Procedures	292	-.02	296	-.04	287	-.06	292	-.09	296	-.07	383	.05	720	.02

Note. Correlations in bold are statistically significant, $p < .05$.