Technical Report 1332

Tier One Performance Screen Initial Operational Test and Evaluation: 2012 Interim Report

Deirdre J. Knapp, Editor
Human Resources Research Organization

Kate A. LaPort, Editor
U.S. Army Research Institute

December 2013

**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 by

Human Resources Research Organization

Technical Review by

J. Douglas Dressel, U.S. Army Research Institute
Nehama Babin, U.S. Army Research Institute

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) December 2013	2. REPORT TYPE Interim	3. DATES COVERED (from. . . to) August 2009 to March 2013

4. TITLE AND SUBTITLE Tier One Performance Screen Initial Operational Test and Evaluation: 2012 Interim Report	5a. CONTRACT OR GRANT NUMBER W5J9CQ-10-C-0031
	5b. PROGRAM ELEMENT NUMBER 62785
6. EDITORS: Deirdre J Knapp; Kate A. LaPort	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 329
	5e. WORK UNIT NUMBER
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 and Social Sciences 6000 6th Street (Bldg 1464 / Mail Stop 5610) Fort Belvoir, VA 22060	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1332

12. DISTRIBUTION/AVAILABILITY STATEMENT: Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

   Contracting Officer's Representative and Subject Matter Expert POC: Dr. Kate LaPort

14. ABSTRACT *(Maximum 200 words)*:

In addition to educational, medical, and moral screens, the U.S. Army relies on a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), the Armed Forces Qualification Test (AFQT), to select new Soldiers. Although the AFQT is useful for selecting new Soldiers, other personal attributes are important to Soldier performance and retention. Based on previous U.S. Army Research Institute (ARI) investigations, the Army selected one promising measure, the Tailored Adaptive Personality Assessment System (TAPAS), for an initial operational test and evaluation (IOT&E), beginning administration to applicants in 2009. Criterion data are being compiled at 6-month intervals from administrative records, from schools for selected military occupational specialties (MOS), and from Soldiers in units. This is the fifth in a series of planned evaluations of the TAPAS. Similar to prior research, the cumulative results thus far suggest that several TAPAS scales significantly predict a number of criteria of interest, indicating that the measure holds promise for both selection and classification purposes.

15. SUBJECT TERMS
 Personnel, Manpower, Selection and classification

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	139	Kate LaPort

Technical Report 1332

Tier One Performance Screen Initial Operational Test and Evaluation: 2012 Interim Report

Deirdre J. Knapp, Editor
Human Resources Research Organization

Kate A. LaPort, Editor
U.S. Army Research Institute

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, VA 22060

December 2013

ACKNOWLEDGEMENTS

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL
TEST AND EVALUATION: 2012 INTERIM REPORT

EXECUTIVE SUMMARY

Research Requirement:

In addition to educational, physical, and moral screens, the U.S. Army relies on the Armed Forces
Qualification Test (AFQT), a composite score from the Armed Services Vocational Aptitude
Battery (ASVAB), to select new Soldiers into the Army. Although the AFQT has proven to be and
will continue to serve as a useful metric for selecting new Soldiers, other personal attributes, in
particular non-cognitive attributes (e.g., temperament, interests, and values), are important to entry-
level Soldier performance and retention (e.g., Campbell & Knapp, 2001; Ingerick, Diaz, & Putka,
2009; Knapp & Heffner, 2009, 2010; Knapp & Tremble, 2007). Based on previous research
(Knapp & Heffner, 2010), the Army selected one particularly promising measure, the Tailored
Adaptive Personality Assessment System (TAPAS), as the basis for an initial operational test and
evaluation (IOT&E) of the *Tier One Performance Screen* (TOPS). The TAPAS capitalizes on the
latest advances in testing technology to assess motivation through the measurement of personality
characteristics.

Procedure:

In May 2009, the Military Entrance Processing Command (MEPCOM) began administering the
TAPAS on the computer adaptive platform for the ASVAB (CAT-ASVAB) at Military Entrance
Processing Stations (MEPS). To evaluate the TAPAS, outcome (criterion) data are being collected at
multiple points in time from Soldiers who took the TAPAS at entry. Specifically, initial military
training (IMT) criterion data are being collected at schools for Soldiers in eight military
occupational specialties (MOS). Project teams are also collecting criterion data from Soldiers
(regardless of MOS) in their units in multiple waves of site visits during the course of the IOT&E.

The criterion measures include job knowledge tests, an attitudinal assessment (the Army Life
Questionnaire), and performance rating scales completed by the Soldiers' cadre members (in IMT) or
supervisors (in units). Course grades, completion rates, and attrition status are obtained from
administrative records for all Soldiers.

A data file containing TAPAS data collected through March 2012 and criterion data collected
through May 2012 is the basis for the analyses documented in this report. It consists of a total of
282,563 applicants who took the TAPAS; 250,884 of these individuals were in the TOPS
"Applicant Sample." The Applicant Sample (used for analysis purposes) excluded Education
Tier 3, AFQT Category V, and prior service applicants. The validation sample sizes were
considerably smaller, with the IMT Validation Sample comprising 13,238 Soldiers, the In-Unit
Validation Sample comprising 604 Soldiers, and the Administrative Validation Sample (which
includes Soldiers with criterion data [e.g., attrition] from at least one administrative source,
comprising 113,618 Soldiers.

Our approach to analyzing the TAPAS' incremental predictive validity was consistent with previous evaluations of this measure and similar experimental non-cognitive predictors (Ingerick et al., 2009; Knapp & Heffner, 2009, 2010, 2011). In brief, this approach involved testing a series of hierarchical regression models, regressing scores on each criterion measure onto Soldiers' AFQT scores or education tier in the first step, followed by their TOPS composite or TAPAS scale scores in the second step. The resulting increment in the multiple correlation value (ΔR) when the TOPS composite or TAPAS scale scores were added to the baseline regression models served as our index of incremental validity. Scale-level correlations between the TAPAS scale scores and selected criteria were also examined.

We evaluated the TAPAS' potential for improving new Soldier classification using (a) Horst's (1954, 1955) index of differential validity (H_d) and (b) Brogden's expected criterion scores of optimally assigned individuals (De Corte, 2000). Similar to the incremental predictive validity analyses, our approach involved comparing these indices when Soldiers were matched to MOS using scores on the ASVAB or scores on the ASVAB and the TAPAS, combined.

Findings:

Results indicate that the TAPAS holds promise for predicting important first-term criteria over and above the AFQT, primarily for predicting will-do criterion measures (i.e., those measuring nontechnical aspects of Soldier performance, such as physical fitness, effort, and personal discipline).The TAPAS exhibited its greatest predictive gains over education tier. Many of the scale-level correlations are consistent with a theoretical understanding of the TAPAS scales, suggesting that the scales are measuring the characteristics that they are intended to measure. Classification results provide further evidence that the TAPAS has the potential to enhance matching new Soldiers to MOS, particularly for minimizing attrition.

Utilization and Dissemination of Findings:

The research findings will be used by the U.S. Army Recruiting Command, Army G-1, ASA(M&RA), and Training and Doctrine Command to evaluate the effectiveness of tools used for Army applicant selection and assignment. With each successive set of findings, the TOPS can be revised and refined to meet Army needs and requirements.

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL
TEST AND EVALUATION: 2012 INTERIM REPORT

CONTENTS

TABLES

CONTENTS (CONTINUED)

FIGURES

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL TEST AND EVALUATION: 2012 INTERIM REPORT

CHAPTER 1: INTRODUCTION

Deirdre J. Knapp (HumRRO), Tonia S. Heffner, Leonard A. White, and Kate LaPort (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 personnel research for the Army. The focus of PARU's research is maximizing the potential of the individual Soldier through effective selection, classification, and retention strategies.

In addition to educational, physical, and moral screens, the U.S. Army relies on the Armed Forces Qualification Test (AFQT), a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), to select new Soldiers into the Army. Although the AFQT has proven to be and will continue to serve as a useful metric for selecting new Soldiers, other personal attributes, in particular non-cognitive attributes (e.g., temperament, interests, and values), are important to entry-level Soldier performance and retention (e.g., Knapp & Tremble, 2007).

In December 2006, the Department of Defense (DoD) ASVAB review panel—a panel of experts in the measurement of human characteristics and performance— released recommendations (Drasgow, Embretson, Kyllonen, & Schmitt, 2006), several of which focused on supplementing the ASVAB with additional measures for use in selection and classification decisions. The ASVAB review panel further recommended that the use of these measures be validated against performance criteria.

Just prior to the release of the ASVAB review panel's findings, ARI had initiated a longitudinal research effort, *Validating Future Force Performance Measures (Army Class),* to examine the prediction potential of several non-cognitive measures (e.g., temperament and person-environment fit) for Army outcomes (e.g., performance, attitudes, attrition). The Army Class research project is a 6-year effort that is being conducted with contract support from the Human Resources Research Organization ([HumRRO]; Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009). Experimental predictors were administered to new Soldiers in 2007 and early 2008. Since then, Army Class researchers have obtained attrition data from Army records and collected training criterion data on a subset of the Soldier sample. Job performance criterion data were collected from Soldiers in the Army Class longitudinal validation sample in 2009 and a second round of job performance data collection was completed in April 2011 (Knapp, Owens, & Allen, 2012). Final analysis and reporting of this program of research has been completed (Allen, Knapp, & Owens, 2013.

After the Army Class research was underway, ARI initiated the *Expanded Enlistment Eligibility Metrics (EEEM)* project (Knapp & Heffner, 2010). The EEEM goals were similar to Army Class, but the focus was specifically on Soldier selection and the time horizon was much shorter. Specifically, EEEM required identification of one or more promising new predictor measures for

immediate implementation. The EEEM project capitalized on the existing Army Class data collection procedure and, thus, the EEEM sample was a subset of the Army Class sample.

As a result of the EEEM findings, Army policy-makers approved an initial operational test and evaluation (IOT&E) of the *Tier One Performance Screen (TOPS)*. This report is the fifth in a series presenting continuing analyses from the IOT&E of TOPS.

The Tier One Performance Screen (TOPS)

Six experimental pre-enlistment measures were included in the EEEM research (Allen, Cheng, Putka, Hunter, & White, 2010). These included several temperament measures, a situational judgment test, and two person-environment fit measures based on values and interests. The most promising measures recommended to the Army for implementation were identified based on the following considerations:

- Incremental validity over AFQT for predicting important performance and retention-related outcomes
- Minimal subgroup differences
- Low susceptibility to response distortion (e.g., faking optimal responses)
- Minimal administration time requirements

The Tailored Adaptive Personality Assessment System ([TAPAS]; Stark, Chernyshenko, & Drasgow, 2010) surfaced as the top choice, with the Work Preferences Assessment ([WPA]; Putka & Van Iddekinge, 2007) identified as another good option that was substantively different from the TAPAS. Specifically, the TAPAS is a measure of personality characteristics (e.g., achievement, sociability) that capitalizes on the latest advances in psychometric theory and provides a good indicator of personal motivation. The WPA asks applicants to indicate their preference for various kinds of work activities and environments (e.g., "A job that requires me to teach others," "A job that requires me to work outdoors"). Although not included in the EEEM research, the Information/Communications Technology Literacy (ICTL) test emerged as a potential test of applicants' familiarity with computers and information technology, which may predict performance in high-technology occupations (Russell & Sellman, 2009).

In May 2009, the Military Entrance Processing Command (MEPCOM) began administering TAPAS on the computer adaptive platform for the ASVAB (CAT-ASVAB). Initially, TAPAS was to be administered only to Education Tier 1, non-prior service applicants.[1] This limitation to Education Tier 1 was removed early in CY2011 so the Army could evaluate TAPAS across all types of applicants.

TOPS uses non-cognitive measures to identify applicants who would likely perform differently (higher or lower) than would be predicted by their ASVAB scores. As part of the TOPS IOT&E, TAPAS scores are being used to screen out a small number of AFQT Category IIIB/ IV applicants.[2]

---

[1] Applicant educational credentials are classified as Tier 1 (primarily high school diploma graduates), Tier 2 (primarily non-diploma graduate), and Tier 3 (not a high school graduate).
[2] 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).

Although the WPA is part of the TOPS IOT&E, WPA scores will not be considered for enlistment eligibility. The WPA is scheduled for administration at MEPS starting in late CY2013.

Although the initial conceptualization for the IOT&E was to use the TAPAS as a tool for "screening in" Education Tier 1 applicants with lower AFQT scores, changing economic conditions spurred a reconceptualization that led to using the TAPAS as a tool that screens out low motivated applicants. Recruiting conditions continue to shift, so both the IOT&E and any subsequent fully operational system will need to adjust to fit with the applicant market. TAPAS composite scores and cut points can be modified as needed to fit recruiting market conditions.

Evaluating TOPS

Figure 1.1 illustrates the TOPS IOT&E research plan. To evaluate the pre-enlistment measures (TAPAS, WPA, and ICTL), the Army is collecting training criterion data on Soldiers in eight target military occupational specialties (MOS) as they complete initial military training (IMT).[3] The criterion measures include job knowledge tests (JKTs); an attitudinal assessment, the Army Life Questionnaire (ALQ); and performance rating scales (PRS) completed by the Soldiers' cadre. These measures are computer-administered at the schools for each of the eight target MOS. The process is overseen by Army personnel with guidance and support from both ARI and HumRRO. Course grades and completion rates are obtained from administrative records for all Soldiers who take the TAPAS, regardless of MOS.

| Predictors | Criteria | | | Evaluation Cycles |
	IMT	In-Unit	Separation	
TEST ASVAB TAPAS WPA ICTL	JKT ALQ PRS Grades Course Completion	JKT ALQ PRS	Attrition Reenlistment	Bi-Annually
SAMPLE All non-prior service applicants at MEPS	11B, 19K, 25U, 31B, 42A, 68W, 88M, 91B	All Soldiers	All Soldiers	
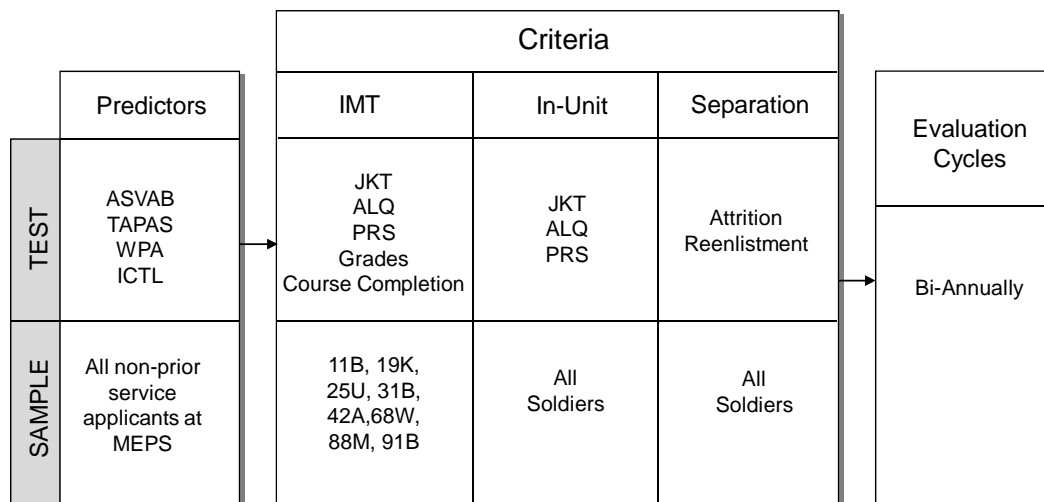
Figure 1.1. TOPS Initial Operational Test & Evaluation (IOT&E).

Criterion data are also being collected from Soldiers and their supervisors during data collection trips to major Army installations. These proctored "in-unit" data collections began in January 2011 and target Soldiers who took the TAPAS prior to enlistment. The in-unit criterion measures

---

[3] The target MOS are Infantryman (11B), Armor Crewman (19K), Signal Support Specialist (25U), Military Police (31B), Human Resources Specialist (42A), Health Care Specialist (68W), Motor Transport Operator (88M), and Light Wheel Vehicle Mechanic (91B). These MOS were selected to include large, highly critical MOS as well as to represent the diversity of work requirements across MOS.

include JKTs, the ALQ, and supervisor ratings of performance. The data collection model closely mirrors that which was used in the Army Class research program (Knapp et al., 2012). Separation status of all Soldiers who took the TAPAS prior to enlistment is tracked throughout the course of the research.

This report describes the fifth iteration of developing a criterion-related validation data file and conducting evaluation analyses using data collected in the TOPS IOT&E initiative. Prior evaluations are described in a series of technical reports (Knapp & Heffner, 2011, 2012; Knapp, Heffner, & White, 2011; Knapp & LaPort, 2013). Additional analysis datasets and validation analyses will be prepared and conducted at 6-month intervals throughout the multi-year IOT&E period.

Overview of Report

Chapter 2 explains how the evaluation analysis data files are constructed and then describes characteristics of the samples resulting from construction of the latest analysis data file. Chapter 3 describes the TAPAS and ASVAB, including content, scoring, and psychometric characteristics. Chapter 4 describes the IMT and in-unit criterion measures included in this evaluation, including their psychometric characteristics. Criterion-related validation analyses are presented in Chapter 5. Chapter 6 examines classification potential of the TAPAS. The report concludes with Chapter 7, which summarizes our continuing efforts to evaluate TOPS and looks toward plans for future iterations of these evaluations.

CHAPTER 2: DATA FILE DEVELOPMENT

D. Matthew Trippe, Laura Ford, Bethany Bynum, and Karen Moriarty (HumRRO)

Overview of Process

The TOPS data file comprises predictor and criterion data obtained from administrative, IMT, and in-unit sources. The IMT and in-unit assessments are described in Chapter 4.

An illustrative view of the TOPS analysis file construction process is provided in Figure 2.1.[4] The lighter boxes within the figure represent source data files, and the darker boxes represent samples on which descriptive or inferential analyses are conducted. Samples are formed by applying filters to a data file such that it includes the observations of interest. The leftmost column in the figure summarizes the predictor data sources used to derive the TOPS Applicant Sample. The other columns summarize the research-only (i.e., non-administrative) and administrative criterion data. Predictor and criterion data are merged to form the IMT or in-unit
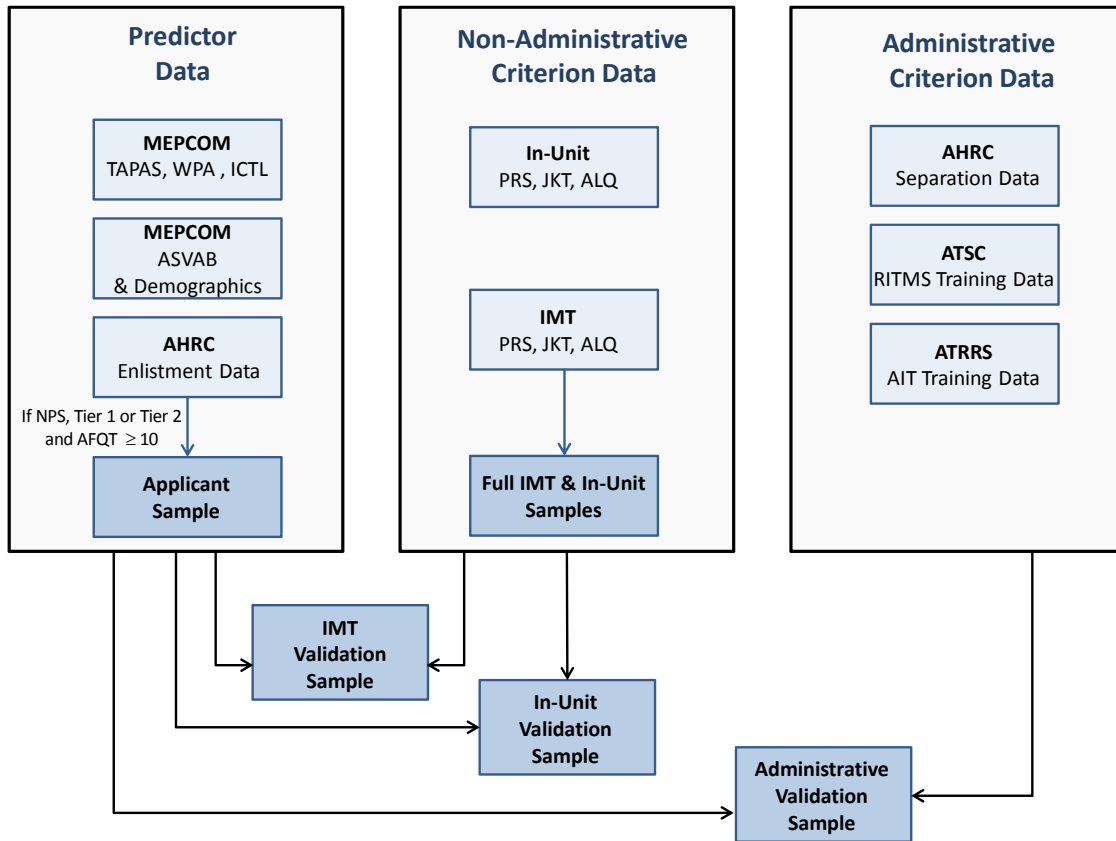
Figure 2.1. Overview of TOPS data file merging and nested sample generation process.

---

[4] Administrative data are collected from the following sources: (a) Military Entrance Processing Command (MEPCOM), (b) Army Human Resources Command (AHRC), (c) Army Training Support Center's (ATSC) Resident Individual Training Management System (RITMS), and (d) Training and Doctrine Command's (TRADOC) Army Training Requirements and Resources System (ATRRS).

validation samples and the large administrative validation sample, which includes all Soldiers who have predictor data and at least one criterion record (e.g., administrative data). The latest version of the TOPS data file does not contain WPA predictor scores or a sufficient number of ICTL scores to warrant inclusion in the analysis data set at this time. Future versions of the data file will include those data.

Description of Data File and Sample Construction

The latest data file created in June 2012 includes TAPAS data collected from May 2009 through March 2012 and criterion data collected through May 2012. Table 2.1 summarizes the relevant characteristics of the total TAPAS sample contained in the June 2012 TOPS data file. The total sample includes applicants who did not enlist in the Army. The TOPS Applicant Sample was defined by limiting records in the total sample data file provided by MEPCOM to those Soldiers who are non-prior service, Education Tier 1 or 2[5], and have an AFQT score of 10 or greater. Among the 282,563 applicants in the total, unfiltered sample, 250,884 (88.8%) met these screens and constituted the Applicant Sample.

Sample sizes reported in all subsequent chapters and appendices will generally be smaller than the initial numbers reported here because of further data filtering or disaggregation that occurs for each particular analysis. Predictor and criterion scores were determined to be valid if they passed multiple data quality screens intended to identify unmotivated applicants. Those additional screens have not yet been applied to the samples described in this chapter because they are often specific to a particular analysis. Further, new versions of the TAPAS were introduced in August 2011. However, there were insufficient numbers of post-August 2011 applicants with criterion data to include in the evaluation at this time. Accordingly, the predictive validity and person-job matching analyses reported in Chapters 5 and 6, respectively, were further limited to applicants who completed the pre-August 2011 versions of the TAPAS.

A detailed breakout of background and demographic characteristics observed in the analytic samples appears in Table 2.2. Regular Army Soldiers comprise a majority of the cases in each sample. The samples are predominantly male, Caucasian, and non-Hispanic; however, a large percentage of Soldiers declined to provide information on race or ethnicity.

The Administrative Validation Sample described in Table 2.2 includes 113,618 Soldiers. Included in this sample are Soldiers who meet all of the inclusion criteria for the TOPS Applicant Sample and also have at least one record in an administrative criterion data source (i.e., Army Training Requirements and Resources System [ATRRS], Resident Individual Training Management System [RITMS], attrition). However, the number of Soldiers included in any individual analysis is generally much smaller. The exact number of Soldiers varies by criterion depending on the availability of valid data on key variables. Specific sample details on each criterion variable are provided in subsequent chapters.

---

[5] Starting with the June 2012 TOPS data file, we incorporated education tier information from a AHRC data source to best capture a Soldier's education tier status at the time of his or her accession. As a result, figures for education tier reported in the current report will differ from corresponding figures in previous reports. The differences were generally minor and did not impact the overall results or findings.

Although there are 44,987 Soldiers in the Full IMT data file, only 13,238 have TAPAS scores. There are two primary reasons for this disconnect. First, early in the research effort most of the Soldiers tested at the schools had taken their pre-enlistment tests before MEPCOM started administering the TAPAS widely to applicants. Second, we rely on name and date of birth to match TAPAS records to the criterion data, which often results in unsuccessful matches. Indeed, the overall match rate at this stage (29.4%) is an improvement over the match rates obtained previously (26.8%, 19.8%, 12.7%, 5.5%; Trippe, Ford, Bynum, & Moriarty, 2012). The match rate for newly collected criterion data was 55%. Similarly, there are 2,529 Soldiers with in-unit data, 604 of whom had matching TAPAS data. There are 113 Soldiers with a TAPAS record and both IMT and in-unit criterion data.

Table 2.1. Full TAPAS Data File Characteristics

Variables	*n*	% of Total Sample *(N = 282,563)*
Education Tier		
Tier 1	263,204	93.2
Tier 2	14,341	5.1
Tier 3	5,013	1.8
Unknown	5	.0
Prior Service		
Yes	6,796	2.4
No or Missing	275,767	97.6
Military Occupational Specialty		
11B/11C/11X/18X	23,058	8.2
19K	1,236	.4
25U	2,297	.8
31B	5,691	2.0
42A	2,979	1.1
68W	6,902	2.4
88M	6,731	2.4
91B	6,158	2.2
Other	78,617	27.8
Unknown [a]	148,894	52.7
AFQT Category		
I	19,035	6.7
II	82,450	29.2
IIIA	54,792	19.4
IIIB [b]	82,188	29.1
IV [b]	39,627	14.0
V	4,461	1.6
Unknown	10	.0
Contract Status		
Signed	168,647	59.7
Not signed	113,916	40.3
Applicant Sample [c]	250,884	88.8

[a] Generally, when the MOS is unknown, it is either because the respondent did not access into the Army or because the information was not yet available in the data sources on which the June 2012 data file was based.

[b] AFQT Category IIIB and IV applicants are oversampled. Figures presented are not representative of Army accessions.

[c] The Applicant Sample size is smaller than the total TAPAS sample because it is limited to non-prior service, Education Tier 1 and 2, and AFQT ≥ 10 applicants with valid TAPAS scores.

Table 2.2. Background and Demographic Characteristics of the TOPS Samples

Characteristic	Applicant [a] *n* = 250,884 *n*	*%*	Administrative Validation [b] *n* = 113,618 *n*	*%*	IMT Validation [c] *n* = 13,238 *n*	*%*	In-Unit Validation [d] *n* = 604 *n*	*%*
Component								
Regular	148,574	59.2	65,974	58.1	7,795	58.9	594	98.3
ARNG	69,172	27.6	32,029	28.2	4,128	31.2	--	--
USAR	33,138	13.2	15,615	13.7	1,315	9.9	--	--
Unknown	--	--	--	--	--	--	10	.2
Education Tier								
Tier 1	237,949	94.8	109,487	96.4	12,824	96.9	599	99.2
Tier 2	12,935	5.2	4,131	3.6	414	3.1	5	.8
Military Occupational Specialty								
11B/11C/11X/18X	22,255	8.3	20,115	17.7	5,648	42.6	137	22.7
19K	1,144	.5	1,094	1.0	154	1.2	10	1.7
25U	2,128	.9	1,868	1.6	11	.1	10	1.7
31B	5,113	2.0	4,555	4.0	2,439	18.4	10	1.7
42A	2,661	1.0	2,366	2.1	209	1.6	11	1.8
68W	6,357	2.5	5,951	5.2	2,323	17.5	22	3.6
88M	6,103	2.4	5,145	4.5	2,019	15.3	27	4.5
91B	5,538	2.2	4,890	4.3	434	3.3	29	4.8
Other	71,278	28.4	67,483	59.4	--	--	348	57.6
Unknown	129,784	51.7	151	.1	--	--	--	--
AFQT Category								
I	16,942	6.8	8,767	7.7	1,031	7.8	48	7.9
II	74,450	29.7	38,838	34.2	5,129	38.7	190	31.5
IIIA	49,577	19.8	25,012	22.0	2,884	21.8	133	22.0
IIIB	74,262	29.6	35,293	31.1	3,624	27.4	213	35.3
IV	35,653	14.2	5,708	5.0	570	4.3	20	3.3
Gender								
Female	49,726	19.8	19,258	16.9	1,710	12.9	75	12.4
Male	199,067	79.4	93,427	82.2	11,460	86.6	529	87.6
Unknown	2,091	.8	933	.8	68	.5	--	--
Race								
African American	45,457	18.1	18,429	16.2	1,484	11.2	85	14.1
American Indian	1,850	.7	794	.7	107	.8	2	.3
Asian	7,990	3.2	3,806	3.3	366	2.8	18	3.0
Hawaiian/Pacific Islander	1,378	.5	717	.6	91	.7	5	.8
Caucasian	180,601	72.0	85,287	75.1	10,520	79.5	464	76.8
Multiple	1,034	.4	508	.4	61	.5	--	--
Unknown	12,574	5.0	4,077	3.6	609	4.6	30	5.0
Ethnicity								
Hispanic/Latino	37,301	14.9	16,349	14.4	1,553	11.7	73	12.1
Not Hispanic	201,511	80.3	93,793	82.6	11,160	84.3	507	83.9
Unknown	12,054	4.8	3,476	3.1	525	4.0	24	4.0

[a] Limited to applicants who had no prior service, Education Tier 1 or 2, and AFQT ≥ 10; served as the core analysis sample.
[b] Soldiers in Applicant Sample with administrative criterion data (i.e., ATTRS, RITMS, or attrition).
[c] Soldiers in Applicant Sample with IMT criterion data collected at the schoolhouses.
[d] Soldiers in Applicant Sample with In-Unit criterion data.

Summary

The TOPS data file is periodically updated by merging new TAPAS scores, administrative records, IMT, and in-unit data into one master data file. The June 2012 data file includes a total of 282,563 applicants who took the TAPAS. Of these, 250,884 were in the TOPS Applicant Sample. The Applicant Sample was determined by excluding Education Tier 3, AFQT Category V, and prior service applicants from the master data file. Of that Applicant Sample, 113,618 (45.3%) had a record in at least one of the administrative criterion data sources; 13,238 had IMT data collected from the schoolhouse and 604 had in-unit criterion data. The IMT match rate represents an improvement over prior reporting cycles. This is likely due to the maturation of criterion data in the source data files. For example, in earlier reporting cycles we likely collected criterion data on a number of Soldiers who were once waiting in the Delayed Enlistment Program (DEP). These Soldiers were recruited prior to when the TAPAS was widely administered and thus did not have a TAPAS score to merge. Nevertheless, the amount of criterion data that is actually used in a given analysis remains small in relation to the amount of available predictor data. Subsequent iterations of the TOPS IOT&E data file will no doubt show progressively stronger sample sizes to support validation and other evaluative analyses.

CHAPTER 3: DESCRIPTION OF THE TOPS IOT&E PREDICTOR MEASURES

Stephen Stark, O. Sasha Chernyshenko, Fritz Drasgow (Drasgow Consulting Group), Matthew T. Allen, and Deirdre J. Knapp (HumRRO)

The purpose of this chapter is to describe the predictor measures investigated to date in the TOPS IOT&E (i.e., TAPAS and ASVAB). The central predictor under investigation in this analysis is the TAPAS (Drasgow, Stark, Chernyshenko, Nye, Hulin, & White, 2012; Stark et al., 2010), while the baseline predictor used by the Army is the ASVAB. Two additional experimental measures, the ICTL (Russell & Sellman, 2009) and Work Preferences Assessment (WPA; Putka & Van Iddekinge, 2005), are not yet included in the analysis data files, and are therefore not discussed further here.

Tailored Adaptive Personality Assessment System (TAPAS)

Description

The TAPAS is a personality measurement tool originally developed by Drasgow Consulting Group (DCG) under the Army's Small Business Innovation Research (SBIR) program. The system builds on the foundational work of the Assessment of Individual Motivation ([AIM]; White & Young, 1998) by incorporating features designed to promote resistance to faking and by measuring narrow personality constructs (i.e., facets) that are known to predict outcomes in work settings. Because the TAPAS uses item response theory (IRT) methods to construct and score items, it can be administered in multiple formats: (a) as a fixed length, *non-adaptive test* where examinees respond to the same sequence of items or (b) as an *adaptive* test where each examinee responds to a unique sequence of items selected to maximize measurement accuracy for that specific examinee.

The TAPAS uses an IRT model for multidimensional pairwise preference items ([MUPP]; Stark, Chernyshenko, & Drasgow, 2005) as the basis for constructing, administering, and scoring personality tests that are designed to reduce response distortion (i.e., faking) and yield normative scores even with tests of high dimensionality (Stark, Chernyshenko, & Drasgow 2012). TAPAS items consist of pairs of personality statements for which a respondent's task is to choose the one that is "more like me." The two statements constituting each item are matched in terms of social desirability and often represent different dimensions. As a result, it is difficult for respondents to discern which answers improve their chances of being enlistment eligible. Because they are less likely to know which dimensions are being used for selection, they are less likely to identify which statements measure those dimensions and they are less likely to be able to keep track of their answers on several dimensions simultaneously so as to provide consistent patterns of responses across the whole test. Without knowing which answers have an impact on their eligibility status, respondents should not be able to increase their scores on selection dimensions as easily as when traditional, single statement measures are used. In short, the TAPAS' features make it difficult for applicants to distort their responses to obtain more desirable scores.

The use of a formal IRT model also greatly increases the flexibility of the assessment process. A variety of test versions can be constructed to measure personality dimensions that are relevant to

specific work contexts, and the measures can be administered via paper-and-pencil or computerized formats. If test content specifications (i.e., test blueprints) are comparable across versions, the respective scores can be readily compared because the metric of the statement parameters has already been established by calibrating response data obtained from a base or reference group (e.g., Army recruits). The same principle applies to adaptive testing, wherein each examinee receives a different set of items chosen specifically to reduce the error in his or her trait scores at points throughout the exam. Adaptive item selection enhances test security because there is less overlap across examinees in terms of the items presented.

Another important feature of the TAPAS is that pools of statements representing more than 20 narrow personality traits are available. The initial TAPAS trait taxonomy was developed using the results of several large scale factor-analytic studies with the goal of identifying a comprehensive set of non-redundant narrow traits. Since the TAPAS was initially developed, additional traits have been added. These narrow traits, if necessary or desired, can be combined to form either the Big Five (the most common organization scheme for narrow personality traits) or any other number of broader traits (e.g., Integrity or Positive Core Self-Evaluations). This is advantageous for applied purposes because TAPAS versions can be created to fit a wide range of applications and are not limited to a particular service branch or criterion. Selection of specific TAPAS dimensions can be guided by consulting the results of a meta-analytic study performed by DCG that mapped 22 TAPAS dimensions to several important organizational criteria for military and civilian jobs (e.g., task proficiency, training performance, attrition) (Chernyshenko & Stark, 2007), as well as subsequent validation research.

Scoring details and the criterion-related validation work that led to the inclusion of the TAPAS in the TOPS IOT&E can be found in the *Expanded Enlistment Eligibility Metrics* report (Knapp & Heffner, 2010) and in earlier evaluation reports in this series (Knapp et al., 2011; Knapp & Heffner, 2011)

Multiple Versions of TAPAS

As part of the TOPS IOT&E, multiple versions of the TAPAS have been administered as ARI explores the value of new and alternative dimensions. (See Table 3.1 shows for a list of dimension names and descriptions.) In the present report, analyses reported in Chapters 5 and 6 are based on two 15-dimension versions of TAPAS, each administering 120 items (i.e., pairs of statements). One version was nonadaptive (static), so all examinees answered the same sequence of items; the other was adaptive, so each examinee answered items tailored to his or her trait level estimates. The TAPAS-15D-Static was administered from mid-July to mid-September of 2009 to all examinees, and later to smaller numbers of examinees at some MEPS. The adaptive version, referred to as TAPAS-15D-CAT, was introduced in September 2009.

In August 2011, three new adaptive versions of the TAPAS were introduced into the MEPS (15D-CAT Version 2, Forms A, B, and C) to replace the original versions. Each version measures 15 dimensions, though not always the same dimensions. Specifically, all three versions assess the same nine core dimensions, to include all of the scales in the TOPS first operational "can-do" and "will-do" composites (described next), plus six of 12 experimental dimensions. The six experimental dimensions assessed vary by version. In total, the newer versions of the

TAPAS collectively measure 21 dimensions. Note also that these versions of TAPAS use statement pools that were created exclusively for ARI. The present report provides psychometric information on these versions of the TAPAS, but they are not included in the selection and classification analyses because associated criterion data are limited.

Table 3.1. TAPAS Dimensions Names and Definitions

Facet Name	Brief Description
Achievement	High scoring individuals are seen as hard working, ambitious, confident, and resourceful.
Adjustment	High scoring individuals are well adjusted, worry free, and handle stress well.
Adventure Seeking	High scoring individuals enjoy participating in extreme sports and outdoor activities.
Aesthetics	High scoring individuals appreciate various forms of art and music and participate in art-related activities more than most people.
Attention Seeking	High scoring individuals tend to engage in behaviors that attract social attention. They are loud, loquacious, entertaining, and even boastful.
Commitment to Serve	High scoring individuals identify with the military and have a strong desire to serve their country.
Consideration	High scoring individuals are affectionate, compassionate, sensitive, and caring.
Cooperation	High scoring individuals are pleasant, trusting, cordial, non-critical, and easy to get along with.
Courage	High scoring individuals stand up to challenges and are not afraid to face dangerous situations.
Curiosity	High scoring individuals are inquisitive and perceptive; they are interested in learning new information and attend courses and workshops whenever they can.
Dominance	High scoring individuals are domineering, "take charge" and are often referred to by their peers as "natural leaders."
Even Tempered	High scoring individuals tend to be calm and stable. They don't often exhibit anger, hostility, or aggression.
Ingenuity	High scoring individuals are inventive and can think "outside of the box."
Intellectual Efficiency	High scoring individuals believe they process information and make decisions quickly; they see themselves (and they may be perceived by others) as knowledgeable, astute, or intellectual.
Non-Delinquency	High scoring individuals tend to comply with rules, customs, norms, and expectations, and they tend not to challenge authority.
Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.
Order	High scoring individuals tend to organize tasks and activities and desire to maintain neat and clean surroundings.

Table 3.1. (Continued)

Facet Name	Brief Description
Physical Conditioning	High scoring individuals tend to engage in activities to maintain their physical fitness and are more likely participate in vigorous sports or exercise.
Responsibility	High scoring individuals are dependable, reliable, and make every effort to keep their promises.
Self Control	High scoring individuals tend to be cautious, levelheaded, able to delay gratification, and patient.
Selflessness	High scoring individuals are generous with their time and resources.
Situational Awareness	High scoring individuals pay attention to their surroundings and rarely get lost or surprised.
Sociability	High scoring individuals tend to seek out and initiate social interactions.
Team Orientation	High scoring individuals prefer working in teams and make people work together better.
Virtue	High scoring individuals strive to adhere to standards of honesty, morality, and "good Samaritan" behavior.

As described further in Chapter 7, these versions of the TAPAS will soon be replaced as well.

Scores have been standardized within TAPAS versions to smooth potential scaling differences and to enable cross-version analyses. Descriptive statistics and intercorrelations of individual TAPAS scales and composite scores are provided in Appendix A. Information is provided for each TAPAS version, by education tier.

TAPAS Composites

An initial Education Tier 1 performance screen was developed from the TAPAS-95s scales for the purpose of testing in an applicant setting (Allen et al., 2010).[6] This was accomplished by (a) identifying key criteria of most interest to the Army, (b) sorting these criteria into "can-do" and "will-do" categories (see below), and (c) selecting composite scales corresponding to the can-do and will-do criteria, taking into account both theoretical rationale and empirical results. The result of this process was two composite scores.

1. Can-Do Composite: The original TOPS Operational Can-Do composite consists of five TAPAS scales and is designed to predict the extent to which Soldiers can perform the technical aspects of their jobs, using indicators such as MOS-specific job knowledge, Advanced Individual Training (AIT) exam grades, and graduation from AIT/One Station Unit Training (OSUT).

2. Will-Do Composite: The original TOPS Operational Will-Do composite consists of five TAPAS scales (three of which overlap with the Can-Do composite) and is

---

[6] TAPAS-95s was a paper-and-pencil, static version of the TAPAS used in the Army Class research.

designed to predict the more motivational elements of job performance, such as maintaining physical fitness, adjusting to Army life, demonstrating effort, and supporting peers.

As more data have become available and the dimensions included in the different TAPAS versions could be evaluated, additional work has been done to create and evaluate new TAPAS composites. As a result of this work, the Army has approved the use of three new composites to screen applicants. In addition to reconfigured Can-Do and Will-Do composites, there is a "Adaptation" composite designed to predict attrition. These new composites will be used starting in FY2014, along with the introduction of three new versions of TAPAS. More information about how the new composites were developed is provided in a limited distribution addendum to this report. Those interested in obtaining a copy of this addendum should contact the editors for further information.

Armed Services Vocational Aptitude Battery (ASVAB)
Content, Structure, and Scoring

The ASVAB is a multiple aptitude battery of tests administered by the MEPCOM. Most military applicants take the computer adaptive version of ASVAB (i.e., the CAT-ASVAB). Scores on the ASVAB tests are combined to create composite scores for use in (a) selecting applicants into the Army and (b) classifying them into an MOS. The AFQT, the composite used for selecting applicants into the Army, comprises the Verbal Expression[7] (VE), Arithmetic Reasoning (AR), and Math Knowledge (MK) tests (AFQT = 2*VE + AR + MK). Applicants must meet a minimum AFQT score to be eligible to serve in the military, and the Services favor high-scoring applicants for enlistment. AFQT percentile scores are divided into the following categories:[8]

- Category I (93-99)
- Category II (65-92)
- Category IIIA (50-64)
- Category IIIB (31-49)
- Category IV (10-30)
- Category V (1-9)

AFQT Category V Soldiers are not eligible for enlistment, Category IV accessions are greatly restricted, and priority is given to Category I-IIIA accessions.

For classification, scores on the ASVAB tests are combined to form 10 Aptitude Area (AA) composites. An applicant must receive a minimum score on the MOS-relevant AA composite(s) to qualify for classification to that MOS. For example, applicants must score a 95 in both the Electronics (EL) and Signal Communications (SC) AA composites to qualify as a Signal Support Specialist (25U). Descriptive statistics for the AFQT, ASVAB tests, and AA composites are reported in Appendix A. AFQT Category frequencies are reported in Chapter 2 (Tables 2.1 and 2.2).

---

[7] Verbal Expression is a scaled combination of the Word Knowledge (WK) and Paragraph Comprehension (PC) tests.
[8] For more information on ASVAB scoring, see the official website of the ASVAB, www.officialasvab.com.

Summary

The purpose of this chapter was to describe the predictor measures used as part of the TOPS IOT&E. The TAPAS is unique among personality measures because it uses forced-choice pairwise items and IRT to promote resistance to faking. Initial validation research conducted as part of EEEM was promising enough to warrant an IOT&E of TAPAS. The ASVAB, which consists of multiple tests that are formed into operational selection (i.e., AFQT) and classification (i.e., AA) composites, is used as the baseline instrument for incremental validity analyses reported in Chapter 5.

CHAPTER 4: DESCRIPTION AND PSYCHOMETRIC PROPERTIES OF CRITERION MEASURES

Taylor Sparks and Chad Peddie (HumRRO)

Criterion scores to validate the TAPAS were derived from measures administered for purposes of this research and from administrative records. The research measures included job knowledge tests (JKTs), performance rating scales (PRS), and a questionnaire measuring self-reported attitudes and performance (Army Life Questionnaire [ALQ]). The original versions of these three measures were developed for the Army Class project (Moriarty, Campbell, Heffner, & Knapp, 2009), and modified, as needed, for inclusion in the TOPS IOT&E. Criterion scores drawn from Soldiers' administrative records included separation status (i.e., attrition), IMT completion, and IMT grades.

In this chapter, we first describe the criterion measures in more detail, including revisions that were made during the course of the data collection. The chapter concludes with a discussion of the distributional and psychometric characteristics of the various measures.

Table 4.1. Summary of IMT and In-Unit Criterion Measures

Criterion Measure	Description
Soldier/Cadre Reported	
Job Knowledge Tests (JKT)	The Warrior Tasks and Battle Drills (WTBD) JKT measures knowledge that is general to all enlisted Soldiers. MOS-specific JKTs measure Soldiers' knowledge of basic facts, principles, and procedures required of Soldiers in training for a particular MOS. Each JKT includes a mix of item formats (e.g., multiple-choice, multiple-response, and rank order).
Performance Rating Scales (PRS)	The IMT PRS measure Soldiers' performance in two domains: (a) MOS-specific (e.g., learns preventive maintenance checks and services, learns to troubleshoot vehicle and equipment problems) and (b) Army-wide (e.g., exhibits effort, supports peers, demonstrates physical fitness). The IMT PRS are completed by drill sergeants or training cadre. In-unit PRS cover Army-wide dimensions only and are completed by supervisors.
Army Life Questionnaire (ALQ)	The ALQ measures Soldiers' self-reported attitudes and experiences in the Army. The IMT and in-unit versions are very similar.
Administrative	
Attrition	Separation data were obtained on participating Regular Army Soldiers at 3 months (attrition near or after the completion of Basic Combat Training), 4 months (attrition during AIT (Advanced Individual Training) /OSUT (One Site Unit Training)), 6 months (attrition near or after completion of AIT/OSUT), and at regular 3-month intervals thereafter. Attrition data through 30 months were available for the current sample.
Initial Military Training (IMT) Criteria	These data provide information about whether Soldiers restarted IMT and for what reasons, the number of times Soldiers restarted training, graduation status, and school grades for Soldiers in AIT.

Criterion Measure Descriptions

Job Knowledge Tests (JKTs)

Multiple sets of JKTs (IMT and in-unit) were adapted from the Select21 (Collins, Le, & Schantz, 2005) and Army Class (Moriarty et al., 2009) projects: one for Warrior Tasks and Battle Drills (WTBD), which is administered to all participating Soldiers, and a set of MOS-specific JKTs for Infantry, Armor, Military Police, Health Care Specialist, Light Wheel Vehicle Mechanic, and Motor Transport Operator Soldiers. MOS-specific JKTs for two additional MOS (Signal Support Specialist and Human Resources Specialist) were developed in the Fall of 2011.

Most of the JKT items are 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 are also used. The items use visual images to make them more realistic and reduce reading requirements for the test.

Prior to finalizing the items for initial use in the TOPS IOT&E, the items were reviewed by project staff and Army SMEs to ensure they were of high quality. The JKTs were reviewed again in the summer of 2011. Poorly performing or outdated items were replaced and new items were added to ensure adequate coverage of content areas identified in the test blueprints. Depending on the MOS, administration of the revised JKTs started in Fall 2011 or Spring 2012. Preliminary analyses were conducted comparing the revised versions of the JKTs to the initial versions, based on data collected through May 2012. Overall, the psychometric properties of the initial and revised versions of the JKTs were comparable. Accordingly, data on the initial and revised versions were combined for analysis purposes.

Performance Rating Scales (PRS)

The PRS, like the JKTs, also were adapted from or based on previous research (see Moriarty et al., 2009 for details). The IMT and in-unit PRS are fairly different, so they will be described separately.

IMT PRS

The IMT PRS target two domains of Soldier performance requirements: (a) Army-wide and (b) MOS-specific. The IMT PRS were completed by cadre members (supervisors/trainers) of participating Soldiers. The initial IMT Army-wide PRS consisted of the eight scales listed in Figure 4.1. As summarized in Figure 4.1, ratings on several of the individual scales were combined into PRS composites, resulting in a total of five Army-wide PRS scores, in addition to the Overall Performance rating. An Overall Army-Wide composite score also was computed by averaging the ratings on the eight individual scales. For the MOS-specific PRS, the number of scales ranged from five to nine. Ratings on the individual MOS-specific PRS were combined to form a single overall score for each MOS.[9] The initial IMT PRS, except for the Overall Performance scale, employed a behaviorally-anchored rating scale (BARS) format. Each scale

---

[9] MOS-specific ratings were combined into a single overall MOS performance score because (a) dimension ratings were highly correlated with each other and (b) inter-reliability estimates obtained in prior research for these scales were low.

ranged from 1 (lowest) to 7 (highest) and included descriptions (or "anchors") of high, medium, and low performance. The scales also included a "not observed" option for those instances where the cadre did not have an opportunity to observe a Soldier's performance on a specific dimension. Figure 4.2 provides an example of one of the BARS on the original IMT PRS.
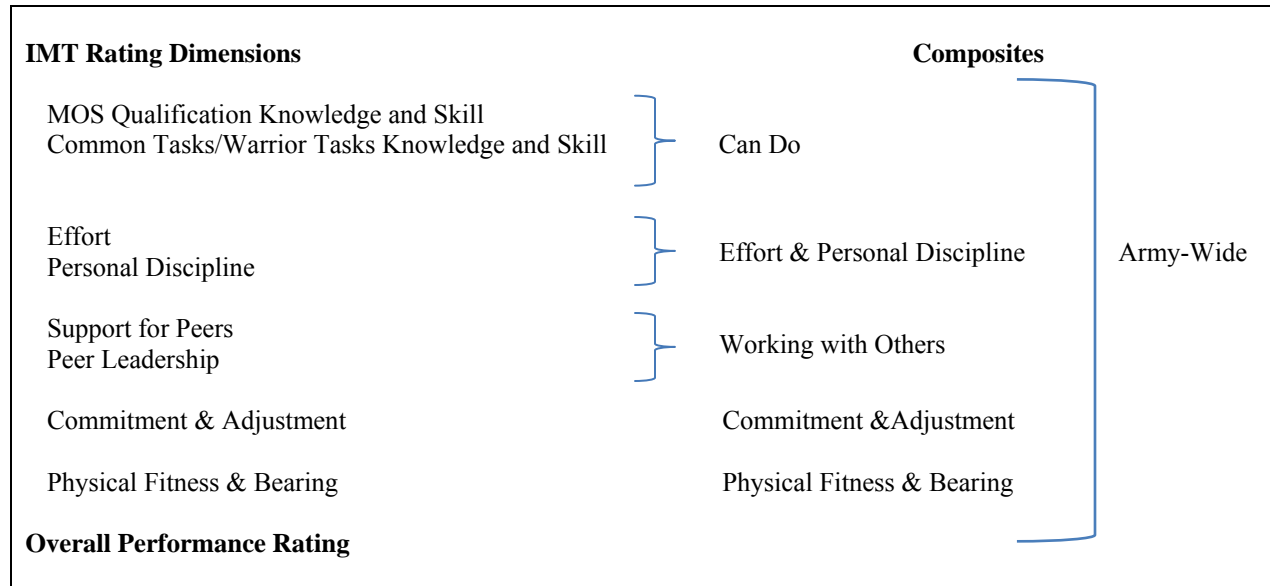
Figure 4.1. IMT Army-Wide Performance Rating Scale dimensions and composite score composition (original scales).

Figure 4.2. Sample IMT 7-point behaviorally-anchored rating scale (original format).

Overall Performance ratings were made on a 5-point relative comparison scale, as shown in Figure 4.3. The initial PRS assessment also included a 3-point "familiarity" scale in which the cadre rated his or her general opportunity to observe each Soldier being rated (i.e., limited, reasonable, or a lot of opportunity to observe).

Overall Performance				
Considering your evaluation of the Soldier on the dimensions important to successful performance, please rate the overall effectiveness of each Soldier compared to his/her peers.				
1	2	3	4	5
Among the Weakest	**Below Average**	**Average**	**Above Average**	**Among the Best**
(in the bottom 20% of Soldiers)	(in the bottom 40% of Soldiers)	(better than the bottom 40% of Soldiers, but not as good as the top 40%)	(in the top 40% of Soldiers)	(in the top 20% of Soldiers)

Figure 4.3. IMT relative overall performance rating scale.

Prior IOT&E evaluations noted low interrater reliability estimates for the PRS. Accordingly, several changes were made to the IMT PRS in an attempt to improve their psychometric characteristics in Fall 2011.[10] First, the BARS format for both the Army-wide and MOS-specific PRS was replaced with a 5-point relative scale format, comparable to the Overall Performance scale shown in Figure 4.3. Second, the number of Army-wide scales was reduced from eight to five, paralleling the five scores generated from the original scales. Instead of including a single "Can Do" scale, the Common Tasks Knowledge and Skill scale (which is highly correlated with MOS knowledge and skill) was dropped. Finally, the "Adjustment" scale was revised for clarity. The configuration of the revised Army-wide scales is shown in Figure 4.4. Finally, the familiarity scale was changed to a 4-point scale enabling raters to more clearly indicate their ability to judge each Soldier's performance.

There were insufficient numbers of Soldiers in the IMT sample rated on the new PRS to conclusively determine if these changes improved the psychometric quality of the ratings data. That question can be examined more closely in the next evaluation cycle.

In-Unit PRS

The in-unit PRS only target Army-wide dimensions of performance (i.e., there are no MOS-specific in-unit PRS). The in-unit PRSs have consistently employed the 7-point BARS format used for the initial IMT scales (see Figure 4.2). The initial in-unit Army-wide PRS measured 12 performance dimensions, plus a Leadership Potential scale (see Figure 4.5). A thirteenth scale was dropped in Fall 2011 because of poor psychometric properties and has since been replaced with an Adjustment scale, comparable to the corresponding IMT scale. The revised 4-point "familiarity" scale used in the new IMT PRS also is used with the in-unit PRS.

---

[10] Interrater reliability was assessed using $G(q,k)$, a reliability metric designed specifically for studies like TOPS where the measurement design is ill-structured (Putka, Le, McCloy, & Diaz, 2008).

IMT Rating Dimensions

    MOS Qualification Knowledge and Skill

    Effort & Personal Discipline

    Working with Others   **Overall Performance Composite**

    Adjustment to the Army

    Physical Fitness and Bearing

Overall Performance Rating
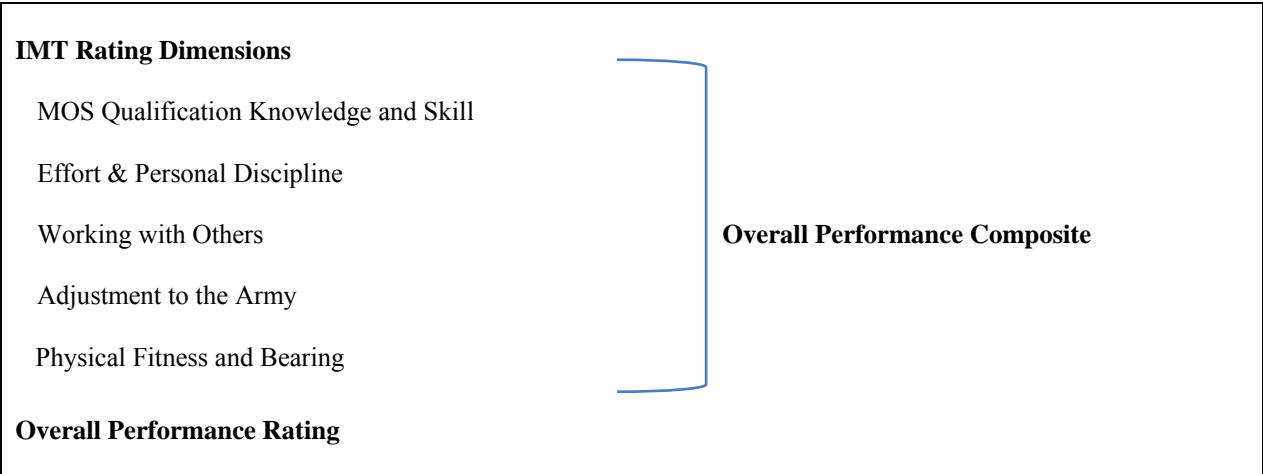
Figure 4.4. IMT Army-Wide Performance Rating Scale dimensions and composite score composition (revised scales).

Ratings on several of the individual scales are combined to form several PRS composites, resulting in a total of five Army-wide PRS scores, in addition to the Leadership Potential rating (see Figure 4.5). Derivation of the composites was a rational exercise to mirror as closely as possible the IMT rating dimensions. An Overall Army-Wide composite score also is computed by averaging the ratings on all 12 scales.

In-Unit Rating Dimensions   **Composites**

    Performing Core Warrior Tasks
    Performing MOS-Specific Tasks
    Processing Information   Can Do
    Solving Problems

    Exhibiting Effort
    Exhibiting Personal Discipline   Effort & Personal Discipline

    Communicating with Others
    Contributing to the Team   Working with Others   Army-Wide

    Following Safety Procedures
    Developing Own Skills   Self-Management
    Managing Personal Matters

    Adjusting to Army Life

    Physical Fitness and Bearing

Overall Leadership Potential Rating

Figure 4.5. In-unit Army-Wide Performance Rating Scale dimensions and composite score composition.

Army Life Questionnaire (ALQ)

The ALQ was designed to measure Soldiers' self-reported attitudes and experiences in the Army. Earlier forms of the IMT and in-unit versions of the ALQ (Van Iddekinge, Putka, & Sager, 2005) were modified slightly for use in the TOPS IOT&E. The ALQ includes scales that cover (a) Soldiers' commitment and retention-related attitudes and (b) Soldiers' performance and adjustment. Each ALQ scale is scored differently depending on the nature of the attribute being measured. The Army Physical Fitness Test (APFT) score is a write-in item. Training Achievements, Training Failures, (both of which appear only on the IMT version of the ALQ), and Disciplinary Incidents are simply a sum of the "yes" responses. The remaining scales (see Table 4.2) are composed of Likert-type response scales and are scored by computing a mean of the constituent item scores. Note that most scales appear on both the IMT and in-unit versions of the scales, though the IMT version has three unique scales and the in-unit version has an MOS Satisfaction scale.

We present the results for a selected subset of IMT ALQ scales in this report because of the large number of scales. The scales not reported on are: Normative Commitment, Army Career Intentions, and Army Reenlistment Intentions. Normative Commitment was excluded because of its conceptual overlap with Attrition Cognitions. The two intentions scales were excluded because Soldiers' attitudes at this early stage of their careers are less predictive of their actual retention behavior than attitudes captured at a later point more proximal to their behavior. Results for all the applicable in-unit ALQ scales are reported.

Administrative Criteria

Attrition is a broad category that encompasses involuntary and voluntary separations for a variety of reasons (e.g., underage enlistment, conduct, family concerns, drugs or alcohol, performance, physical standards or weight, mental disorder, or violations of the Uniform Code of Military Justice [UCMJ]). The reason for separation was determined by the Soldiers' Separation Program Designator (SPD) code. Soldiers who were classified as "attrits" for reasons outside of the Soldiers' or the Army's control were excluded in our analyses (e.g., death or serious injury incurred while performing one's duties).

Data on IMT school performance and completion were extracted from ATRRS and RITMS data files (see Chapter 2). Soldiers' IMT completion status and whether he or she graduated from IMT without a restart were extracted from ATRRS. Soldiers' final AIT course grades were extracted from RITMS. Final grades from One Station Unit Training (OSUT) courses were excluded from the data file because the variance in the grades was highly restricted or based on a pass-fail metric that was redundant with the data from ATRRS.

Table 4.2. Army Life Questionnaire (ALQ) Likert-Type Scales

Scale Name	Description	Number of Items	Example Item	Likert Scale Anchors
Affective Commitment	Measures Soldiers' emotional attachment to the Army.	7	I feel like I am part of the Army 'family.'	1 (strongly disagree) to 5 (strongly agree)
Normative Commitment [a]	Measures Soldiers' feelings of obligation toward staying in the Army until the end of their current term of service.	5	I would feel guilty if I left the Army before the end of my current term of service.	1 (strongly disagree) to 5 (strongly agree)
Career Intentions	Measures Soldiers' intentions to reenlist and to make the Army a career.	3	How likely is it that you will make the Army a career?	Varies by item: 1 (strongly disagree) to 5 (strongly agree); 1 (not at all confident) to 5 (extremely confident); 1 (extremely unlikely) to 5 (extremely likely)
Reenlistment Intentions	Measures Soldiers' intention to reenlist in the Army.	4	How likely is it that you will leave the Army after completing your current term of service?	1 (strongly disagree) to 5 (strongly agree)
Attrition Cognitions	Measures the degree to which Soldiers think about attriting before the end of their first term.	4	How likely is it that you will complete your current term of service?	Varies by item: 1 (strongly disagree) to 5 (strongly agree); 1 (never) to 5 (very often)
MOS Satisfaction [b]	Measures Soldiers' satisfaction with their MOS.	6	Given my skills and abilities, I think I am in the right MOS.	1(strongly disagree) to 5 (strongly agree)
Army Life Adjustment [a]	Measures Soldiers' transition from civilian to Army life.	9	Looking back, I was not prepared for the challenges of training in the Army.	1 (strongly disagree) to 5 (strongly agree)
MOS Fit	Measures Soldiers' perceived fit with their MOS.	9	My MOS provides the right amount of challenge for me.	1 (strongly disagree) to 5 (strongly agree)
Army Fit [c]	Measures Soldiers' perceived fit with the Army.	8	The Army is a good match for me.	1 (strongly disagree) to 5 (strongly agree)

[a] Appears only on the IMT ALQ.
[b] Appears only on the in-unit ALQ.
[c] Scale has 6 items on the in-unit ALQ.

Basic Descriptive Statistics and Psychometric Properties for IMT and In-Unit Criteria

In this section, we briefly summarize the basic descriptive statistics and psychometric properties of the IMT and in-unit criterion measures for Soldiers comprising the Validation Sample (i.e., Education Tier 1 and 2, non-prior service, AFQT Category IV or above Soldiers with matching criterion data). Basic descriptive statistics and psychometric properties of the IMT criterion measures for the Full IMT and In-Unit Samples are reported in Appendix C. The basic descriptive statistics and reliability estimates observed in the validation samples were generally comparable to those for the full samples.

Job Knowledge Tests (JKTs)

A single, overall raw score was computed for each JKT by summing the total number of points Soldiers earned across the JKT items. Each of the multiple-choice items was worth one point. Depending on the format of the non-traditional items (e.g., multiple response), they were worth one or more points. To facilitate comparisons across MOS, we computed a percent correct score based on the maximum number of points that could be obtained on each MOS test. For the criterion-related validity analyses, we converted the total raw score to a standardized score (or *z*-score) by standardizing the scores *within* each MOS.

JKT scores were flagged as not useable if the Soldier (a) omitted more than 10% of the assessment items, (b) took fewer than 5 minutes to complete the entire assessment, or (c) selected an implausible response to one of the careless responding items.[11]

Table 4.3 summarizes the percent correct scores and internal consistency reliability estimates for the MOS-specific and the WTBD JKTs by education tier. Descriptive statistics for the in-unit JKTs are provided in Table 4.4. Note that the psychometric properties for the in-unit MOS-specific JKTs are not reported by individual MOS because of insufficient sample size.

Performance Rating Scales (PRS)

For the IMT MOS-specific PRS, a composite score was computed by averaging ratings across the individual scales. Ratings on the original IMT Army-wide PRS were re-scaled to a 5-point scale whose distributional properties matched the format of the new PRS and scale scores were combined or dropped to align with the revised PRS. Implementing these changes enabled us to combine ratings data from across the two versions of PRS. Accordingly, all IMT PRS results reported are based on data from the initial and revised PRS, combined, and use the names shown in Figure 4.4.

A Soldier's PRS ratings were dropped if the rater provided a familiarity rating of one (i.e., "I have had little opportunity to observe this Soldier" in the original PRS and "Not enough to judge the Soldier's performance" in the revised PRS). PRS data also were flagged as unusable if the rater omitted more than 10% of the assessment items or indicated that he or she had not observed

---

[11] The 5-minute criterion was established during the first in-unit phase of the Army Class project, which employs highly similar assessments administered via the same platform. See Knapp et al. (2012) for details.

the Soldier on more than 50% of the dimensions. Data also were removed if a rater engaged in "flat responding"—that is, ratings were removed from the data file if a rater rated 10 or more Soldiers on a particular scale and 90% or more of those rating profiles were exactly the same. Approximately 20% of Soldiers with IMT PRS were rated by more than one cadre member. Soldiers in units were rated by only one supervisor so interrater reliability estimates could not be estimated for the in-unit PRS.

Table 4.3. Descriptive Statistics and Reliability Estimates for the Job Knowledge Tests (JKTs) by Education Tier in the IMT Validation Sample

Domain/ JKT	*n*	*M*	*SD*	*Min*	*Max*	r_{WTBD}	r_{AFQT}	α
			Tier 1					
MOS-Specific								
11B/C/X + 18X	3,939	61.38	10.34	25.6	86.0	.56	.43	.77
19K	103	61.43	12.29	20.3	85.7	.53	.45	.66
31B	2,062	68.14	8.27	38.8	91.3	.49	.47	.75
68W	1,926	73.33	10.91	31.5	92.4	.53	.28	.88
88M	1,422	63.50	10.39	31.9	88.9	.55	.43	.75
91B	261	57.50	13.64	27.8	90.7	.56	.29	.90
All MOS Combined	9,713	65.39	11.26	20.3	92.4	.56	.46	.79
WTBD (Army-Wide)	11,492	64.89	12.86	6.5	96.8	--	.44	.67
			Tier 2					
MOS-Specific								
11B/C/X + 18X	129	59.48	11.32	26.1	87.0	.68	.31	.78
31B	53	69.57	7.42	52.4	85.4	.39	.31	.73
All MOS Combined	293	63.46	11.95	26.1	87.0	.63	.28	.77
WTBD (Army-Wide)	349	64.08	12.73	16.1	93.5	--	.27	.67
			Tier 1 + 2 (Combined)					
MOS-Specific								
11B/C/X + 18X	4,068	61.32	10.38	25.6	87.0	.57	.43	.77
19K	112	61.37	12.02	20.3	85.7	.51	.44	.66
31B	2,115	68.17	8.25	38.8	91.3	.49	.46	.75
68W	1,969	73.26	10.97	31.5	92.4	.53	.28	.88
88M	1,471	63.50	10.38	31.9	88.9	.56	.43	.75
91B	271	57.45	13.75	27.8	90.7	.56	.26	.90
All MOS Combined	10,006	65.33	11.28	20.3	92.4	.56	.45	.79
WTBD (Army-Wide)	11,841	64.87	12.85	6.5	96.8	--	.44	.67

*Note. M*s, *SD*s, *Min*, and *Max* reflect percent correct. α = coefficient alpha. WTBD = Warrior Tasks and Battle Drills. r_{WTBD} = correlation with WTBD JKT scores. r_{AFQT} = correlation with AFQT scores. Statistics based on fewer than 50 Soldiers are not reported. All correlations are statistically significant ($p < .01$, one-tailed).

Table 4.4. Descriptive Statistics and Reliability Estimates for the Job Knowledge Tests (JKTs) by Education Tier in the In-Unit Validation Sample

Domain/Setting/JKT	*n*	*M*	*SD*	*Min*	*Max*	r_{WTBD}	r_{AFQT}	α
			Tier 1					
MOS-Specific								
All MOS Combined	220	64.92	10.82	26.8	90.2	.54	.45	--
WTBD (Army-Wide)	591	68.01	12.03	23.1	96.2	--	.46	.56
			Tier 1 + 2 (Combined)[b]					
MOS-Specific								
All MOS Combined	220	64.92	10.82	26.8	90.2	.54	.45	--
WTBD (Army-Wide)	596	67.99	11.99	23.1	96.2	--	.46	.56

*Note. M*s, *SD*s, *Min*, and *Max* reflect percent correct. α = coefficient alpha. WTBD = Warrior Tasks and Battle Drills. r_{WTBD} = correlation with WTBD JKT scores. r_{AFQT} = correlation with AFQT scores. Statistics based on fewer than 50 Soldiers are not reported. All correlations are statistically significant ($p < .01$, one-tailed).
[b] Statistics for Tier 2 Soldiers were not reported separately because of insufficient sample size ($n < 10$).

Table 4.5 summarizes the basic descriptive statistics and reliability estimates (internal consistency and interrater reliability) for the IMT PRS by domain and education tier. Mean ratings were consistently above the mid-point, a common finding in research involving performance ratings. The IMT PRS also were highly intercorrelated (see Appendix C). Table 4.6 reports the basic descriptive statistics and reliability estimates (internal consistency only) for the in-unit Army-wide PRS by domain and education tier.

As reported in Table 4.5 the single-rater interrater reliability estimates were generally low (.30 or less). The estimates ranged from .09 to .32 for the Army-wide scales in the full IMT sample (see Table C.2). The low estimates on the MOS-specific PRS composites are particularly problematic since these composites are based on multiple PRS. We attribute these low coefficients to several interrelated issues. First, the number of ratees per rater is high, averaging about 14 for the Full IMT Sample. As a result, raters likely became fatigued during the rating task. Second, within-rater variance was generally limited, perhaps reflecting raters' inability to differentiate among individual Soldiers. Third, these data collections were not proctored, unlike prior research (e.g., Knapp & Heffner, 2009, 2010). Finally, the number of raters per ratee was small, averaging less than two, which limits the generalizability of single-rater reliability estimates. Although not all of these potential issues with the PRS can be addressed within the practical constraints of the research (e.g., collecting ratings in an unproctored setting), the interrater reliability may be improved by the PRS format changes introduced in Fall 2011. We will not be able to evaluate this change, however, until we have more data collected using the new format.

Table 4.5. Descriptive Statistics and Reliability Estimates for the Performance Rating Scales (PRS) by Education Tier in the IMT Validation Sample

Domain / PRS	*n*	*M*	*SD*	*Min*	*Max*	*α*	*IRR*
Tier 1							
Army-Wide							
Adjustment to Army	3,844	3.31	1.03	1.00	5.00	--	--
Effort & Personal Discipline	3,849	3.10	1.00	1.00	5.00	--	--
MOS Qualification Knowledge	3,666	3.25	.96	1.00	5.00	--	--
Physical Fitness & Bearing	3,834	3.18	1.02	1.00	5.00	--	--
Working with Others	3,838	3.06	1.00	1.00	5.00	--	--
Overall Performance	3,831	3.55	.83	1.00	5.00	--	--
Overall Army-Wide [a]	3,849	3.18	.88	1.00	5.00	.94	--
MOS-Specific							
11B/C/X + 18X	1,213	3.11	.84	1.00	5.00	.95	--
19K	70	3.35	.60	1.71	4.86	.87	--
31B	823	3.29	.79	1.13	5.00	.95	--
68W	633	2.78	.77	1.00	5.00	.95	--
88M	124	2.87	.75	1.20	5.00	.92	--
91B	55	3.10	1.25	1.00	5.00	.98	--
All MOS Combined	2,918	3.08	.84	1.00	5.00	.95	--
Tier 2							
Army-Wide							
Adjustment to Army	83	3.34	.97	1.00	5.00	--	--
Effort & Personal Discipline	83	3.11	.91	1.00	5.00	--	--
MOS Qualification Knowledge	81	3.41	.93	2.00	5.00	--	--
Physical Fitness & Bearing	83	3.04	.87	1.00	5.00	--	--
Working with Others	83	3.06	1.03	1.00	5.00	--	--
Overall Performance	82	3.50	.80	1.00	5.00	--	--
Overall Army-Wide [a]	83	3.19	.81	1.20	5.00	.94	--
MOS-Specific							
All MOS Combined	68	2.98	.79	1.00	5.00	--	--
Tier 1 + Tier 2 (Combined)							
Army-Wide							
Adjustment to Army	3,927	3.31	1.02	1.00	5.00	--	.18
Effort & Personal Discipline	3,932	3.10	.99	1.00	5.00	--	.22
MOS Qualification Knowledge	3,747	3.26	.96	1.00	5.00	--	.12
Physical Fitness & Bearing	3,917	3.17	1.02	1.00	5.00	--	.20
Working with Others	3,921	3.06	1.00	1.00	5.00	--	.18
Overall Performance	3,913	3.55	.83	1.00	5.00	--	.35
Overall Army-Wide [a]	3,932	3.18	.88	1.00	5.00	.94	.19

Note. PRS ratings range from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. α = coefficient alpha. IRR = Single-rater interrater reliability, estimated using G(q,k) (Putka, Le, McCloy, & Diaz, 2008). MOS-specific results based on fewer than 50 Soldiers are not reported.
[a] Composite of all scales except "Overall Performance."

Table 4.5. (Continued)

Domain / PRS	n	M	SD	Min	Max	α	IRR
MOS-Specific							
11B/C/X + 18X	1,235	3.11	.84	1.00	5.00	.95	.20
19K	75	3.35	.59	1.71	4.86	.87	.56
31B	842	3.29	.79	1.13	5.00	.95	.11
68W	650	2.78	.78	1.00	5.00	.95	.00
88M	129	2.88	.76	1.20	5.00	.93	.00
91B	55	3.10	1.25	1.00	5.00	.98	--
All MOS Combined	2,986	3.08	.83	1.00	5.00	.95	.13

Note. PRS ratings range from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. α = coefficient alpha. IRR = Single-rater interrater reliability, estimated using G(*q,k*) (Putka, Le, McCloy, & Diaz, 2008). MOS-specific results based on fewer than 50 Soldiers are not reported.
[a] Composite of all scales except "Overall Performance."

Table 4.6. Descriptive Statistics and Reliability Estimates for the Performance Rating Scales (PRS) in the In-Unit Validation Sample

Domain/PRS	n	M	SD	Min	Max	α
Tier 1						
Army-Wide						
Can Do [a]	301	4.89	1.21	1.00	7.00	.84
Effort & Personal Discipline [a]	301	5.19	1.36	1.50	7.00	.80
Physical Fitness & Bearing	300	5.29	1.54	1.00	7.00	--
Self-Management [a]	300	5.26	1.16	1.33	7.00	.83
Working with Others [a]	301	5.23	1.25	2.00	7.00	.74
Overall Leadership Potential	298	4.81	1.72	1.00	7.00	--
Army-Wide [a]	301	5.12	1.12	1.50	7.00	.95
Tier 1 + Tier 2 (Combined) [b]						
Army-Wide						
Can Do [a]	304	4.89	1.21	1.00	7.00	.84
Effort & Personal Discipline [a]	304	5.19	1.36	1.50	7.00	.80
Physical Fitness & Bearing	303	5.28	1.54	1.00	7.00	--
Self-Management [a]	303	5.27	1.16	1.33	7.00	.83
Working with Others [a]	304	5.23	1.24	2.00	7.00	.74
Overall Leadership Potential	301	4.80	1.73	1.00	7.00	--
Army-Wide [a]	304	5.12	1.11	1.50	7.00	.95

Note. PRS ratings range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. α = coefficient alpha. Interrater reliability (IRR) estimates were not computed because a limited number of Soldiers were rated by more than one supervisor. Statistics were not reported on the new PRS, Adjustment to Army Life, because *n* < 50, as of May 2012.
[a] Ratings composite comprised of two or more Army-wide PRS (see Figure 4.4).
[b] Statistics for Tier 2 Soldiers were not reported separately because of insufficient sample size (*n* < 10).

Army Life Questionnaire (ALQ)

In most cases, ALQ subscale scores were computed by averaging the responses to the items constituting each scale, after accounting for reverse coded items. The number of Training Failures, Training Achievements, Accelerated Developments, and Disciplinary Incidents were computed by summing the total number of "yes" responses. Dichotomous versions of Training Failures and Disciplinary Incidents also were constructed based on the number of versions of these measures (e.g., a Soldier with at least one self-reported restart or disciplinary incident was classified as "yes"). Award Points Earned (Weighted) was computed based on the number of promotion points a Soldier earns for the awards he or she has achieved, consistent with current Army Enlisted promotion policy (Army Regulation 600-8-19). ALQ data were flagged as unusable if the Soldier (a) omitted more than 10% of the assessment items, (b) took fewer than 5 minutes to complete the entire assessment, or (c) chose an implausible response to the careless responding item.

Table 4.7 summarizes the basic descriptive statistics and internal consistency reliability estimates for the IMT ALQ scales by criterion domain and education tier. Table 4.8 summarizes corresponding estimates for the in-unit ALQ scales. See Table 4.2 for information on the IMT and in-unit ALQ scales.

Administrative Criteria

Administrative criterion data captured separation-related information as well as performance information. Separation (attrition) data were limited to Regular Army Soldiers only because of questions regarding the reliability of administrative separation data for Reserve and National Guard Soldiers. Attrition encompasses involuntary and voluntary separations before the end of the enlistment term for a variety of reasons (e.g., underage enlistment, conduct, family concerns, drugs or alcohol, performance, physical standards or weight, mental disorder, or UCMJ violations). Soldiers who separated were coded as 1 (attrit). Soldiers who attrited for reasons that were categorized as being beyond their or the Army's control were excluded in our analyses (e.g., death or serious injury). The current analyses cover attrition through 24 months of service.

Administrative criterion data also included information about Soldiers' performance during IMT. IMT performance data were generally available for all Soldiers, regardless of component. For the variable Restarted IMT, Soldiers who graduated IMT but who had to restart at least once were coded as 1 (restart). Soldiers who had not had an opportunity to complete their IMT at the time the data were extracted were excluded from our analyses. Final AIT Grade represents the final cumulative grade administratively recorded for the Soldier upon graduation from Advanced Individual Training. A standardized version of Final AIT Grade was computed for those MOS graduating 15 or more Soldiers. Tables 4.9 and 4.10 summarize the basic descriptive statistics for the administrative criteria by education tier.

Table 4.7. Descriptive Statistics and Reliability Estimates for the Army Life Questionnaire (ALQ) by Education Tier in the IMT Validation Sample

Domain/Scale	n	M	SD	Min	Max	α
			Tier 1			
Retention						
Affective Commitment	12,039	3.87	.67	1.00	5.00	.86
Army Life Adjustment	12,039	4.09	.66	1.00	5.00	.87
Army Fit	12,039	4.07	.59	1.00	5.00	.86
Attrition Cognitions	12,039	1.52	.60	1.00	5.00	.76
MOS Fit	12,039	3.79	.83	1.00	5.00	.92
Achievement/Performance						
Disciplinary Incidents (#)	10,338	.27	.61	0.00	7.00	--
Disciplinary Incident (Y/N)	10,338	.20	.40	0.00	1.00	--
Last APFT Score	11,912	251.90	30.81	118.00	300.00	--
Training Achievement	12,028	.42	.62	0.00	2.00	--
Training Restarts (#)	12,038	.38	.62	0.00	4.00	--
Training Restart (Y/N)	12,038	.31	.46	0.00	1.00	--
			Tier 2			
Retention						
Affective Commitment	362	3.99	.64	1.57	5.00	.86
Army Life Adjustment	362	4.05	.65	1.44	5.00	.86
Army Fit	362	4.11	.61	2.13	5.00	.88
Attrition Cognitions	362	1.49	.59	1.00	3.75	.75
MOS Fit	362	3.82	.80	1.11	5.00	.92
Achievement/Performance						
Disciplinary Incidents (#)	326	.29	.66	0.00	4.00	--
Disciplinary Incident (Y/N)	326	.20	.40	0.00	1.00	--
Last APFT Score	357	243.48	31.97	136.00	300.00	--
Training Achievement	362	.40	.60	0.00	2.00	--
Training Restarts (#)	362	.38	.61	0.00	3.00	--
Training Restart (Y/N)	362	.32	.47	0.00	1.00	--
		Tier 1 + Tier 2 (Combined)				
Retention						
Affective Commitment	12,401	3.88	.67	1.00	5.00	.86
Army Life Adjustment	12,401	4.09	.66	1.00	5.00	.87
Army Fit	12,401	4.07	.59	1.00	5.00	.86
Attrition Cognitions	12,401	1.52	.60	1.00	5.00	.76
MOS Fit	12,401	3.79	.83	1.00	5.00	.92
Achievement/Performance						
Disciplinary Incidents (#)	10,664	.27	.61	0.00	7.00	--
Disciplinary Incident (Y/N)	10,664	.20	.40	0.00	1.00	--
Last APFT Score	12,269	251.66	30.88	118.00	300.00	--
Training Achievement	12,390	.42	.62	0.00	2.00	--
Training Restarts (#)	12,400	.38	.62	0.00	4.00	--
Training Restarts (Y/N)	12,400	.31	.46	0.00	1.00	--

Note. α = coefficient alpha. APFT = Army Physical Fitness Test.

Table 4.8. Descriptive Statistics and Reliability Estimates for the Army Life Questionnaire (ALQ) by Education Tier in the In-Unit Validation Sample

Domain/Scale	*n*	*M*	*SD*	*Min*	*Max*	*α*
		Tier 1				
Retention						
Affective Commitment	595	3.69	.73	1.00	5.00	.86
Army Career Intentions	595	2.70	1.11	1.00	5.00	.91
Army Fit	595	4.02	.63	1.83	5.00	.78
Attrition Cognitions	595	1.58	.63	1.00	5.00	.74
MOS Fit	595	3.35	.92	1.00	5.00	.92
MOS Satisfaction	595	3.61	.84	1.00	5.00	.91
Reenlistment Intentions	595	3.06	1.12	1.00	5.00	.78
Achievement/Performance						
Accelerated Development	591	.24	.52	0.00	3.00	--
Award Pts Earned (Weighted)	595	3.55	7.62	0.00	45.00	--
Disciplinary Incidents (#)	595	.26	.68	0.00	7.00	--
Disciplinary Incidents (Y/N)	595	.18	.38	0.00	1.00	--
Last APFT Score	578	246.69	30.10	120.00	300.00	--
		Tier 1 + Tier 2 (Combined) [a]				
Retention						
Affective Commitment	600	3.69	.73	1.00	5.00	.86
Army Career Intentions	600	2.70	1.12	1.00	5.00	.91
Army Fit	600	4.02	.63	1.83	5.00	.78
Attrition Cognitions	600	1.57	.62	1.00	5.00	.74
MOS Fit	600	3.35	.92	1.00	5.00	.92
MOS Satisfaction	600	3.61	.84	1.00	5.00	.91
Reenlistment Intentions	600	3.07	1.12	1.00	5.00	.78
Achievement/Performance						
Accelerated Development	596	.24	.52	0.00	3.00	--
Awards Pts Earned (Weighted)	600	3.52	7.59	0.00	45.00	--
Disciplinary Incidents (#)	600	.26	.69	0.00	7.00	--
Disciplinary Incidents (Y/N)	600	.18	.39	0.00	1.00	--
Last APFT Score	583	246.67	30.02	120.00	300.00	--

Note. α = coefficient alpha. APFT = Army Physical Fitness Test.
[a] Statistics for Tier 2 Soldiers were not reported separately because of insufficient sample size (*n* < 10).

Table 4.9. Base Rates for Administrative Attrition Criteria by Education Tier in the Validation Sample (Regular Army Only)

	Tier 1			Tier 2			Tier 1 + Tier 2 (Combined)		
Cumulative Attrition	N	N_{Attrit}	$\%_{Attrit}$	N	N_{Attrit}	$\%_{Attrit}$	N	N_{Attrit}	$\%_{Attrit}$
3-Month	51,529	3,121	6.1	1,219	115	9.4	52,748	3,236	6.1
6-Month	45,231	4,225	9.3	861	114	13.2	46,092	4,339	9.4
9-Month	38,986	4,140	10.6	395	58	14.7	39,381	4,198	10.7
12-Month	31,606	3,727	11.8	224	39	17.4	31,830	3,766	11.8
15-Month	22,549	2,985	13.2	89	17	19.1	22,638	3,002	13.3
18-Month	16,887	2,390	14.2	86	17	19.8	16,973	2,407	14.2
21-Month	10,182	1,548	15.2	82	17	20.7	10,264	1,565	15.2
24-Month	4,993	817	16.4	80	17	21.3	5,073	834	16.4

N = number of Regular Army Soldiers with attrition data at the time data were extracted. N_{Attrit} = number of Regular Army Soldiers who attrited at 3 through 24 months of service. $\%_{Attrit}$ = percentage of Soldiers who attrited at 3 through 24 months of service [(N_{Attrit} /N) x 100].

Table 4.10. Base Rates or Basic Descriptive Statistics for Administrative IMT Criteria by Education Tier in the Validation Sample

	Tier 1			Tier 2			Tier 1 + Tier 2 (Combined)		
Restarted Initial Military Training (IMT) [a]	N	$N_{Restarted}$	$\%_{Restarted}$	N [a]	$N_{Restarted}$	$\%_{Restarted}$	N [a]	$N_{Restarted}$	$\%_{Restarted}$
At Least Once During IMT	53,041	6,866	12.9	1,537	264	17.2	54,578	7,130	13.1
For Academic or Other Pejorative Reasons	52,185	5,996	11.5	1,516	243	16.0	53,701	6,239	11.6
For Academic Reasons	51,315	5,137	10.0	1,451	178	12.3	52,766	5,315	10.1
Final Advanced Individual Training (AIT) Grade [b]	N	M	SD	N [b]	M	SD	N [b]	M	SD
Overall Average (Unstandardized)	20,373	91.79	7.91	591	91.39	6.74	20,964	91.78	7.88
Overall Average (Standardized within MOS)	20,212	.05	.81	584	-.08	.81	20,796	.05	.82

[a] N = number of Soldiers with IMT data at the time data were extracted. $N_{Restarted}$ = number of Soldiers who restarted at least once during IMT. $\%_{Restarted}$ = percentage of Soldiers who restarted at least once during IMT [($N_{Restarted}$ /N) x 100].
[b] N = number of Soldiers with AIT school grade data at time data were extracted. Standardized grades exclude MOS with insufficient sample size ($n < 15$).

Summary

Criterion data, such as attrition, training restarts, and AIT course grades were gathered from administrative records. In addition, three types of criterion measures were adapted from previous Army research to validate the TAPAS: (a) the JKTs, (b) the PRS, and (c) the ALQ. The JKTs are completed by Soldiers in eight target MOS and measure MOS-specific and WTBD knowledge. The PRS are completed by trainer cadre (IMT) or supervisors (in-unit) and measure MOS-specific competence (IMT Soldiers only) and Army-wide constructs such as effort and leadership. Finally, the ALQ asks Soldiers to complete verifiable self-report performance items (e.g., their APFT scores) and self-report attitudinal items (e.g., Adjustment to Army life). For the validation analyses reported in subsequent chapters, scores from these measures were combined into a smaller number of meaningful composite criterion scores.

In general, the criterion measures described in this chapter exhibited acceptable and theoretically consistent psychometric properties. The exception to this was the Army-wide and MOS-specific PRS, which exhibited very low interrater reliability coefficients. Results involving the PRS should be interpreted with caution until the recently implemented improvements to the PRS can be more comprehensively evaluated.

CHAPTER 5: EVIDENCE FOR THE PREDICTIVE VALIDITY OF THE TAPAS

Chad Peddie and Michael J. Ingerick (HumRRO)

This chapter summarizes the potential of the TAPAS to predict Soldiers' performance and retention through their first unit of assignment. We begin with a brief description of our analytic approach. Next, we summarize the main findings from (a) incremental validity analyses involving the existing and newly developed TOPS composites (Can-Do, Will-Do, Adaptation) as well as an optimally weighted composite of scores on all 15 TAPAS scales and (b) bivariate correlations between the TAPAS scales and selected performance and retention criteria.

Analysis Approach

To evaluate the TAPAS' potential to enhance new Soldier selection, we examined its incremental validity in predicting first-term outcomes important to the Army over the two baseline predictors: (a) AFQT and (b) education tier. Consistent with the Army's personnel goals, we selected performance and retention-related outcomes that, as a group, provided representative coverage of the criterion space. Specific outcome measures were selected based on extant models of first-term Soldier performance and retention (Campbell, Hanson, & Oppler, 2001; Campbell, McHenry, & Wise, 1990; Knapp & Tremble, 2007; Strickland, 2005).[12]

Our analysis approach was generally consistent with previous evaluations of the TAPAS or similar experimental non-cognitive predictors (Ingerick et al., 2009; Knapp & Heffner, 2009; 2010; Trippe, Caramagno, Allen, & Ingerick, 2011). In brief, this approach involved testing a series of two-step hierarchical regression models, where scores on each criterion measure were regressed onto Soldiers' AFQT scores or education tier in the first step, followed by scores on the TOPS composites or TAPAS scales in the second step. The resulting increment in the multiple correlation (ΔR) from adding the TAPAS in the second step served as the index of incremental validity.

For the continuously scaled criteria, the models were estimated using Ordinary Least Squares (OLS) regression. Logistic regression was used for the dichotomous criteria (e.g., attrition). For the logistic regression models, we estimated point-biserial correlations (r_{pb}) in place of the traditional pseudo-R estimates. These point-biserial correlations reflected the correlation between a Soldier's predicted probability of exhibiting a selected behavior and his or her actual behavior (e.g., attriting). We estimated these correlations because of the well-known conceptual and statistical issues associated with traditional pseudo-R estimates.

A series of six regression models were estimated for each criterion measure (where k = the number of predictors in the model): (a) a model consisting of all TAPAS scales ($k = 15$), (b) the original TOPS Can-Do composite ($k = 1$), (c) the original TOPS Will-Do composite ($k = 1$), (d) the revised TOPS Can-Do composite ($k = 1$), (e) the revised TOPS Will-Do composite ($k = 1$),

---

[12] Readers are reminded that the interrater reliability estimates for the PRS were generally low. Therefore, the predictive validity evidence associated with the PRS should be interpreted with caution. The PRS measures have been revised in an attempt to mitigate this issue.

and (f) the newly developed TOPS Adaptation composite ($k = 1$).[13] Estimates for the first model, the model consisting of all TAPAS scales, represent the best-case scenario of the TAPAS predictive potential, whereas estimates for the TOPS composite models capture the predictive potential of the composites as configured for operational usage. Table 5.1 provides a summary of each of the regression models. All models were estimated separately by education tier, in addition to Tiers 1 and 2 combined.

Table 5.1. Summary of the Regression Models

Model Name	# of Predictors (*k*)	Description
All TAPAS Scales	15	Model consists of an optimally weighted composite of scores on all 15 TAPAS dimensions (or facets).
Original TOPS Can-Do Composite	1	Model consists of scores on the existing TOPS Can-Do composite.
Original TOPS Will-Do Composite	1	Model consists of scores on the existing TOPS Will-Do composite.
Revised TOPS Can-Do Composite	1	Model consists of scores on the new TOPS Can-Do composite.
Revised TOPS Will-Do Composite	1	Model consists of scores on the new TOPS Will-Do composite.
New TOPS Adaptation Composite	1	Model consists of scores on the new TOPS Adaptation composite.

To enable comparisons across the different models and education tier levels, we adjusted the observed R and ΔR for shrinkage. Comparing the aforementioned models directly, minus the shrinkage adjustments, would have been challenging and could have led to incorrect conclusions. The reason for this is because the observed R is inflated (upwards) when the sample size is small and as more predictors are entered in the model, regardless of the "true" relationship between the predictors and the criterion. This over inflation can be particularly problematic when both conditions are present – a large number of predictors and small sample size. Accordingly, adjusting the estimates for shrinkage enables a direct comparison across models for the same criterion or by education tier. We adjusted the observed R and ΔR estimates using Burket's (1964) formula for shrinkage (cf. Formula 8; Schmitt & Ployhart, 1999). This adjustment was implemented as follows:

1. Using the observed (unadjusted) correlations among the TAPAS, AFQT, or education tier and the selected criterion previously estimated, correlations between the predictors and the performance-related criteria (e.g., JKT) were adjusted for sample size and number of predictors using Burket's (1964) formula for shrinkage:

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

   where *k* equals the number of elements in the model.

2. The shrinkage-adjusted incremental validity estimates for the experimental predictors were calculated by subtracting the adjusted R associated with the AFQT-only or the education tier-only model from the adjusted R obtained from the full model (i.e., the AFQT + Experimental Predictor model or the Education Tier + Experimental Predictor model).

---

[13] Analyses were limited to Soldiers with valid TAPAS 15D-Static or 15D-CAT score data (June 2009-July 2011).

In addition to the incremental validity analyses, we examined the predictive validity of the individual TAPAS scales based on the bivariate correlations between scores on the TAPAS scales and the selected criterion measures.

Findings

Tables 5.2 to 5.18 summarize the main findings from the incremental validity analyses. Findings are organized by criterion domain: (a) IMT performance, (b) in-unit performance, and (c) attrition (data captured quarterly). Within each section, the results are further organized by model (i.e., all TAPAS scales, the original TOPS Can-Do and Will-Do composites, the revised Can-Do and Will-Do composites, and the Adaptation composite). A few notes related to interpretation of the findings:

- The results for Tier 2 Soldiers should be interpreted with caution at this stage of the TOPS evaluation because of limited criterion data for those Soldiers. Accordingly, our discussion primarily focuses on the shrinkage adjusted results for Tier 1 and 2 Soldiers (combined).[14]

- The shrinkage adjusted results presented in this report should not be directly compared to previously reported results that were not adjusted for shrinkage.

- Most of our discussion focuses on the models involving the TOPS composites because these models best evaluate the TAPAS' current operational format as well as its format envisioned for the future. Overall, the models based on an optimally weighted composite of all 15 TAPAS dimensions yielded incremental validity estimates that were comparable to or lower than the reduced models, on average, after adjusting the estimates for shrinkage.

Predicting IMT Performance

Tables 5.2 to 5.6 summarize the incremental validity results for predicting IMT performance criteria over and above the AFQT. Consistent with expectations and previous analyses, the original TOPS Can-Do composite evidenced no notable increments over the AFQT in predicting technical IMT performance (with Adj. ΔR consistently less than .01). The new TOPS Can-Do composite exhibited a similar pattern with the technical IMT performance criteria. However, the original Can-Do composite did exhibit small to modest gains over the AFQT in predicting multiple nontechnical criteria. The largest gains for the original TOPS Can-Do composite were evidenced in predicting Soldiers' attitudes towards the Army (Army Life Adjustment Adj. ΔR = .08; Army Fit Adj. ΔR = .07). The revised TOPS Can-Do composite yielded more modest predictive gains with these attitudes.

---

[14] Results for Tier 1 and Tier 2 applicants combined were generally comparable to Tier 1 only results.

Table 5.2. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting IMT Technical Performance Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1 AFQT R	Tier 1 AFQT + TAPAS R	Tier 1 ΔR	Tier 1 Adj ΔR	Tier 2 AFQT R	Tier 2 AFQT + TAPAS R	Tier 2 ΔR	Tier 2 Adj ΔR	Combined AFQT R	Combined AFQT + TAPAS R	Combined ΔR	Combined Adj ΔR
WTBD JKT	*n = 10,427*				*n = 300*				*n = 10,727*			
All TAPAS Scales (15)	**.44**	**.44**	.00	.00	**.27**	**.36**	.09	.00	**.43**	**.44**	.00	.00
Original Can-Do	**.44**	**.44**	.00	.00	**.27**	**.27**	.00	.00	**.43**	**.43**	.00	.00
Original Will-Do	**.44**	**.44**	.00	.00	**.27**	**.27**	.00	.00	**.43**	**.43**	.00	.00
Revised Can-Do	**.44**	**.44**	.00	.00	**.27**	**.28**	.01	.00	**.43**	**.44**	.00	.00
Revised Will-Do	**.44**	**.44**	.00	.00	**.27**	**.29**	.01	.01	**.43**	**.43**	.00	.00
New Adaptation	**.44**	**.44**	.00	.00	**.27**	**.30**	.03	.02	**.43**	**.43**	.00	.00
MOS-Specific JKT	*n = 8,856*				*n = 250*				*n = 9,106*			
All TAPAS Scales (15)	**.37**	**.38**	**.01**	.01	**.20**	.32	.12	.00	**.37**	**.38**	**.01**	.01
Original Can-Do	**.37**	**.37**	.00	.00	**.20**	**.22**	.01	.00	**.37**	**.37**	.00	.00
Original Will-Do	**.37**	**.37**	.00	.00	**.20**	**.21**	.00	.00	**.37**	**.37**	.00	.00
Revised Can-Do	**.37**	**.38**	.00	.01	**.20**	**.22**	.01	.00	**.37**	**.37**	.00	.00
Revised Will-Do	**.37**	**.37**	.00	.00	**.20**	**.23**	.03	.02	**.37**	**.37**	.00	.00
New Adaptation	**.37**	**.37**	.01	.00	**.20**	**.22**	.01	.00	**.37**	**.37**	.00	.00
Final AIT Grade	*n = 18,252*				*n = 505*				*n = 18,757*			
All TAPAS Scales (15)	**.30**	**.31**	**.01**	.01	**.28**	**.34**	.06	.01	**.30**	**.31**	**.01**	.01
Original Can-Do	**.30**	**.30**	.00	.00	**.28**	**.29**	.01	.00	**.30**	**.30**	.00	.00
Original Will-Do	**.30**	**.30**	.00	.00	**.28**	**.29**	.01	.00	**.30**	**.30**	.00	.00
Revised Can-Do	**.30**	**.30**	.00	.00	**.28**	**.29**	.01	.00	**.30**	**.30**	.00	.00
Revised Will-Do	**.30**	**.30**	.00	.00	**.28**	**.28**	.00	.00	**.30**	**.30**	.00	.00
New Adaptation	**.30**	**.30**	.00	.00	**.28**	**.28**	.00	.00	**.30**	**.30**	.00	.00

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes WTBD = Warrior Tasks and Battle Drills. JKT = Job Knowledge Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT and selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, $p < .01$ (one-tailed). Adjusted ΔR = Increment in R adjusted for shrinkage using Burket's (1964) formula.

Table 5.3. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting IMT Effort and Disciplinary Incidents Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT R/r_{pb}	AFQT + TAPAS R/r_{pb}	$\Delta R/r_{pb}$	Adj $\Delta R/r_{pb}$	AFQT R/r_{pb}	AFQT + TAPAS R/r_{pb}	$\Delta R/r_{pb}$	Adj $\Delta R/r_{pb}$	AFQT R/r_{pb}	AFQT + TAPAS R/r_{pb}	$\Delta R/r_{pb}$	Adj $\Delta R/r_{pb}$
Effort & Personal Discipline PRS	*n = 3,478*				*n = 67*				*n = 3,545*			
All TAPAS Scales (15)	**.07**	**.13**	**.06**	.03	.12	.38	.26	.00	**.07**	**.13**	**.06**	.04
Original Can-Do	**.07**	**.08**	.01	.01	.12	.13	.01	.00	**.07**	**.08**	.01	.01
Original Will-Do	**.07**	**.09**	**.02**	.02	.12	.12	.00	.00	**.07**	**.09**	**.02**	.02
Revised Can-Do	**.07**	**.07**	.00	.00	.12	.14	.09	.00	**.07**	**.07**	.00	.00
Revised Will-Do	**.07**	**.11**	**.04**	.04	.12	.12	.00	.00	**.07**	**.11**	**.04**	.04
New Adaptation	**.07**	**.09**	**.01**	.01	.12	.13	.01	.00	**.07**	**.08**	**.01**	.01
Disciplinary Incidents (#)(ALQ)	*n = 9,343*				*n = 276*				*n = 9,619*			
All TAPAS Scales (15)	**.04**	**.12**	**.08**	.07	.09	.24	.15	.03	**.04**	**.12**	**.08**	.07
Original Can-Do	**.04**	**.06**	**.03**	.02	.09	.10	.01	.00	**.04**	**.06**	**.03**	.02
Original Will-Do	**.04**	**.09**	**.05**	.05	.09	.09	.01	.00	**.04**	**.09**	**.05**	.05
Revised Can-Do	**.04**	**.04**	.00	.00	.09	.09	.00	.00	**.04**	**.04**	.00	.00
Revised Will-Do	**.04**	**.11**	**.07**	.07	.09	.09	.00	.00	**.04**	**.11**	**.07**	.07
New Adaptation	**.04**	**.08**	**.04**	.04	.09	.09	.00	.00	**.04**	**.08**	**.04**	.04
Training Restart (Y/N)(Administrative)	*n = 88,513*				*n = 2,881*				*n = 91,394*			
All TAPAS Scales (15)	.01	**.05**	**.04**	.04	**.05**	**.09**	.04	.00	.00	**.04**	**.04**	.00
Original Can-Do	.01	.01	.00	.00	**.05**	**.05**	.00	.00	.00	.01	.01	.00
Original Will-Do	.01	**.02**	**.01**	.01	**.05**	**.05**	.00	.00	.00	**.02**	**.02**	.00
Revised Can-Do	.01	.01	.00	.00	**.05**	**.05**	.00	.00	.00	.01	.01	.00
Revised Will-Do	.01	**.03**	**.02**	.02	**.05**	**.05**	.00	.00	.00	**.03**	**.03**	.00
New Adaptation	.01	**.03**	**.02**	.02	**.05**	**.06**	.01	.01	.00	**.03**	**.03**	.00

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. Training Restart (ATRRS) is a dichotomous criterion. Estimates for this criterion represent the observed point-biserial correlation (r_{pb}) between Soldiers' predicted probability of recycling during IMT and their actual behavior. AFQT + TAPAS = Multiple correlation (R/r_{pb}) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. $\Delta R/r_{pb}$ = Increment in R/r_{pb} from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, $p < .01$ (one-tailed). Adjusted $\Delta R/r_{pb}$ = Increment in R/r_{pb} adjusted for shrinkage using Burket's (1964) formula.

Table 5.4. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting IMT Physical Fitness Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1 AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	Tier 2 AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	Tier 1 + Tier 2 (Combined) AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
Physical Fit & Bearing PRS	*n = 3,464*				*n = 67*				*n = 3,531*			
All TAPAS Scales (15)	**.06**	**.18**	**.12**	.10	.12	.62	.40	.31	**.06**	**.18**	**.12**	.10
Original Can-Do	**.06**	**.06**	.00	.00	.12	.26	.07	.07	**.06**	**.06**	.00	.00
Original Will-Do	**.06**	**.09**	**.03**	.03	.12	.24	.12	.12	**.06**	**.08**	**.03**	.02
Revised Can-Do	**.06**	**.08**	.01	.01	.12	.13	.01	.01	**.06**	**.08**	.02	.02
Revised Will-Do	**.06**	**.13**	**.05**	.05	.12	.15	.03	.03	**.06**	**.12**	**.06**	.06
New Adaptation	**.06**	**.09**	.01	.01	.12	.12	.00	.00	**.06**	**.08**	.02	.02
Last APFT Score (ALQ)	*n = 10,799*				*n = 306*				*n = 11,105*			
All TAPAS Scales (15)	**.08**	**.31**	**.23**	.23	.07	**.34**	**.27**	.17	**.08**	**.32**	**.23**	.23
Original Can-Do	**.08**	**.08**	.00	.00	.07	.10	.03	.01	**.08**	**.08**	.00	.00
Original Will-Do	**.08**	**.12**	**.04**	.04	.07	.16	**.09**	.09	**.08**	**.13**	**.05**	.05
Revised Can-Do	**.08**	**.09**	**.01**	.01	.07	.07	.00	.00	**.08**	**.09**	**.01**	.01
Revised Will-Do	**.08**	**.27**	**.18**	.18	.07	**.30**	**.23**	.23	**.08**	**.27**	**.19**	.19
New Adaptation	**.08**	**.18**	**.10**	.10	.07	**.17**	**.10**	.10	**.08**	**.18**	**.10**	.10

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. APFT = Army Physical Fitness Test. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, $p < .01$ (one-tailed). Adjusted ΔR = Increment in *R* adjusted for shrinkage using Burket's (1964) formula.

Table 5.5. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting IMT Peer Support Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
Working with Others PRS	*n = 3,467*				*n = 67*				*n = 3,534*			
All TAPAS Scales (15)	**.06**	**.12**	**.06**	.03	.22	.51	.29	.00	**.05**	**.12**	**.06**	.04
Original Can-Do	**.06**	**.06**	.00	.00	.22	.24	.02	.00	**.05**	**.06**	.00	.00
Original Will-Do	**.06**	**.07**	.01	.01	.22	.22	.00	.00	**.05**	**.06**	.01	.01
Revised Can-Do	**.06**	**.06**	.00	.00	.22	.24	.02	.00	**.05**	**.05**	.00	.00
Revised Will-Do	**.06**	**.10**	**.04**	.04	.22	.22	.00	.00	**.05**	**.09**	**.04**	.04
New Adaptation	**.06**	**.07**	**.01**	.01	.22	.24	.02	.00	**.05**	**.07**	**.02**	.02

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. PRS = Performance Rating Scales. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, $p < .01$ (one-tailed). Adjusted ΔR = Increment in R adjusted for shrinkage using Burket's (1964) formula.

Table 5.6. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting IMT Commitment and Adjustment to the Army by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
Adjustment to the Army PRS	*n = 3,473*				*n = 67*				*n = 3,540*			
All TAPAS Scales (15)	**.04**	**.11**	**.08**	.04	.04	.51	.47	.39	.04	**.11**	**.08**	.04
Original Can-Do	**.04**	**.06**	**.02**	.02	.04	.12	.08	.08	.04	**.06**	**.02**	.02
Original Will-Do	**.04**	**.06**	.01	.01	.04	.05	.01	.00	.04	**.06**	.01	.01
Revised Can-Do	**.04**	.04	.00	.00	.04	.12	.08	.08	.04	.04	.00	.00
Revised Will-Do	**.04**	**.08**	**.03**	.03	.04	.06	.02	.00	.04	**.08**	**.03**	.03
New Adaptation	**.04**	.05	.00	.00	.04	.06	.02	.00	.04	.04	.00	.00
Army Life Adjustment (ALQ)	*n = 10,916*				*n = 310*				*n = 11,226*			
All TAPAS Scales (15)	**.07**	**.23**	**.16**	.16	.01	.21	.20	.20	**.07**	**.23**	**.16**	.16
Original Can-Do	**.07**	**.15**	**.08**	.08	.01	.09	.08	.08	**.07**	**.14**	**.08**	.07
Original Will-Do	**.07**	**.12**	**.05**	.05	.01	.10	.09	.09	**.07**	**.12**	**.05**	.05
Revised Can-Do	**.07**	**.10**	**.03**	.03	.01	.02	.01	.00	**.07**	**.10**	**.03**	.03
Revised Will-Do	**.07**	**.21**	**.14**	.14	.01	.11	.09	.09	**.07**	**.21**	**.14**	.14
New Adaptation	**.07**	**.12**	**.05**	.05	.01	.04	.03	.03	**.07**	**.11**	**.05**	.04
Army Fit (ALQ)	*n = 10,916*				*n = 310*				*n = 11,226*			
All TAPAS Scales (15)	**.04**	**.17**	**.13**	.12	.04	.21	.17	.00	**.04**	**.16**	**.13**	.11
Original Can-Do	**.04**	**.11**	**.07**	.07	.04	.06	.02	.00	**.04**	**.11**	**.07**	.07
Original Will-Do	**.04**	**.09**	**.05**	.05	.04	.04	.00	.00	**.04**	**.09**	**.05**	.05
Revised Can-Do	**.04**	**.05**	**.01**	.01	.04	.06	.02	.00	**.04**	**.05**	**.01**	.01
Revised Will-Do	**.04**	**.13**	**.10**	.09	.04	.05	.01	.00	**.04**	**.13**	**.09**	.09
New Adaptation	**.04**	**.05**	**.01**	.01	.04	.07	.03	.02	**.04**	**.05**	**.01**	.01

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, $p < .01$ (one-tailed). Adjusted ΔR = Increment in R adjusted for shrinkage using Burket's (1964) formula.

Same as the TOPS Can-Do composites, the existing and new Will-Do composites failed to predict IMT technical performance. The Will-Do composites, however, did evidence incremental validity in predicting nontechnical performance criteria. The greatest predictive gains for the existing Will-Do composites involved Soldiers' attitudes towards the Army (Army Life Adjustment Adj. $\Delta R = .05$; Army Fit Adj. $\Delta R = .05$). The new Will-Do composites incremented predictive validity in these attitudes (Army Life Adjustment Adj. $\Delta R = .14$; Army Fit Adj. $\Delta R = .09$), as well as the Army Physical Fitness Test (APFT) scores (Last APFT [ALQ]: Adj. $\Delta R = .19$). The new Will-Do composite predicted outcomes more effectively than the existing composite across all nontechnical performance domains. The gains in incremental validity of the new Will-Do composite over the existing composite were substantial at times, ranging from .01 to .14. As expected, the new Adaptation composite did not predict technical performance, but did increment predictive validity estimates of most nontechnical performance domains.

Predicting In-Unit Performance

Tables 5.7 to 5.11 summarize the incremental validity results for predicting in-unit performance over the AFQT used alone. Separate analyses were not conducted for Tier 2 Soldiers because those Soldiers had limited in-unit criterion data ($n < 10$). Consistent with the previous sections, our discussion focuses on the shrinkage adjusted results for Tier 1 and 2 Soldiers combined. The existing TOPS Can-Do composite exhibited meaningful gains in the prediction of in-unit Soldiers' commitment and satisfaction with the Army (Army Fit, Adj. $\Delta R = .12$; Affective Commitment, Adj. $\Delta R = .15$), and reenlistment intentions (Adj. $\Delta R = .12$). Estimates for the newly developed Can-Do composite demonstrated predictive validity increments above the AFQT for these outcomes as well. However, these additions in incremental validity were more modest than those in the existing composites. The new composite demonstrated predictive validity in additional outcomes such as technical performance (Warrior Tasks and Battle Drills [WTBD] Job Knowledge Test [JKT], Adj. $\Delta R = .02$). The existing TOPS Will-Do composite exhibited increments over the AFQT in predicting technical performance (Can Do Performance Rating Scale [PRS], Adj. $\Delta R = .04$), effort and discipline (PRS, Adj. $\Delta R = .08$), APFT scores (Last APFT, Adj. $\Delta R = .07$), commitment and satisfaction with the Army (Army Fit, Adj. $\Delta R = .10$; Affective Commitment, Adj. $\Delta R = .10$), and Reenlistment Intent (Adj. $\Delta R = .10$). The new TOPS Will-Do composite incremented the predictive validity of Disciplinary Incidents (Adj. $\Delta R = .04$), Accelerated Development (Adj. $\Delta R = .03$), commitment and satisfaction with the Army (Army Fit, Adj. $\Delta R = .06$; Affective Commitment, Adj. $\Delta R = .06$), and Reenlistment Intent (Adj. $\Delta R = .06$).

Table 5.7. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting In-Unit Technical Performance Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
WTBD JKT			*n = 564*				*n = 568*	
All TAPAS Scales (15)	**.46**	**.51**	**.05**	.01	**.46**	**.51**	**.05**	.01
Original Can-Do	**.46**	**.47**	**.01**	.01	**.46**	**.47**	**.01**	.01
Original Will-Do	**.46**	**.47**	**.01**	.01	**.46**	**.47**	**.01**	.01
Revised Can-Do	**.46**	**.48**	**.02**	.02	**.46**	**.48**	**.02**	.02
Revised Will-Do	**.46**	**.47**	.00	.01	**.46**	**.47**	.01	.01
New Adaptation	**.46**	**.47**	**.01**	.01	**.46**	**.47**	**.01**	.01
Can Do PRS			*n = 284*				*n = 286*	
All TAPAS Scales (15)	.08	.24	.16	.00	.08	.24	.16	.00
Original Can-Do	.08	.11	.03	.01	.08	.11	.03	.01
Original Will-Do	.08	.13	.05	.04	.08	.13	.05	.04
Revised Can-Do	.08	.09	.01	.00	.08	.09	.01	.00
Revised Will-Do	.08	.11	.03	.01	.08	.11	.03	.01
New Adaptation	.08	.12	.04	.03	.08	.12	.04	.03

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results are not reported because *n* < 10. WTBD = Warrior Tasks and Battle Drills. JKT = Job Knowledge Test. PRS = Performance Rating Scales.. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, *p* < .05 (one-tailed). Adjusted ΔR = Increment in *R* adjusted for shrinkage using Burket's (1964) formula.

Table 5.8. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting In-Unit Effort and Disciplinary Incidents Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
Effort & Personal Discipline PRS		*n = 284*				*n = 286*		
All TAPAS Scales (15)	.03	.17	.13	.00	.04	.17	.13	.00
Original Can-Do	.03	.07	.04	.04	.04	.07	.03	.02
Original Will-Do	.03	.12	**.08**	.08	.04	.12	**.08**	.08
Revised Can-Do	.03	.05	.02	.00	.04	.05	.02	.00
Revised Will-Do	.03	.06	.03	.03	.04	.06	.02	.00
New Adaptation	.03	.07	.04	.04	.04	.07	.03	.02
Disciplinary Incidents (ALQ)		*n = 566*				*n = 570*		
All TAPAS Scales (15)	.03	.16	.13	.01	.03	.15	.12	.00
Original Can-Do	.03	.04	.01	.00	.03	.04	.01	.00
Original Will-Do	.03	.04	.01	.00	.03	.04	.01	.00
Revised Can-Do	.03	.03	.01	.00	.03	.03	.00	.00
Revised Will-Do	.03	.07	.04	.04	.03	.07	.04	.04
New Adaptation	.03	.03	.00	.00	.03	.03	.00	.00
Accelerated Development (ALQ)		*n = 562*				*n = 566*		
All TAPAS Scales (15)	.03	.16	.13	.01	.03	.15	.12	.00
Original Can-Do	.03	.04	.01	.00	.03	.04	.01	.00
Original Will-Do	.03	.04	.01	.00	.03	.04	.01	.00
Revised Can-Do	.03	.03	.00	.00	.03	.03	.00	.00
Revised Will-Do	.03	.07	.04	.04	.03	.06	.03	.03
New Adaptation	.03	.03	.00	.00	.03	.03	.00	.00

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results are not reported because *n* < 10. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, *p* < .05 (one-tailed). Adjusted ΔR = Increment in *R* adjusted for shrinkage using Burket's (1964) formula.

Table 5.9. Incremental Validity Estimates for the TAPAS s over the AFQT for Predicting In-Unit Physical Fitness Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
Physical Fit & Bearing PRS		*n = 283*				*n = 285*		
All TAPAS Scales (15)	.00	**.32**	**.32**	.15	.00	**.32**	**.32**	.15
Original Can-Do	.00	.01	.01	.00	.00	.01	.01	.00
Original Will-Do	.00	.11	.11	.05	.00	.11	.10	.05
Revised Can-Do	.00	.02	.02	.00	.00	.02	.02	.00
Revised Will-Do	.00	**.24**	**.24**	.21	.00	**.24**	**.24**	.21
New Adaptation	.00	**.16**	**.16**	.12	.00	**.16**	**.16**	.12
Last APFT (ALQ)		*n = 550*				*n = 554*		
All TAPAS Scales (15)	.10	**.31**	**.21**	.14	.10	**.31**	**.21**	.14
Original Can-Do	.10	.12	.02	.01	.10	.12	.02	.01
Original Will-Do	.10	**.17**	**.07**	.07	.10	**.17**	**.07**	.07
Revised Can-Do	.10	.11	.00	.00	.10	.11	.00	.00
Revised Will-Do	.10	.13	.03	.02	.10	.13	.03	.02
New Adaptation	.10	.12	.02	.01	.10	.12	.02	.01

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results are not reported because $n < 10$. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. APFT = Army Physical Fitness Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, $p < .05$ (one-tailed). Adjusted ΔR = Increment in R adjusted for shrinkage using Burket's (1964) formula.

Table 5.10. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting In-Unit Peer Support and Leadership Potential Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
Working with Others PRS	n = 284				n = 286			
All TAPAS Scales (15)	.12	.30	.18	.03	.12	.30	.18	.03
Original Can-Do	.12	.12	.01	.00	.12	.12	.01	.00
Original Will-Do	.12	.15	.04	.01	.12	.16	.04	.03
Revised Can-Do	.12	.12	.00	.00	.12	.12	.00	.00
Revised Will-Do	.12	.15	.03	.01	.12	.15	.03	.01
New Adaptation	.12	.16	.04	.03	.12	.16	.04	.03
Leadership Potential PRS	n = 281				n = 283			
All TAPAS Scales (15)	.12	.31	**.19**	.04	**.12**	**.32**	**.18**	.06
Original Can-Do	.12	.12	.01	.00	**.12**	.13	.01	.00
Original Will-Do	.12	.15	**.04**	.01	**.12**	.16	**.04**	.03
Revised Can-Do	.12	.12	.00	.00	**.12**	.12	.08	.00
Revised Will-Do	.12	.16	**.04**	.02	**.12**	**.17**	**.05**	.04
New Adaptation	.12	.16	**.04**	.02	**.12**	.16	**.04**	.03

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results are not reported because *n* < 10. PRS = Performance Rating Scales. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, *p* < .05 (one-tailed). Adjusted ΔR = Increment in *R* adjusted for shrinkage using Burket's (1964) formula.

Table 5.11. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting In-Unit Commitment and Continuance with the Army by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR	AFQT R	AFQT + TAPAS R	ΔR	Adj ΔR
Affective Commitment (ALQ)			*n = 566*				*n = 570*	
All TAPAS Scales (15)	.02	.20	.18	.13	.02	.20	.18	.13
Original Can-Do	.02	**.17**	**.15**	.15	.02	**.17**	**.15**	.15
Original Will-Do	.02	**.12**	**.10**	.10	.02	**.12**	**.11**	.10
Revised Can-Do	.02	.08	.06	.06	.02	**.09**	.07	.07
Revised Will-Do	.02	.08	.06	.06	.02	.08	.06	.06
New Adaptation	.02	.03	.01	.00	.02	.03	.01	.00
Army Fit (ALQ)			*n = 566*				*n = 570*	
All TAPAS Scales (15)	.02	.16	.14	.05	.02	.16	.14	.05
Original Can-Do	.02	**.13**	**.11**	.11	.02	**.14**	**.12**	.12
Original Will-Do	.02	.12	**.10**	.10	.02	.12	**.10**	.10
Revised Can-Do	.02	.06	.04	.04	.02	.06	.05	.05
Revised Will-Do	.02	.08	.06	.06	.02	.08	.06	.06
New Adaptation	.02	.04	.02	.02	.02	.04	.02	.02
Reenlistment Intent (ALQ)			*n = 566*				*n = 570*	
All TAPAS Scales (15)	.02	.17	.15	.07	.02	.17	.15	.07
Original Can-Do	.02	**.13**	**.11**	.11	.02	**.14**	**.12**	.12
Original Will-Do	.02	**.12**	**.10**	.10	.02	**.12**	**.10**	.10
Revised Can-Do	.02	.06	.04	.04	.02	**.07**	**.05**	.05
Revised Will-Do	.02	.08	.06	.06	.02	.08	.06	.06
New Adaptation	.02	.04	.02	.02	.02	.04	.02	.02

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results are not reported because $n < 10$ ALQ = Army Life Questionnaire. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates in bold were statistically significant, $p < .05$ (one-tailed). Adjusted ΔR = Increment in R adjusted for shrinkage using Burket's (1964) formula.

Predicting Attrition

Table 5.12 summarizes the incremental validity beyond AFQT used alone for predicting attrition from 6 to 24 months of service. Incremental validity estimates for the TOPS composites in predicting attrition at 6-, 12-, 18- and 24-months of service were small (ΔR = .00 - .05). This pattern was true for the adjusted and unadjusted estimates of the ΔR. Results were comparable for an optimally weighted composite of all 15 TAPAS scales, except for Tier 2 Soldiers. Among Tier 2 Soldiers, incremental validity estimates based on the full TAPAS were .11 or higher, after the shrinkage adjustment. Table 5.13 summarizes the incremental validity estimates of TAPAS using education tier as a baseline predictor. The results should be treated as exploratory because of the limited variability in education tier in the current sample (about 97-98% of the sample is Tier 1). The incremental validity estimates for the TOPS composites were generally small (Δr_{pb} = .00-.09); however, prediction of attrition was stronger at later points in time (e.g., 24-months). Results were slightly larger for the 15 TAPAS scales, with incremental validity estimates of .12 for attrition at 24 months.

Bivariate Correlations of TAPAS Scales with Selected Performance and Retention Criteria

Tables 5.13 to 5.18 summarize the bivariate correlations of the AFQT, the individual TAPAS scales (rather than the composites), and selected IMT criteria by education tier. Consistent with the incremental validity results and previous research, AFQT scores were strongly positively correlated with IMT technical performance (WTBD JKT, r = .43, MOS-Specific JKT, r = .37, Final AIT Grade, r = .30; p < .01). As expected, TAPAS scales correlated more strongly with the nontechnical performance criteria, on average, than AFQT scores, although these correlations were generally smaller in magnitude (less than .20). The largest correlations were associated with Physical Conditioning scores. Physical Conditioning scores were positively correlated with self-reported APFT score (r = .29, p < .01) and Adjustment to Army Life (r = .14, p < .01), and negatively correlated with number of training restarts (r = -.04, p < .01). Intellectual Efficiency was positively correlated with WTBD JKT (r = .20, p < .01), MOS-specific JKT (r = .17, p < .01), Final AIT Grade (r = .14, p < .01), and Adjustment to Army Life (r = .12, p < .01). Multiple TAPAS scales correlated significantly with attrition at 6-, 12- 18-, and 24-months of service in the expected direction. However, none of the correlations were larger than ±.06, except for Physical Conditioning (r's = -.06 - .07).

Table 5.19 summarizes the bivariate correlations between the AFQT, the individual TAPAS scales, and selected in-unit criteria. Consistent with the IMT correlational results and previous research, AFQT scores were strongly positively correlated with in-unit technical performance (WTBD JKT, r = .46, MOS-Specific JKT, r = .35; p < .05). TAPAS scales correlated more strongly with the nontechnical performance criteria, on average, than AFQT scores. The largest correlations were associated with Achievement scores which were positively correlated with Affective Commitment (r = .12, p < .05), Army Fit (r = .14, p < .05), and Reenlistment Intentions (r = .12, p < .05). The Intellectual Efficiency scale seemed to have the strongest relationship with outcomes and was positively correlated with WTBD JKT (r = .26, p < .05), MOS-specific JKT (r = .28, p < .05), Affective Commitment (r = .09, p < .05), and Army Fit (r = .09, p < .05).

Table 5.12. Incremental Validity Estimates for the TAPAS over the AFQT for Predicting Cumulative Attrition through 24 Months of Service by Education Tier

Attrition Criterion Measure / Model	Tier 1 AFQT r_{pb}	Tier 1 AFQT + TAPAS r_{pb}	Tier 1 Δr_{pb}	Tier 1 Adj Δr_{pb}	Tier 2 AFQT r_{pb}	Tier 2 AFQT + TAPAS r_{pb}	Tier 2 Δr_{pb}	Tier 2 Adj Δr_{pb}	Combined AFQT r_{pb}	Combined AFQT + TAPAS r_{pb}	Combined Δr_{pb}	Combined Adj Δr_{pb}
6-Months	*n = 41,714*				*n = 735*				*n = 42,449*			
All TAPAS Scales (15)	**.05**	**.10**	**.04**	.04	.02	**.18**	.15	.11	**.05**	**.10**	**.04**	.04
Original Can-Do	**.05**	**.05**	**.00**	.00	.02	.05	.03	.03	**.05**	**.05**	**.00**	.00
Original Will-Do	**.05**	**.06**	**.00**	.01	.02	.02	.01	.00	**.05**	**.06**	**.00**	.00
Revised Can-Do	**.05**	**.06**	**.01**	.01	.02	.02	.05	.00	**.05**	**.05**	**.01**	.00
Revised Will-Do	**.05**	**.08**	**.03**	.03	.02	.05	.03	.03	**.05**	**.08**	**.03**	.03
New Adaptation	**.05**	**.08**	**.03**	.03	.02	.03	01	.00	**.05**	**.08**	**.03**	.03
12-Months	*n = 29,802*				*n = 199*				*n = 30,001*			
All TAPAS Scales (15)	**.06**	**.11**	**.05**	.05	.06	**.33**	.20	.12	**.06**	**.11**	**.05**	.05
Original Can-Do	**.06**	**.06**	**.00**	.00	.06	.06	.00	.00	**.06**	**.06**	**.00**	.00
Original Will-Do	**.06**	**.06**	**.01**	.00	.06	.06	.01	.00	**.06**	**.06**	**.00**	.00
Revised Can-Do	**.06**	**.06**	**.01**	.00	.06	**.13**	.10	.08	**.06**	**.06**	**.00**	.00
Revised Will-Do	**.06**	**.09**	**.03**	.03	.06	**.10**	.04	.02	**.06**	**.09**	**.03**	.03
New Adaptation	**.06**	**.09**	**.03**	.03	.06	.09	.06	.00	**.06**	**.09**	**.03**	.03
18-Months	*n = 15,809*				*n = 84*				*n = 15,893*			
All TAPAS Scales (15)	**.08**	**.13**	**.04**	.04	.09	**.57**	.48	.48	**.08**	**.13**	**.05**	.05
Original Can-Do	**.08**	**.08**	**.00**	.00	.09	.10	.01	.00	**.08**	**.08**	**.00**	.00
Original Will-Do	**.08**	**.09**	**.01**	.01	.09	.26	.17	.03	**.08**	**.09**	**.01**	.01
Revised Can-Do	**.08**	**.08**	**.00**	.00	.09	.16	.07	.00	**.08**	**.08**	**.00**	.00
Revised Will-Do	**.08**	**.11**	**.03**	.03	.09	.23	.14	.00	**.08**	**.11**	**.03**	.03
New Adaptation	**.08**	**.11**	**.03**	.03	.09	.26	.17	.13	**.08**	**.11**	**.03**	.03
24-Months	*n = 4,351*				*n = 78*				*n = 4,429*			
All TAPAS Scales (15)	**.16**	**.20**	**.04**	.02	.13	**.60**	.47	.29	**.16**	**.19**	**.03**	.01
Original Can-Do	**.16**	**.16**	**.00**	.00	.13	.13	.00	.00	**.16**	**.16**	**.00**	.00
Original Will-Do	**.16**	**.16**	**.00**	.00	.13	.28	.15	.15	**.16**	**.16**	**.00**	.00
Revised Can-Do	**.16**	**.16**	**.00**	.00	.13	.20	.07	.04	**.16**	**.16**	**.00**	.00
Revised Will-Do	**.16**	**.18**	**.02**	.02	.13	.26	.13	.13	**.16**	**.17**	**.01**	.01
New Adaptation	**.16**	**.18**	**.02**	.02	.13	**.30**	.17	.17	**.16**	**.18**	**.02**	.02

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. Δr_{pb} = Increment in r_{pb} from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates are the observed point-biserial correlation (r_{pb}) between Soldiers' predicted probability of an event (e.g., attrition, graduating IMT without a restart) and their actual behavior. Large, positive r_{pb} values mean that the TOPS composite or scale positively predicted Soldiers' actual behavior. Estimates in bold were statistically significant, $p < .01$ (one-tailed). Adjusted ΔR = Increment in R adjusted for shrinkage using Burket's (1964) formula.

Table 5.13. Incremental Validity Estimates for the TAPAS over Education Tier for Predicting Cumulative Attrition through 24 Months of Service by Education Tier

Attrition Criterion Measure / Model	Ed Tier r_{pb}	Ed Tier + TAPAS r_{pb}	Δr_{pb}	Adj Δr_{pb}
6-Months		*n = 42,449*		
All TAPAS Scales (15)	**.02**	**.09**	**.07**	.07
Original Can-Do	**.02**	**.02**	**.00**	.00
Original Will-Do	**.02**	**.04**	**.02**	.02
Revised Can-Do	**.02**	**.03**	**.01**	.01
Revised Will-Do	**.02**	**.02**	**.00**	.00
New Adaptation	**.02**	**.06**	**.04**	.04
12-Months		*n = 30,001*		
All TAPAS Scales (15)	**.01**	**.09**	**.07**	.07
Original Can-Do	**.01**	**.02**	**.01**	.01
Original Will-Do	**.01**	**.03**	**.02**	.02
Revised Can-Do	**.01**	**.03**	**.02**	.02
Revised Will-Do	**.01**	**.07**	**.05**	.05
New Adaptation	**.01**	**.07**	**.05**	.05
18-Months		*n = 11,226*		
All TAPAS Scales (15)	.01	**.11**	**.10**	.10
Original Can-Do	.01	**.02**	.01	.01
Original Will-Do	.01	**.04**	**.03**	.03
Revised Can-Do	.01	**.04**	**.03**	.03
Revised Will-Do	.01	**.07**	**.06**	.06
New Adaptation	.01	**.09**	**.08**	.08
24-Months		*n = 4,429*		
All TAPAS Scales (15)	.02	**.14**	**.12**	.11
Original Can-Do	.02	**.05**	**.03**	.03
Original Will-Do	.02	**.04**	**.02**	.00
Revised Can-Do	.02	**.06**	**.04**	.04
Revised Will-Do	.02	**.10**	**.08**	.08
New Adaptation	.02	**.11**	**.09**	.09

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes Ed Tier. Ed Tier = Educational Tier. Ed Tier + TAPAS = Multiple correlation (R) between the Ed Tier and selected TAPAS/TOPS composite scales with the targeted criterion measure. Δr_{pb} = Increment in r_{pb} from adding the selected TAPAS/TOPS composite scales over Ed Tier to the regression model ([Ed Tier + TAPAS]— Ed Tier Only). Estimates are the observed point-biserial correlation (r_{pb}) between Soldiers' predicted probability of an event (e.g., attrition, graduating IMT without a restart) and their actual behavior. Large, positive r_{pb} values mean that the TOPS composite or scale positively predicted Soldiers' actual behavior. Estimates in bold were statistically significant, $p < .01$ (one-tailed). Adjusted ΔR = Increment in R adjusted for shrinkage using Burket's (1964) formula.

Table 5.14. Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria for Tier 1 Soldiers

Measure/Scale	Technical Performance			Effort & Personal Discipline			Physical Fitness	Peer Support	Commitment & Adjustment to the Army		
	WTBD JKT	MOS-Specific JKT	Final AIT Grade	Disciplinary Incidents (ALQ)	Recycled at Least Once (Admin)	Effort & Personal Discipline (PRS)	Last APFT Score (ALQ)	Working with Others (PRS)	Army Life Adjustment (ALQ)	Army Fit (ALQ)	Adjustment to the Army (PRS)
AFQT	**.44**	**.37**	**.30**	**-.04**	.01	**.07**	**.08**	**.06**	**.07**	**-.04**	**.03**
Individual TAPAS Scales											
Achievement	**.05**	**.05**	**.08**	**-.07**	.00	**.08**	**.09**	**.06**	**.14**	**.12**	**.07**
Adjustment	**.06**	**.05**	.01	-.02	.01	-.02	.02	-.01	**.08**	.02	.00
Attention Seeking	**.04**	.01	.00	.00	.00	.02	**.08**	.03	**.07**	**.03**	.02
Cooperation	-.01	-.01	.01	.00	.00	.00	-.01	-.01	.00	.00	.01
Dominance	**.04**	.00	**.03**	**-.05**	-.01	.02	**.11**	.03	**.14**	**.10**	.04
Even Tempered	**.05**	.03	**.03**	-.02	**.01**	.01	**-.05**	.01	.02	.01	.00
Intellectual Efficiency	**.20**	**.17**	**.14**	-.02	**.01**	.01	**.03**	.00	**.12**	**.03**	.00
Non-Delinquency	.01	-.01	**.03**	-.03	.00	.00	**-.04**	-.01	.00	**.03**	.00
Optimism	.00	.00	.01	**-.03**	.00	**.05**	**.05**	.04	**.10**	**.06**	**.05**
Order	**-.08**	**-.07**	**-.03**	.00	.00	.00	**.03**	-.02	-.01	.01	-.01
Physical Conditioning	.01	-.02	-.01	**-.09**	**-.04**	**.06**	**.29**	**.06**	**.14**	**.05**	**.07**
Self Control	.02	.01	**.03**	**-.03**	**.01**	.02	.00	-.01	**.03**	**.04**	.00
Selflessness	**-.03**	**-.04**	-.01	.01	.01	-.02	-.01	-.03	-.01	**.05**	-.02
Sociability	**-.08**	**-.09**	**-.07**	.00	.00	-.01	**.04**	.01	**.05**	**.04**	.01
Tolerance	**-.03**	-.03	**-.03**	.01	**.01**	.00	.00	-.01	.02	**.05**	-.01
N	10,427	8,856	18,252	9,343	88,513	3,478	10,799	3,467	10,916	10,916	3,473

Note. Correlations in bold are statistically significant (*p* < .01, two-tailed). JKT = Job Knowledge Test. WTBD = Warrior Tasks and Battle Drills. Admin = Administrative. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. APFT = Army Physical Fitness Test.

Table 5.15. Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected Attrition Criteria for Tier 1 Soldiers

Measure/Scale	Attrition					
	3-Months	6-Months	9-Months	12-Months	18-Months	24-Months
AFQT	**-.04**	**-.06**	**-.06**	**-.06**	**-.09**	**-.17**
Individual TAPAS Scales						
Achievement	-.01	**-.01**	**-.02**	**-.02**	-.02	-.03
Adjustment	**-.02**	**-.01**	**-.02**	**-.02**	-.02	-.01
Attention Seeking	**-.03**	**-.03**	**-.03**	**-.02**	-.02	**-.04**
Cooperation	.00	.00	.00	.00	-.01	.01
Dominance	-.01	**-.02**	**-.02**	**-.02**	**-.02**	**-.06**
Even Tempered	-.01	-.01	-.01	-.01	-.01	-.01
Intellectual Efficiency	-.01	-.01	-.01	-.01	-.02	-.04
Non-Delinquency	**.01**	**.01**	.01	.01	.01	.02
Optimism	**-.02**	**-.02**	**-.03**	**-.02**	-.02	-.04
Order	.01	**.01**	**.02**	**.02**	**.02**	**.04**
Physical Conditioning	**-.05**	**-.07**	**-.07**	**-.07**	**-.09**	**-.11**
Self Control	.00	.00	.00	.00	.00	-.01
Selflessness	**.03**	**.03**	**.03**	**.03**	**.04**	.02
Sociability	-.01	-.01	.00	.01	.02	.01
Tolerance	.01	.01	.01	.01	.01	.00
N	46,938	41,714	36,401	29,802	15,809	4,351

Note. Correlations in bold are statistically significant (*p* < .01, two-tailed).

Table 5.16. *Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria for Tier 2 Soldiers*

Measure/Scale	Technical Performance			Effort & Personal Discipline			Physical Fitness	Peer Support	Commitment & Adjustment to the Army		
	WTBD JKT	MOS-Specific JKT	Final AIT Grade	Disciplinary Incidents (ALQ)	Recycled at Least Once (Admin)	Effort & Personal Discipline (PRS)	Last APFT Score (ALQ)	Working with Others (PRS)	Army Life Adjustment (ALQ)	Army Fit (ALQ)	Adjustment to the Army (PRS)
AFQT	**.27**	**.20**	**.28**	-.09	**.05**	-.12	.07	-.22	-.01	-.04	-.12
Individual TAPAS Scales											
Achievement	.05	.00	.07	-.02	.01	-.05	**.17**	-.04	.05	.02	-.06
Adjustment	.10	.05	.10	.05	-.01	-.15	.04	-.17	-.06	-.07	-.14
Attention Seeking	-.05	-.01	.00	.03	-.02	-.01	.02	.04	-.02	.05	-.08
Cooperation	-.04	-.04	.07	.04	.02	-.08	-.02	.05	-.02	-.03	.11
Dominance	.00	-.09	.05	-.02	.02	-.12	**.16**	-.18	.07	.08	-.11
Even Tempered	.06	.05	.07	.13	.02	-.04	-.08	.02	.03	-.08	-.03
Intellectual Efficiency	.05	-.02	**.21**	-.06	-.03	.03	.05	-.01	.04	.07	-.10
Non-Delinquency	.01	-.02	.05	.02	.00	-.14	-.01	.01	.06	.08	-.03
Optimism	-.06	-.10	-.01	-.01	-.02	-.07	.06	.14	.06	.01	.19
Order	-.03	-.05	.05	.10	.01	.06	-.04	-.07	.05	.00	-.08
Physical Conditioning	**-.15**	-.10	.00	-.02	-.04	.10	**.25**	.08	.08	.00	.05
Self Control	.04	-.03	.04	.03	.02	.04	.03	.03	-.07	-.03	.10
Selflessness	.01	.09	.00	.09	.02	.02	.08	-.04	-.05	-.04	.05
Sociability	.03	.02	-.03	.01	-.02	.00	.07	.14	.04	.06	.26
Tolerance	.03	-.04	-.07	-.06	.00	-.18	.07	-.17	-.01	-.02	-.24
N	300	250	505	276	2,881	67	306	67	310	310	67

Note. Correlations in bold are statistically significant (*p* < .01, two-tailed). JKT = Job Knowledge Test. WTBD = Warrior Tasks and Battle Drills. Admin = Administrative. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. APFT = Army Physical Fitness Test.

Table 5.17. Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected Attrition Criteria for Tier 2 Soldiers

	Attrition					
Measure/Scale	3-Months	6-Months	9-Months	12-Months	18-Months	24-Months
AFQT	-.01	-.02	-.06	-.07	-.09	-.12
Individual TAPAS Scales						
Achievement	-.01	-.03	-.02	.02	.03	.20
Adjustment	-.04	-.05	-.04	.11	.06	.07
Attention Seeking	-.04	-.10	-.11	-.08	**-.30**	**-.31**
Cooperation	.01	.02	.01	.05	-.12	-.14
Dominance	.02	-.01	.02	.05	.07	.08
Even Tempered	-.04	-.01	-.01	-.04	-.13	-.14
Intellectual Efficiency	.02	.00	.05	.09	-.01	-.01
Non-Delinquency	-.01	-.05	-.10	-.11	.01	.00
Optimism	-.07	-.06	-.04	.03	.04	.04
Order	.00	.00	-.02	-.02	-.03	-.04
Physical Conditioning	-.02	-.03	.01	.09	**.31**	.34
Self Control	.07	.07	.14	.22	.26	.25
Selflessness	.01	.03	.02	.03	-.03	-.02
Sociability	-.04	.00	-.04	-.04	-.13	-.14
Tolerance	-.04	-.04	.00	.03	-.05	-.05
N	1,046	735	349	199	84	78

Note. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table 5.18. Summary of the Bivariate Correlations between AFQT, Education Tier, TAPAS, and Selected IMT Criteria for Tier 1 + Tier 2 Soldiers

Measure/Scale	Technical Performance			Effort & Personal Discipline			Physical Fitness	Peer Support	Commitment & Adjustment to the Army		
	WTBD JKT	MOS-Specific JKT	Final AIT Grade	Disciplinary Incidents (ALQ)	Recycled at Least Once (Admin)	Effort & Personal Discipline (PRS)	Last APFT Score (ALQ)	Working with Others (PRS)	Army Life Adjustment (ALQ)	Army Fit (ALQ)	Adjustment to the Army (PRS)
AFQT	**.43**	**.37**	**.30**	**-.04**	.00	**.07**	**.08**	**.06**	**.07**	**-.04**	.02
Education Tier	-.01	-.02	**-.03**	.01	.00	.00	**-.05**	.00	-.01	.01	.00
Individual TAPAS Scales											
Achievement	**.05**	**.04**	**.08**	**-.07**	.00	**.07**	**.09**	**.06**	**.14**	**.12**	**.07**
Adjustment	**.06**	**.05**	.01	-.02	.01	-.02	.02	-.01	**.08**	.02	.00
Attention Seeking	**.04**	.01	.00	.00	.00	.02	**.07**	.03	**.07**	**.03**	.01
Cooperation	-.01	-.01	.01	.00	.00	.00	-.01	-.01	.00	.00	.01
Dominance	**.04**	.00	**.03**	**-.05**	-.01	.02	**.12**	.02	**.14**	**.10**	.04
Even Tempered	**.05**	**.03**	**.03**	-.01	**.01**	.01	**-.05**	.01	.02	.00	-.01
Intellectual Efficiency	**.20**	**.17**	**.14**	-.02	**.01**	.01	**.03**	.00	**.11**	**.03**	.00
Non-Delinquency	.01	-.01	**.03**	-.03	.00	.00	**-.04**	-.01	.00	**.03**	.00
Optimism	.00	-.01	.01	**-.03**	-.01	**.04**	**.05**	.04	**.10**	**.06**	**.05**
Order	**-.08**	**-.07**	**-.02**	.00	.01	.00	.02	-.02	-.01	.01	-.02
Physical Conditioning	.01	-.02	.00	**-.09**	**-.04**	**.06**	**.29**	**.06**	**.14**	**.05**	**.07**
Self Control	.02	.01	**.03**	**-.03**	**.01**	.02	.00	-.01	.02	**.04**	.01
Selflessness	**-.03**	**-.03**	-.01	.01	.01	-.02	-.01	-.03	-.01	**.05**	-.02
Sociability	**-.07**	**-.09**	**-.07**	.00	.00	-.01	**.04**	.01	**.05**	**.04**	.01
Tolerance	**-.03**	-.03	**-.03**	.00	**.01**	.00	.00	-.02	.02	**.05**	-.02
N	10,727	9,106	18,757	9,619	91,394	3,545	11,105	3,534	11,226	11,226	3,540

Note. Correlations in bold are statistically significant ($p < .01$, two-tailed). JKT = Job Knowledge Test. WTBD = Warrior Tasks and Battle Drills. Admin = Administrative. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. APFT = Army Physical Fitness Test.

Table 5.19. Summary of the Bivariate Correlations between AFQT, Education Tier, TAPAS, and Selected Attrition Criteria for Tier 1 + Tier 2 Soldiers

Measure/Scale	Attrition					
	3-Months	6-Months	9-Months	12-Months	18-Months	24-Months
AFQT	**-.04**	**-.06**	**-.06**	**-.06**	**-.09**	**-.17**
Education Tier	**.02**	**.02**	**.02**	.02	.01	.02
Individual TAPAS Scales						
Achievement	-.01	**-.01**	**-.02**	**-.02**	-.02	-.03
Adjustment	**-.02**	**-.02**	**-.02**	**-.02**	-.02	-.01
Attention Seeking	**-.03**	**-.03**	**-.03**	**-.02**	-.02	**-.05**
Cooperation	.00	.00	.00	.00	-.01	.01
Dominance	-.01	**-.02**	**-.02**	**-.02**	**-.02**	**-.06**
Even Tempered	-.01	-.01	-.01	-.01	-.01	-.01
Intellectual Efficiency	.00	-.01	-.01	-.01	-.02	-.04
Non-Delinquency	.01	.01	.01	.01	.01	.02
Optimism	**-.02**	**-.03**	**-.03**	**-.02**	-.02	-.04
Order	.01	**.01**	**.02**	**.02**	**.02**	**.04**
Physical Conditioning	**-.05**	**-.07**	**-.07**	**-.07**	**-.08**	**-.10**
Self Control	.01	.00	.00	.00	.00	.00
Selflessness	**.03**	**.03**	**.03**	**.03**	**.04**	.02
Sociability	-.01	.00	.00	.01	.01	.01
Tolerance	.01	.01	.01	.01	.01	.00
N	47,984	42,449	36,750	30,001	15,893	4,429

Note. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table 5.20. Summary of the Bivariate Correlations between AFQT, Education Tier, TAPAS, and Selected In-Unit Criteria for Tier 1 + Tier 2 Soldiers

Measure/Scale	Technical Performance		Effort & Personal Discipline		Physical Fitness	Leader-ship	Commitment to Army		
	WTBD JKT	MOS-Specific JKT	Disciplinary Incidents (ALQ)	Effort & Personal Discipline (PRS)	Last APFT Scores (ALQ)	Leadership Potential (PRS)	Affective Commitment (ALQ)	Army Fit (ALQ)	Reenlist-ment Intentions (ALQ)
AFQT	**.46**	**.35**	-.04	.10	.00	**.12**	.02	-.01	-.02
Education Tier	-.05	--	**.09**	.02	-.01	.03	**.08**	**.09**	**.08**
Individual TAPAS Scales									
Achievement	**.13**	.13	-.06	.11	.06	.07	**.12**	**.14**	**.12**
Adjustment	.05	-.03	-.02	-.02	-.03	-.07	.06	.07	.02
Attention Seeking	-.01	.04	.08	-.08	-.03	-.08	.00	**.10**	-.02
Cooperation	.04	-.05	-.06	-.10	-.04	-.11	.04	.05	.06
Dominance	**.12**	-.01	.05	-.04	**.10**	-.01	.06	**.09**	.05
Even Tempered	**.11**	.00	-.06	-.03	**-.09**	.04	**.11**	**.09**	**.09**
Intellectual Efficiency	**.26**	**.28**	-.04	.06	.01	.00	**.09**	**.09**	.03
Non-Delinquency	.04	.09	-.04	-.03	-.07	-.04	**.09**	.07	.05
Optimism	-.01	.02	-.02	.07	.04	.11	.06	**.13**	.06
Order	**-.09**	-.04	-.01	.02	.03	-.05	-.02	-.02	-.01
Physical Conditioning	.03	-.05	-.06	.08	**.28**	.08	-.01	.01	.00
Self Control	.08	.11	-.05	.06	.01	.02	.08	.04	.03
Selflessness	**-.10**	**-.14**	-.03	-.03	-.04	-.02	.02	.01	.03
Sociability	**-.13**	**-.25**	.00	-.05	.02	-.03	.04	.05	.01
Tolerance	**-.09**	**-.15**	.03	-.01	.02	.01	.06	.08	.04
N	568	209	570	286	554	283	570	570	570

Note. Correlations in bold are statistically significant ($p < .05$, two-tailed). The correlation of Education Tier with MOS-specific JKT scores could not be computed because no Tier 2 Soldiers had JKT scores, at this time. JKT = Job Knowledge Test. WTBD = Warrior Tasks and Battle Drills. PRS = Performance Rating Scales. ALQ = Army Life Questionnaire. APFT = Army Physical Fitness Test.

Summary

This chapter summarized results from the fifth cycle of evaluating criterion-related validity in the TOPS IOT&E. Overall the existing and new TOPS composites demonstrated modest incremental validity over the AFQT in predicting first-term Soldier performance and retention. Incremental validity estimates (adjusted for shrinkage) were consistently .03 or less, on average, for the new and existing Can-Do and Will-Do composites over the AFQT used alone. Prior evaluation cycles have shown the existing Will-Do composite to exhibit somewhat higher incremental validity, on average, than the Can-Do composite. Whereas that same general trend continued in this cycle, the newly developed Will-Do composite was associated with the greatest incremental validity gains, on average across all domains. Overall, incremental validity gains associated with the New Will-Do composite were the highest (or tied for highest) of any composite on 15 of the 23 total non-technical outcomes across the IMT, attrition, and in-unit criterion domains. At the scale-level, Intellectual Efficiency emerged as the single best correlate of technical performance criteria (r's ranging from .17 to .28) across IMT and in-unit. In IMT, Physical Conditioning represented the single best and most consistent correlate of nontechnical performance and attrition (r's ranging from -.09 to .29), while Achievement was the strongest predictor in-unit (r's ranging from .12 to .14).

CHAPTER 6: EVALUATION OF TAPAS POTENTIAL FOR CLASSIFICATION PURPOSES

Bethany H. Bynum, Matthew Trippe, Ted Diaz, and Michael Ingerick (HumRRO)

Introduction

Similar to previous research (Ingerick et al., 2009; Knapp & Heffner, 2011; Tripp et al., 2011), we evaluated the classification potential of the TAPAS using (a) Horst's (1954, 1955) index of differential validity (H_d) and (b) Brogden's expected criterion scores of optimally matched individuals (De Corte, 2000). Conceptually, H_d provides an index of the TAPAS' potential 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, Brogden's expected criterion scores reflect the predicted criterion scores for Soldiers optimally matched to a sample of jobs using the TAPAS. A common way to summarize predicted criterion scores is with the mean predicted criterion score (MPCS). The greater the MPCS, the higher Soldiers are predicted to perform or persist, on average, when matched to a sample of jobs using the TAPAS. Predicted criterion scores are traditionally expressed in a standardized metric with a known distribution that is common across MOS. In contrast, we report results in the metric of the targeted criterion (e.g., predicted percentage of Soldiers attriting) to facilitate interpretation.

Although the two classification 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 TAPAS. Whereas H_d provides information on cross-job differences (or variability) in Soldiers' predicted criterion scores resulting from the use of the TAPAS to classify Soldiers into a sample of jobs, the MPCS speaks to the average level at which Soldiers are predicted to score on the targeted criterion (e.g., performance, retention). Accordingly, H_d can be viewed as a descriptive measure of classification potential that does not include all of the factors modeled in Brogden's expected criterion scores. MPCS is a more comprehensive measure that accounts for a number of additional factors, including the percentage of Soldiers allocated to each MOS and the optimal matching of Soldiers to MOS to maximize (or minimize) a targeted criterion.

Approach to Estimating the Classification Potential

Comparable to the incremental predictive validity analyses, we estimated the increment in H_d and MPCS resulting from using the TAPAS over an optimally weighted combination of the existing ASVAB subtests to enhance new Soldier classification.[15] Consistent with the Army's personnel management objectives, we investigated the TAPAS' potential for enhancing both performance and retention-related criteria.

---

[15] ASVAB subtests include General Science (GS), Arithmetic Reasoning (AR), Math Knowledge (MK), Electronics Information (EI), Auto Shop (AS), Mechanical Comprehension (MC), Verbal composite (VE) of Word Knowledge (WK) and Paragraph Comprehension (PC). Assembling Objects (AO) was not included because (a) it is not currently part of any existing Aptitude Area composites and (b) missing data are prevalent in this subtest.

As in the last cycle of classification analyses (Knapp & Heffner, 2012), we extended Brogden's classification framework to handle dichotomous criteria. Overall, the steps in the extended approach closely follow those for the traditional approach to continuous criteria. The main difference between the traditional and extended approach is in how the classification composites used for optimally matching Soldiers to jobs are derived. In the traditional approach, the classification composites are modeled using linear multiple regression (e.g., ordinary least squares or conditional normal regression). In the extended approach, a logistic probability model is applied to obtain predicted probabilities. The predicted probabilities are then transformed using the logit function to derive a linear composite of predictors. The linear composites then become input to DeCorte's (2000) multivariate normal formulation of Brogden's classification framework.

In sum, estimating the classification potential of the TAPAS for optimizing each criterion consisted of the following general steps:

1. Estimate the linear predictor (classification) composite for each MOS.

2. Estimate the observed (uncorrected) predictor-linear composite covariance matrix for each MOS.

3. Correct the predictor-linear composite covariances from Step 2 for multivariate range restriction on the ASVAB and TAPAS using the Accession sample as the reference population (Lawley, 1943).[16]

4. Compute the multiple correlation of the linear composites using the corrected predictor-linear composite covariance from Step 3.

5. Correct the multiple correlations of linear composites for shrinkage (Burket, 1964).

6. Using the corrected covariance matrices from Step 5, compute two indices of classification potential: (a) (H_d) and (b) Brogden's expected criterion scores of optimally assigned individuals (DeCorte, 2000).

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). Accordingly, the estimates reported reflect the *potential* of the TAPAS to enhance new Soldier classification and not the *actual* expected gains if the measure was used operationally. Second, the ASVAB baseline consists of an optimally weighted composite of scores on the ASVAB subtests, excluding AO, keyed to each criterion and not the existing Aptitude Area (AA) composites. Third, the results reported could differ if an alternative sample of MOS or set of criterion measures were examined. Accordingly, we focused our analyses on the MOS targeted in the TOPS IOT&E. Fourth, there are no standards or conventions for interpreting the magnitude of or gain in H_d relative to some baseline. As for MPCS, previous simulation research indicates that increments in MPCS as low as .10 carry significant and practical operational 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;

---

[16] The Accession sample is comprised of Soldiers from the Applicant Sample who signed an enlistment contract.

Scholarios, Johnson, & Zeidner, 1994). Those MPCS values, however, were reported in a standardized metric, while the results presented in the current evaluation are in the metric of the criterion analyzed. Accordingly, the current results are best interpreted by the degree of overlap in distributions of predicted criterion scores. Specifically, the less the overlap in the distributions of predicted criterion scores, the greater the classification gains.

Results

Table 6.1 provides a summary of the classification potential of the TAPAS relative to the ASVAB by criterion measure. Criterion measures were selected based on expectations of cross-MOS differences and sample sizes for the target MOS. Six criterion measures were selected: (a) attrition (3-months through 18-months of service); (b) MOS-specific Job Knowledge Test (JKT); (c) restarted IMT at least once; (d) last Army Physical Fitness Test (APFT) score; (e) perceived MOS fit; and (f) perceived Army fit.[17] MPCS values reported in Table 6.1 reported overall means computed across MOS and weighted by the corresponding MOS allocation percentages (see notes in Tables 6.2 through 6.3 for the allocation percentages). Overall, the TAPAS showed the greatest potential to increment the ASVAB in minimizing attrition and IMT restarts. The TAPAS demonstrated minimal to no gains relative to the ASVAB in maximizing the other IMT performance criteria and retention-related attitudes.

Table 6.1. Summary of Overall Classification Potential Indices of the TAPAS Relative to the ASVAB by Criterion Measure

Domain/Measure	Number of MOS	N	H_d ASVAB	H_d ASVAB+ TAPAS	MPCS ASVAB	MPCS ASVAB+ TAPAS
Attrition						
3-Month	8	55,052	.022	.067	6.0%	5.3%
6-Month	8	55,052	.018	.051	9.8%	8.8%
12-Month	8	55,052	.020	.051	11.6%	10.4%
18-Month	8	55,052	.042	.079	13.4%	11.5%
IMT Performance						
MOS-Specific JKT	6	49,776	.026	.042	64.2%	64.4%
Restarted IMT at Least Once	8	55,052	.019	.050	12.6%	11.7%
Last APFT Score	7	52,755	.012	.034	252.21	252.48
Retention-Related Attitudes						
MOS Fit	7	52,755	.007	.006	3.75	3.74
Army Fit	7	52,755	.022	.026	3.89	4.08

Note. Number of MOS = Number of target MOS modeled. N = Number of Soldiers in the Validation Sample from the target MOS with non-missing predictor data. H_d = Average standardized mean difference between all possible pairs of predicted criterion scores for the target MOS modeled. MPCS = Mean predicted criterion scores for Soldiers optimally matched to the target MOS using the ASVAB or ASVAB+TAPAS. JKT = Job Knowledge Test. APFT = Army Physical Fitness Test. MPCS values were reported in the metric of the criterion measure. For example, the MPCS for attrition represents the aggregated predicted attrition rate, while the MPCS for the MOS-specific JKT represents the average percent correct score.

---

[17] 24-month attrition was excluded from these analyses because not all of the eight target MOS had sufficient base rates or sample sizes for estimating the classification composites.

Tables 6.2 through 6.4 summarize the MPCS results, overall and by target MOS.[18]

Table 6.1 shows that when the ASVAB subtests are used to classify new Soldiers to the target MOS, the overall predicted attrition rate is 6.0%, 9.8%, 11.6%, and 13.4% for 3-, 6-, 12- and 18-months, respectively. The overall predicted attrition rate drops to 5.3%, 8.8%, 10.4%, and 11.5%, respectively, when the TAPAS is added. Although an overall reduction in 3-month, 6-month, and 12-month attrition of just 1% is certainly modest, the reduction in 18-month attrition was nearly 2%. The incremental decrease in attrition when TAPAS is included in the classification increases with time. That is, the difference between ASAVAB and ASVAB + TAPAS increases from .70% at 3 months to 1.9% at 18 months.

Table 6.2 summarizes the mean predicted 3-month, 6-month, 12-month and 18-month attrition rates across all eight target MOS. Evaluation of the distribution of attrition rates reveals more substantive improvements particularly for 42A, 19K, and 91B. For example, the mean predicted 3-month attrition rate for 42A is 5.8% when the ASVAB alone is used to classify, but this rate is reduced to 2.7% when the TAPAS is added to the classification model. That is, the mean predicted 3-month attrition rate is reduced by roughly half for this MOS. The mean predicted 12-month attrition rate for 19K is 10.8% when ASVAB alone is used to classify, but drops to 5.3% when TAPAS is also considered.

Table 6.2 also reveals that the attrition rate for 11B remains virtually unchanged when the TAPAS is added to a classification model containing the ASVAB. There are a few possible explanations for this finding. First, the classification model must allocate nearly half (42%) of Soldiers to this MOS. Because 11B represents such a large proportion with respect to the other MOS, the model cannot be as selective on who is classified into this MOS. Similarly, 11B attrition is harder to predict than other MOS in this model. The overall amount of variance explained in attrition by the ASVAB and the TAPAS is lower in 11B than for the other MOS.[19] Thus, when the model tries to optimize (reduce) attrition in this MOS, it is difficult to achieve gains for 11B. Stated more generally, MOS with relatively small allocation percentages and relatively high multiple correlations have the most to gain in the classification model. In the analysis presented in Table 6.2, 11B has both a higher allocation percentage and a lower multiple correlation, relative to the other MOS. Moreover, we applied a correction for cross validity (Burket, 1964) to the multiple correlations that are part of the basis for the classification estimates, which will penalize the less parsimonious model including both the ASVAB and TAPAS scales.

---

[18] See Appendix D for a more extensive summary of the MPCS results by and within MPCS.
[19] Multiple correlations (R) serve as input to the classification model. The estimated R's are not reported in Tables 6.2 and 6.3.

Table 6.2. Classification Potential of the TAPAS Relative to the ASVAB for Minimizing 3-Month, 6-Month, 12-Month and 18-Month Attrition

| | *Predicted Percentage (%) of Soldiers Attriting* | | | | | | | | | | | |
| | 3-Month | | | 6-Month | | | 12-Month | | | 18-Month | | |
	Obs	*ASVAB Only*	*ASVAB + TAPAS*	*Obs*	*ASVAB Only*	*ASVAB + TAPAS*	*Obs*	*ASVAB Only*	*ASVAB + TAPAS*	*Obs*	*ASVAB Only*	*ASVAB + TAPAS*
Overall	7.2	6.0	5.3%	11.2	9.8	8.8	13.4	11.6	10.4	16.3	13.4	11.5
11B	7.8	7.4	7.6	12.0	11.4	12.0	14.1	13.4	14.1	17.1	16.4	17.0
19K	5.0	3.2	1.0	11.1	9.4	5.3	14.2	10.8	5.3	17.7	13.3	5.5
25U	8.1	5.8	4.9	11.0	9.4	7.6	11.8	12.1	10.9	16.4	10.9	10.0
31B	6.6	4.4	4.3	12.5	8.7	7.7	14.0	10.2	9.2	15.3	10.9	7.1
42A	6.2	5.8	2.7	9.8	7.2	4.6	8.7	7.3	3.6	11.5	6.6	2.0
68W	6.0	6.5	6.0	9.0	9.7	8.5	12.0	12.6	11.6	13.4	13.8	11.6
88M	7.7	4.9	3.6	11.6	9.4	7.1	13.1	9.8	8.6	16.4	13.2	12.1
91B	5.8	4.5	2.8	8.9	7.0	4.2	12.9	10.5	7.6	16.7	14.1	12.7

Note. Obs = Observed aggregate attrition rate by MOS or for the target MOS overall. ASVAB Only = Predicted aggregate attrition rates when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Predicted aggregate attrition rates when the ASVAB + TAPAS was used to classify Soldiers into the MOS listed. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (11B = 42%, 19K = 2%, 25U = 4%, 31B = 10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, *n* = 23,058; 19K, *n* = 1,236; 25U, *n* = 2,297; 31B, *n* = 5,691; 42A, *n* = 2,979; 68W, *n* = 6,902; 88M, *n* = 6,731; 91B, *n* = 6,158).

Table 6.3 summarizes the classification potential of the TAPAS for maximizing or minimizing selected IMT performance criteria (MOS-specific JKT scores, IMT restarts, last APFT score). Results for the MOS-specific JKT scores evidenced minimal to no classification gains for the TAPAS (*MPCS* = 64.4%) over the ASVAB (*MPCS* = 64.2%), overall or by MOS. This pattern of results could be due in part to the reduced number of target MOS. More likely, however, these findings can be explained by the lack of incremental validity of the TAPAS over the ASVAB in predicting MOS-specific JKT scores (see Chapter 5). Results for the last APFT score were comparable. The TAPAS (*MPCS* = 252.5) showed no significant gains over the ASVAB (*MPCS* = 252.2) in maximizing AFPT scores, overall or by MOS. Conversely, for IMT restarts, the TAPAS (*MPCS* = 11.7%) did show classification gains over the ASVAB (*MPCS* = 12.6%). Adding the TAPAS reduced the overall predicted IMT restart rate by a percentage point, on average. At the MOS-level, the most significant reductions in predicted IMT restarts rates were associated with 19K (from 10.3% to 5.5%), 42A (from 3.5% to 2.2%), and 88M (from 10.4% to 8.9%).

Table 6.4 summarizes the classification potential of the TAPAS for maximizing retention-related attitudes predictive of re-enlistment (perceived MOS fit, perceived Army fit). The TAPAS (*MPCS* = 3.74) did not significantly increment the ASVAB (*MPCS* = 3.75) in maximizing MOS fit, both overall and by MOS, on average. In contrast, the TAPAS (*MPCS* = 4.08) exhibited significant gains in maximizing perceived Army fit over the ASVAB (*MPCS* = 3.89). At the MOS-level, 88M and 91B showed the largest gains in perceived fit, on average. Several factors explain these results. First, the H_d values found in Table 6.1 indicate that the cross-MOS differences in the predicted criterion scores were generally low, on average. Second, the observed values on perceived MOS and Army fit were high to start with, with most Soldiers in the target MOS reporting high perceived MOS and Army fit. Finally, sample sizes for 19K and 42A were low compared to other models. The low sample sizes, when combined with the shrinkage adjustment, over-penalized the TAPAS because of the large number of dimension scores in the model.

Table 6.3. Classification Potential of the TAPAS Relative to the ASVAB for Maximizing IMT Performance Criteria

	MOS-Specific JKT (% Correct)			Restarted IMT (% with a Restart)			Last APFT Score (100-300)		
	Obs	*ASVAB Only*	*ASVAB + TAPAS*	*Obs*	*ASVAB Only*	*ASVAB + TAPAS*	*Obs*	*ASVAB Only*	*ASVAB + TAPAS*
Overall	65.5	64.2	64.4	13.8	12.6	11.7	251.6	252.2	252.5
11B	61.5	60.6	61.0	15.5	15.1	15.4	250.5	252.2	250.9
19K	60.8	59.7	59.6	15.0	10.3	5.5	253.9	256.3	258.6
25U	--	--	--	14.5	10.0	8.4	--	--	--
31B	68.5	74.0	73.4	15.0	13.3	11.6	258.4	258.9	258.7
42A	--	--	--	4.4	3.5	2.2	247.6	249.8	250.9
68W	73.4	69.1	70.0	16.1	17.8	17.8	251.4	250.8	252.4
88M	63.9	66.4	67.0	10.9	10.4	8.9	248.2	250.8	251.8
91B	57.5	58.5	58.1	9.4	9.1	8.5	245.0	246.6	252.4

Note. Obs = Observed mean predicted criterion score by MOS or for the target MOS overall. ASVAB Only = Mean predicted criterion score when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Mean predicted criterion score when the ASVAB + TAPAS was used to classify Soldiers into the MOS listed. JKT = Job Knowledge Test. APFT = Army Physical Fitness Test. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (MOS-Specific JKT, 11B = 46%, 19K = 2%, 31B=11%, 68W=14%, 88M = 14%, 91B = 12%; Restarted IMT, 11B = 42%, 19K=2%, 25U=4%, 31B=10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%; Last APFT Score, 11B = 44%, 19K=2%, 31B=11%, 42A =6%, 68W=13%, 88M = 13%, 91B =12%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, *n* = 23,058; 19K, *n* = 1,236; 25U, *n* = 2,297; 31B, *n* = 5,691; 42A, *n* = 2,979; 68W, *n* = 6,902; 88M, *n* = 6,731; 91B, *n* = 6,158).

Table 6.4. Classification Potential of the TAPAS Relative to the ASVAB for Maximizing Retention-Related Attitudes

	MOS Fit (1=Low Fit-5 = High Fit)			Army Fit (1=Low Fit-5 = High Fit)		
	Obs	*ASVAB Only*	*ASVAB + TAPAS*	*Obs*	*ASVAB Only*	*ASVAB + TAPAS*
Overall	3.80	3.75	3.74	4.08	3.89	4.08
11B	3.92	3.92	3.91	4.11	3.93	4.12
19K	3.42	3.39	3.42	4.07	3.92	4.10
25U	--	--	--	--	--	--
31B	3.87	3.87	3.87	4.13	3.95	4.14
42A	3.69	3.65	3.63	4.04	3.82	3.99
68W	3.96	3.94	3.93	3.92	3.75	3.92
88M	3.22	3.29	3.29	4.16	3.96	4.17
91B	3.51	3.47	3.48	3.87	3.71	3.96

Note. Obs = Observed mean predicted criterion score by MOS or for the target MOS overall. ASVAB Only = Mean predicted criterion score when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Mean predicted criterion score when the ASVAB + TAPAS was used to classify Soldiers into the MOS listed. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (IMT MOS-Specific JKT, 11B = 46%, 19K = 2%, 31B = 11%, 68W = 14%, 88M = 14%, 91B = 12%; Restarted IMT, 11B = 42%, 19K = 2%, 25U = 4%, 31B=10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%; Last APFT Score, 11B = 44%, 19K = 2%, 31B = 11%, 42A = 6%, 68W=13%, 88M = 13%, 91B =12%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, *n* = 23,058; 19K, *n* = 1,236; 31B, *n* = 5,691; 42A, *n* = 2,979; 68W, *n* = 6,902; 88M, *n* = 6,731; 91B, *n* = 6,158).

Summary

Overall, the results presented in this chapter demonstrate that the TAPAS has some potential for incrementing the ASVAB when matching new Soldiers to entry-level MOS to optimize first-term outcomes. Specifically, the TAPAS evidenced the greatest classification gains over the ASVAB in minimizing aggregate attrition and IMT restart rates and maximizing perceived Army fit. The TAPAS exhibited minimal to no significant gains over the ASVAB, on average, in maximizing other IMT performance criteria and retention-related attitudes.

The current results were attributable to some extent to the modeling approach employed here, in particular the percentages used to allocate Soldiers to the target MOS to estimate MPCS. Infantry (11B) accounted for a large portion of the allocation. One strategy for mitigating the potential overweighting of 11B in the future would be to introduce an "other" MOS category that Soldiers could be assigned if Soldiers do not fit well into one of the target MOS. Doing so allows for a specified percentage of Soldiers to remain unclassified or matched to MOS outside of the target MOS, which would relax a constraint that prevented the classification composites from demonstrating their full potential.

CHAPTER 7: SUMMARY AND A LOOK AHEAD

Deirdre J. Knapp (HumRRO), Kate LaPort, Tonia S. Heffner, and Leonard A. White (ARI)

Summary of the TOPS IOT&E Method

In an effort to expand the basis on which applicants are evaluated for enlistment, the Army is conducting an initial operational test and evaluation (IOT&E) of the Tier One Performance Screen (TOPS). The TOPS assessments, including the Tailored Adaptive Personality Assessment System (TAPAS), the Information/Communications Technology Literacy (ICTL) test, and starting in FY2014, the Work Preferences Assessment (WPA), are being administered to non-prior service applicants testing at MEPS locations.

To evaluate the TAPAS, ICTL, and WPA, the Army is collecting training criterion data on Soldiers in selected MOS as they complete their IMT. The criterion measures include job knowledge tests (JKTs), an attitudinal person-environment fit assessment (the Army Life Questionnaire; ALQ), and performance rating scales (PRS) completed by the Soldiers' cadre members. Course grades and completion rates are obtained from administrative records for all Soldiers, regardless of MOS. The plan is to construct analysis datasets and conduct cumulative validation analyses at 6-month intervals throughout the IOT&E period.

Job performance data are also being collected from Soldiers in their units to gather data on Soldiers who completed the TAPAS (and WPA and ICTL) at entry. These measures again include JKTs, the ALQ, and supervisor ratings. Finally, the separation status of all Soldiers who took the TAPAS at entry is being tracked throughout the course of the research.

The June 2012 data file (containing data collected through May 2012), which was the basis for analyses documented in this report, includes a total of 282,563 applicants who took the TAPAS between June 2009 and March 2012. Of these total applicants, 250,884 were in the TOPS Applicant Sample. The Applicant Sample excluded Education Tier 3, AFQT Category V, and prior service applicants. The validation sample sizes were considerably smaller, with the IMT Validation Sample comprising 13,238 Soldiers, the In-Unit Validation Sample comprising 604 Soldiers, and the Administrative Validation Sample (Soldiers with criterion data from at least one administrative source) comprising 113,618 Soldiers.

The JKT, ALQ, and administrative criterion measures exhibited acceptable and theoretically consistent psychometric properties. The Army-wide and MOS-specific PRS, however, continued to exhibit very low interrater reliability. The PRS instruments have been revised to change both content and format in an attempt to improve their psychometric characteristics. Details of these changes were summarized in Chapter 3. Results based on supervisor ratings should be interpreted with caution until we have sufficient data to evaluate these changes.

Summary of Evaluation Results to Date

Evaluation results thus far suggest that, while the magnitude of the predictive validity estimates are not as large as those found in the experimental *Expanded Enlistment Eligibility Metrics* (EEEM) research (Knapp & Heffner, 2010), the TAPAS holds promise for new Soldier selection. Results of the incremental validity analyses indicate that the TAPAS predicts important first-term criteria over and above the AFQT, especially measures tapping non-technical aspects of Soldier performance, such as physical fitness, effort, and personal discipline. The TAPAS exhibited its greatest predictive gains over education tier. Many of the scale-level coefficients are consistent with a theoretical understanding of the TAPAS scales, suggesting that the scales are measuring the characteristics that they are intended to measure. Results of the classification analyses indicated that the TAPAS has the potential to enhance matching new Soldiers to MOS, particularly for minimizing attrition.

Looking Ahead

Changes to Predictor Measures

In FY2014, a third series of new adaptive forms of the TAPAS will be introduced at the MEPS. Each form measures 13 dimensions. All three 13D forms assess the same 10 core dimensions, plus three of seven experimental dimensions. The seven experimental dimensions assessed vary by version. In total, the newer versions of the TAPAS collectively measure 17 dimensions. The experimental dimensions will be evaluated for potential use in revised or new TOPS composites, once sufficient data are available.

Along with the new TAPAS versions, a new TOPS screen also will be fielded to select new Soldiers. The new TOPS screen will be based on the new TOPS composites (Can-Do, Will-Do, Adaptation). The new TOPS composites incorporate several enhancements over the existing composites and were constructed from analyses of the most current IOT&E data. Chapter 5 summarized the predictive validity of the new TOPS composites, compared to the existing composites. Overall, the new Will-Do and Adaption composites evidenced higher predictive validity, on average, than the existing composites, although the magnitude of the gains varied by criterion.

In the next evaluation report, there will also be a sufficiently large sample of Soldiers for whom we have both ICTL and criterion data to include this experimental predictor in the evaluation results. To date, the WPA is still not being administered to Army applicants, so inclusion of this predictor in the IOT&E remains on hold.

Changes to Criterion Measures

We revised both the training and in-unit performance rating scales in an effort to improve their psychometric properties, as summarized in Chapter 4. For example, we changed the format of the IMT MOS-specific rating scales to a 5-point relative performance rating rather than a 7-point absolute performance rating and greatly reduced the amount of reading required. The IMT Army-

wide PRS have been similarly changed, and the number of dimensions rated has been reduced. The impact of changes to the revised IMT PRS will be evaluated once sufficient data are available.

Finally, in the next evaluation cycle we will introduce a revised set of criterion measure composite scores. These scores will be based on theoretically-derived factor analytic work and consideration of specific areas of Army policy-maker interests.

Analyses

The semi-annual reports will continue to include basic psychometric, validation, and incremental validation analyses. As needed, we will examine the comparability of new TAPAS versions to prior forms before determining if the data can be combined for purposes of analysis. We will evaluate changes to the criterion measures, in particular the IMT PRS, once sufficient data have been collected on the revised measures. We also will try an alternative approach to modeling MOS classification outcomes that may result in evaluation results that can more meaningfully inform policy decisions.

The next set of TOPS evaluation analyses will be conducted based on a data file constructed in December 2012. We will continue to update or to modify our analysis plans as the Army's goals for the TOPS IOT&E evolve or to better meet the informational needs of Army stakeholders.

REFERENCES

Allen, M.T., Cheng, Y.A., Putka, D.J., Hunter, A., & White L. (2010). Analysis and findings. In D.J. Knapp & T.S. Heffner (Eds.). *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267, pp. 29-54). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Allen, M. T., Knapp, D. J., & Owens, K. S. (2013). *Validating future force measures (Army Class): Concluding analyses* (Technical Report in preparation). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Burket, G. R. (1964). A study of reduced rank models for multiple prediction. *Psychometric Monograph* (No. 12).

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* (pp. 307-333). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Chernyshenko, O.S., & Stark, S. (October, 2007). *Criterion validity evidence for narrow temperament clusters: A meta-analysis of military studies*. Paper presented at the 49th annual conference of the International Military Testing Association. Gold Coast, AU.

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* (Technical Report 1168, pp. 49-58). 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.

Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical review of the Armed Services Vocational Aptitude Battery (ASVAB)* (FR-06-25). Alexandria, VA: Human Resources Research Organization.

Drasgow, F., Stark, S., Chernyshenko, O.S., Nye, C.D., Hulin, C.L., & White, L.A. (2012). Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions (Technical Report 1311). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Horst, P. (1954). A technique for the development of a differential predictor battery. *Psychometrika*, 68.

Horst, P. S. (1955). A technique for the development of an absolute prediction battery. *Psychometrika*, 69.

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.

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.

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., & Heffner, T. S. (Eds.) (2012). *Tier One Performance Screen initial operational test and evaluation: 2011 interim report* (Technical Report 1306). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Heffner, T. S., & White, L. (Eds.) (2011). *Tier One Performance Screen initial operational test and evaluation: Early results* (Technical Report 1283). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & LaPort, K. (Eds.) (2013). *Tier One Performance Screen initial operational test and evaluation: 2011 annual report* (Technical Report 1325). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Owens, K. S., & Allen, M. T. (Eds.) (2012). *Validating future force performance measures (Army Class): In-unit performance longitudinal validation* (Technical Report 1314). 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.

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 utility effects of alternative selection and classification policies. In J. Zeidner & C.D. Johnson (Eds.), *The economic benefit of predicting job performance: Vol 5. Estimating the gains of alternative policies* (pp. 73-131). New York: Praeger.

Putka, D. J., Le, H., McCloy, R. A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology*, *93*, 959-981.

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, pp. 135-156). 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, Inc.

Russell, T.L, & Sellman, W.S. (Eds.) (2009). *Development and pilot testing of an information and communications technology literacy test for military enlistees: Volume I technical report* (FR 08-128). Alexandria, VA: Human Resources Research Organization.

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

Stark, S. E., Chernyshenko, O. S., & Dragow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise preference model. *Applied Psychological Measurement, 29*, 184-201.

Stark, S. E., Chernyshenko, O. S., & Dragow, F. (2012). Adaptive testing with multi-unidimensional pairwise preference Items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods*, *15*, 463-487.

Stark, S. E., Chernyshenko, O. S., & Dragow, F. (2010). Tailored adaptive personality assessment system (TAPAS-95s). In D. J. Knapp & T. S. Heffner (Eds.) *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267, pp. 15-22). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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

Trippe, D. M., Caramagno, J. P., Allen, M. T., & Ingerick, M. J. (2011). Initial evidence for the predictive validity and classification potential of the TAPAS. In D. J. Knapp, T. S. Heffner, & L. White (Eds.), *Tier One Performance Screen initial operational test and evaluation: Early results* (Technical Report 1283, pp. 40-53). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Trippe, D. M., Ford, F., Bynum, B., & Moriarty, K. O. (2012). Data file development. In D. J. Knapp & T. S. Heffner (Eds.), *Tier One Performance Screen initial operational test and evaluation: 2011 interim report* (Technical Report 1306, pp. 5-10). Fort Belvoir: U.S. 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* (Technical Report 1168, pp. 89-104). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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.

PREDICTOR MEASURE PSYCHOMETRIC PROPERTIES IN THE APPLICANT SAMPLE

Table A.1. Raw Mean and Standard Deviations for the TOPS Composites and TAPAS Scales on the 13D-CAT Version 1 (June 2009-August 2011)

TAPAS Scale/ TOPS Composite	13D-CAT-1					
	Tier 1 (*n* = 1,339)		Tier 2 (*n* = 48)		Tier 1+ 2 (Combined) (*n* = 1,387)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Individual TAPAS Scales						
Achievement	.230	.495	.287	.466	.232	.494
Adjustment	--	--	--	--	--	--
Attention Seeking	-.224	.556	-.174	.603	-.223	.557
Cooperation	.026	.392	.034	.414	.027	.392
Dominance	.063	.600	.106	.605	.064	.600
Even Tempered	.130	.514	.169	.577	.131	.516
Intellectual Efficiency	.096	.609	.137	.559	.097	.607
Non-Delinquency	.107	.457	-.042	.554	.102	.461
Optimism	.168	.464	.268	.445	.171	.463
Order	-.414	.568	-.285	.439	-.409	.565
Physical Conditioning	-.016	.616	-.088	.602	-.018	.615
Self-Control	--	--	--	--	--	--
Selflessness	-.172	.429	-.190	.460	-.173	.430
Sociability	-.028	.621	-.078	.607	-.030	.620
Tolerance	-.238	.592	-.326	.461	-.241	.588
TOPS Composites (Original)						
Can-Do	112.761	16.828	113.813	17.307	112.797	16.839
Will-Do	105.658	15.768	103.625	16.059	105.588	15.777

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data. Scores on the new TOPS composites were not computed because the 13D-CAT-1 excluded one or more of the requisite TAPAS scales.

Table A.2. Raw Mean and Standard Deviations for the TOPS Composites and TAPAS Scales on the 15D-Static (June 2009-August 2011)

TAPAS Scale/ TOPS Composite	15D-Static					
	Tier 1 (*n* = 13,934)		Tier 2 (*n* = 875)		Tier 1+ 2 (Combined) (*n* = 14,809)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Individual TAPAS Scales						
Achievement	.272	.497	.292	.499	.273	.497
Adjustment	.161	.587	.182	.587	.162	.587
Attention Seeking	-.263	.532	-.259	.523	-.262	.532
Cooperation	-.065	.391	-.078	.398	-.066	.391
Dominance	-.025	.585	-.066	.585	-.027	.585
Even Tempered	.254	.483	.325	.489	.258	.483
Intellectual Efficiency	-.103	.588	-.153	.545	-.106	.586
Non-Delinquency	.126	.453	.047	.483	.122	.456
Optimism	.284	.505	.290	.500	.284	.504
Order	-.402	.576	-.380	.567	-.401	.576
Physical Conditioning	-.035	.617	-.204	.552	-.045	.614
Self-Control	.098	.527	.147	.524	.101	.527
Selflessness	-.184	.449	-.175	.444	-.184	.449
Sociability	-.215	.595	-.220	.566	-.215	.593
Tolerance	-.262	.591	-.258	.586	-.262	.590
TOPS Composites (Original)						
Can-Do	114.249	16.544	113.942	17.284	114.231	16.588
Will-Do	108.163	14.730	106.454	15.113	108.062	14.758
TOPS Composites (Revised)						
Can-Do	100.150	19.946	99.321	19.377	100.101	19.913
Will-Do	100.556	20.098	97.151	19.067	100.355	20.054
Adaptation	100.833	19.734	97.632	18.733	100.644	19.690

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data.

Table A.3. Raw Mean and Standard Deviations for the TOPS Composites and TAPAS Scales on the 15D-CAT Version 1 (June 2009-August 2011)

TAPAS Scale/ TOPS Composite	Tier 1 (*n* = 147,787)		Tier 2 (*n* = 6,166)		Tier 1+ 2 (Combined) (*n* = 153,953)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Individual TAPAS Scales						
Achievement	.150	.482	.186	.482	.152	.482
Adjustment	-.017	.569	.056	.579	-.014	.570
Attention Seeking	-.210	.531	-.240	.523	-.211	.531
Cooperation	-.062	.372	-.072	.375	-.063	.372
Dominance	.028	.588	-.012	.601	.027	.589
Even Tempered	.156	.475	.213	.470	.159	.475
Intellectual Efficiency	-.031	.583	.010	.564	-.029	.582
Non-Delinquency	.099	.459	.025	.485	.096	.460
Optimism	.137	.458	.147	.451	.137	.458
Order	-.422	.544	-.429	.526	-.422	.543
Physical Conditioning	.049	.625	-.085	.602	.044	.625
Self-Control	.061	.529	.130	.537	.064	.530
Selflessness	-.199	.429	-.186	.428	-.199	.429
Sociability	-.045	.592	-.091	.591	-.047	.592
Tolerance	-.228	.567	-.205	.555	-.227	.567
TOPS Composites (Original)						
Can-Do	110.226	16.469	111.037	16.665	110.258	16.478
Will-Do	105.294	14.830	104.421	15.156	105.259	14.844
TOPS Composites (Revised)						
Can-Do	99.731	20.112	102.006	19.143	99.822	20.079
Will-Do	100.551	20.077	98.182	19.639	100.456	20.065
Adaptation	100.672	20.068	99.116	19.243	100.610	20.038

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data.

Table A.4. Raw Mean and Standard Deviations for the TOPS Composites and TAPAS Scales on 15D CAT Version 2, Form A (August 2011-March 2012)

TAPAS Scale/ TOPS Composite	15D-CAT-2A					
	Tier 1 (*n* = 13,134)		Tier 2 (*n* = 1,067)		Tier 1+ 2 (Combined) (*n* = 14,201)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Individual TAPAS Scales						
Achievement	.235	.490	.304	.455	.241	.487
Adjustment	.097	.393	.169	.395	.102	.394
Adventure Seeking [a]	--	--	--	--	--	--
Attention Seeking	-.320	.582	-.263	.615	-.316	.584
Commitment to Serve [a]	--	--	--	--	--	--
Cooperation	.205	.531	.148	.498	.200	.529
Courage [a]	--	--	--	--	--	--
Dominance	.376	.513	.373	.491	.376	.511
Even Tempered	.267	.491	.356	.501	.274	.492
Intellectual Efficiency	.058	.539	.170	.520	.066	.538
Non-Delinquency	.159	.522	.142	.554	.158	.524
Optimism	.297	.449	.277	.467	.295	.450
Order	-.264	.547	-.250	.523	-.263	.545
Physical Conditioning	.166	.562	.082	.537	.159	.560
Responsibility [a]	--	--	--	--	--	--
Self-Control	-.255	.482	-.170	.505	-.249	.484
Selflessness	.049	.430	-.002	.448	.045	.432
Situational Awareness [a]	--	--	--	--	--	--
Sociability	-.174	.563	-.169	.574	-.174	.564
Team Orientation [a]	--	--	--	--	--	--
Tolerance	-.005	.521	-.009	.504	-.005	.520
TOPS Composites (Original)						
Can-Do	108.662	20.00	111.834	20.499	108.900	20.054
Will-Do	108.892	21.283	109.327	21.034	108.925	21.264
TOPS Composites (Revised)						
Can-Do	98.830	19.497	102.866	19.064	99.133	19.493
Will-Do	99.548	20.008	98.522	18.674	99.471	19.913
Adaptation	99.899	20.688	98.681	19.057	99.807	20.572

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data.
[a] Scale not included in this version of the TAPAS.

Table A.5. Raw Mean and Standard Deviations for the TOPS Composites and TAPAS Scales on Version 2, Form B (August 2011-March 2012)

TAPAS Scale/ TOPS Composite	15D-CAT-2B					
	Tier 1 ($n = 26,443$)		Tier 2 ($n = 2,222$)		Tier 1+ 2 (Combined) ($n = 28,665$)	
	M	SD	M	SD	M	SD
Individual TAPAS Scales						
Achievement	.228	.477	.271	.481	.231	.478
Adjustment	.094	.383	.170	.372	.099	.382
Adventure Seeking	-.254	.588	-.219	.577	-.252	.587
Attention Seeking	-.311	.582	-.310	.615	-.311	.585
Commitment to Serve	.157	.516	.259	.487	.165	.514
Cooperation	.173	.520	.157	.524	.172	.520
Courage [a]	--	--	--	--	--	--
Dominance	.330	.498	.287	.515	.326	.499
Even Tempered	.245	.471	.338	.488	.252	.473
Intellectual Efficiency	.035	.521	.102	.516	.041	.521
Non-Delinquency	.147	.526	.147	.564	.147	.529
Optimism	.251	.444	.235	.447	.250	.444
Order	-.241	.539	-.279	.531	-.244	.538
Physical Conditioning	.125	.547	.040	.526	.119	.546
Responsibility [a]	--	--	--	--	--	--
Self-Control [a]	--	--	--	--	--	--
Selflessness	.067	.437	.034	.449	.065	.438
Situational Awareness	.016	.494	.100	.498	.023	.495
Sociability [a]	--	--	--	--	--	--
Team Orientation [a]	--	--	--	--	--	--
Tolerance [a]	--	--	--	--	--	--
TOPS Composites (Original)[b]						
Can-Do	106.983	19.956	109.554	20.599	107.182	20.018
Will-Do	107.445	21.001	108.678	21.503	107.540	21.042

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data.
[a] Scale not included in this version of the TAPAS.
[b] Could not compute new composite scores because of missing scales.

Table A.6. Raw Mean and Standard Deviations for the TOPS Composites and TAPAS Scales on Version 2, Form C (August 2011-March 2012)

TAPAS Scale/ TOPS Composite	15D-CAT-2C					
	Tier 1 (*n* = 26,739)		Tier 2 (*n* = 2,137)		Tier 1+ 2 (Combined) (*n* = 28,876)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Individual TAPAS Scales						
Achievement	.202	.475	.246	.454	.206	.474
Adjustment	.053	.378	.100	.372	.056	.377
Adventure Seeking [a]	--	--	--	--	--	--
Attention Seeking	-.307	.576	-.309	.613	-.307	.579
Commitment to Serve [a]	--	--	--	--	--	--
Cooperation [a]	--	--	--	--	--	--
Courage	.106	.541	.194	.537	.113	.542
Dominance	.321	.490	.283	.490	.318	.490
Even Tempered	.282	.502	.355	.516	.287	.504
Intellectual Efficiency	.037	.531	.110	.511	.042	.530
Non-Delinquency	.157	.529	.126	.546	.155	.531
Optimism	.266	.436	.257	.449	.265	.437
Order [a]	--	--	--	--	--	--
Physical Conditioning	.115	.551	.015	.508	.108	.549
Responsibility	.334	.458	.376	.483	.337	.460
Self-Control	-.243	.453	-.187	.449	-.239	.453
Selflessness [a]	--	--	--	--	--	--
Situational Awareness [a]	--	--	--	--	--	--
Sociability	-.170	.542	-.159	.562	-.169	.543
Team Orientation	-.068	.470	-.052	.491	-.067	.472
Tolerance	-.035	.509	-.020	.514	-.034	.509
TOPS Composites (Original)[b]						
Can-Do	105.577	20.366	109.639	20.387	107.730	20.375
Will-Do	107.624	21.555	107.809	21.724	107.638	21.567

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data.
[a] Scale not included in this version of the TAPAS.
[b] Could not compute new composite scores because of missing scales.

A-7

Table A.7. Correlations among TAPAS Scale Scores on 15D-CAT, Version 1 (June 2009-August 2011)

TAPAS Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Achievement														
2. Adjustment	.09													
3. Attention Seeking	.05	.11												
4. Cooperation	.11	.12	.06											
5. Dominance	.33	.10	.20	.01										
6. Even Tempered	.11	.19	.00	.25	-.05									
7. Intellectual Efficiency	.26	.18	.08	.04	.25	.09								
8. Non-Delinquency	.18	.00	-.13	.17	-.02	.18	.01							
9. Optimism	.19	.27	.17	.17	.17	.18	.10	.08						
10. Order	.15	-.08	-.09	.00	.05	-.02	.02	.10	-.02					
11. Physical Conditioning	.15	.07	.12	-.01	.18	-.07	.05	-.02	.10	.03				
12. Self Control	.22	.07	-.11	.12	.05	.19	.18	.23	.06	.18	-.05			
13. Selflessness	.09	-.02	-.08	.19	.01	.12	-.01	.13	.04	.04	-.04	.08		
14. Sociability	.05	.11	.35	.18	.22	.04	.00	-.04	.23	-.04	.13	-.11	.07	
15. Tolerance	.11	.02	.03	.15	.06	.13	.06	.06	.09	.03	-.06	.11	.32	.12

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid 15D-CAT score data (June 2009-August 2011), *n* = 153,953. Correlations in bold are statistically significant, *p* < .01 (two-tailed).

Table A.8. Correlations among TAPAS Scale Scores on 15D-CAT Version 1 by Education Tier (June 2009-August 2011)

TAPAS Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Achievement		.08	.05	.13	.29	.11	.21	.26	.18	.13	.15	.24	.10	.09	.11
2. Adjustment	.09		.11	.14	.07	.17	.15	.02	.26	-.08	.04	.08	.00	.13	.04
3. Attention Seeking	.05	.11		.09	.17	.00	.07	-.10	.15	-.09	.12	-.11	-.07	.36	.07
4. Cooperation	.11	.12	.06		.02	.25	.04	.20	.20	.00	-.02	.14	.19	.20	.15
5. Dominance	.33	.10	.20	.01		-.06	.20	-.01	.13	.05	.17	.05	-.01	.21	.02
6. Even Tempered	.11	.19	.00	.25	-.05		.07	.22	.20	-.02	-.08	.19	.12	.08	.14
7. Intellectual Efficiency	.26	.18	.08	.04	.25	.09		.03	.09	.04	.06	.20	-.01	.01	.06
8. Non-Delinquency	.17	.00	-.13	.17	-.02	.17	.01		.13	.11	-.03	.25	.18	-.01	.12
9. Optimism	.19	.27	.17	.16	.17	.18	.10	.08		-.05	.08	.06	.09	.22	.13
10. Order	.15	-.08	-.09	.00	.05	-.02	.02	.10	-.02		.06	.19	.04	-.05	.01
11. Physical Conditioning	.15	.07	.12	-.01	.18	-.07	.05	-.02	.10	.03		-.02	-.03	.12	-.04
12. Self Control	.22	.07	-.11	.12	.05	.19	.18	.23	.06	.18	-.05		.08	-.07	.09
13. Selflessness	.08	-.02	-.08	.19	.01	.12	-.02	.13	.04	.04	-.05	.08		.06	.31
14. Sociability	.05	.11	.35	.18	.22	.03	.00	-.04	.23	-.04	.13	-.11	.07		.13
15. Tolerance	.11	.02	.03	.15	.06	.13	.06	.06	.09	.03	-.06	.11	.32	.11	

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid 15D-CAT score data (June 2009-August 2011). Correlations below the diagonal are for Education Tier 1 applicants, $n = 147,787$. Correlations above the diagonal are for Education Tier 2 applicants, $n = 6,166$. Correlations in bold are statistically significant, $p < .01$ (two-tailed).

Table A.9. Correlations among TAPAS Scale Scores on the 15D-CAT Version 2 Forms (August 2011-March 2012)

TAPAS Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Achievement															
2. Adjustment	**.10**														
3. Adventure Seeking	**.10**	**.16**													
4. Attention Seeking	**.04**	**.10**	**.17**												
5. Commitment to Serve	**.12**	**.05**	**.04**	-.01											
6. Cooperation	**.08**	-.01	**-.13**	**-.14**	**.04**										
7. Courage	**.21**	**.15**	--	**.10**	--	--									
8. Dominance	**.28**	**.09**	**.13**	**.25**	**.06**	**-.12**	**.22**								
9. Even Tempered	**.09**	**.23**	**-.05**	**-.07**	**.05**	**.32**	**.03**	**-.06**							
10. Intellectual Efficiency	**.24**	**.19**	**.07**	**.10**	.01	**-.05**	**.18**	**.26**	**.09**						
11. Non-Delinquency	**.19**	**.02**	**-.18**	**-.16**	**.12**	**.31**	**.03**	**-.03**	**.26**	**.04**					
12. Optimism	**.12**	**.23**	**.03**	**.05**	.01	**.11**	**.03**	**.09**	**.17**	**.10**	**.14**				
13. Order	**.18**	**-.06**	**-.10**	**-.01**	**.06**	**.13**	--	**.07**	**.02**	**.07**	**.15**	**.02**			
14. Physical Conditioning	**.18**	**.05**	**.25**	**.09**	.01	**-.08**	**.12**	**.17**	**-.09**	**.05**	**-.07**	**.02**	**.04**		
15. Responsibility	**.32**	**.12**	--	**-.05**	--	--	**.14**	**.16**	**.15**	**.17**	**.23**	**.15**	--	**.05**	
16. Self-Control	**.25**	**.11**	--	**-.08**	--	**.21**	**.09**	**.06**	**.24**	**.18**	**.29**	**.13**	**.22**	-.01	**.23**
17. Selflessness	**.14**	**-.07**	-.03	**-.05**	**.04**	**.22**	--	**.07**	**.10**	.00	**.18**	**.08**	**.08**	-.01	--
18. Situational Awareness	**.19**	**.15**	**.10**	**.04**	**.07**	.00	--	**.12**	**.11**	**.26**	**.11**	**.09**	**.15**	**.05**	--
19. Sociability	**.08**	**.07**	--	**.31**	--	**.02**	**.08**	**.21**	.01	**.06**	-.01	**.10**	-.02	.02	.04
20. Team Orientation	**.07**	**.04**	--	**.12**	--	--	**.03**	**.10**	**.06**	**-.04**	**.05**	**.06**	--	**.06**	.03
21. Tolerance	**.07**	**.04**	--	**.03**	--	**.07**	**.03**	**.02**	**.13**	**.13**	**.06**	**.09**	.02	**-.04**	**.08**

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data (August 2011-March 2012), n = 14,201-71,742. Correlations in bold are statistically significant, $p < .01$ (two-tailed). Missing values reflect scales that do not appear on all TAPAS versions.

Table A.9. Correlations among TAPAS Scale Scores on the 15D-CAT Version 2 Forms (August 2011-March 2012) (cont'd)

TAPAS Scale		16	17	18	19	20
1.	Achievement					
2.	Adjustment					
3.	Adventure Seeking					
4.	Attention Seeking					
5.	Commitment to Serve					
6.	Cooperation					
7.	Courage					
8.	Dominance					
9.	Even Tempered					
10.	Intellectual Efficiency					
11.	Non-Delinquency					
12.	Optimism					
13.	Order					
14.	Physical Conditioning					
15.	Responsibility					
16.	Self-Control					
17.	Selflessness	**.11**				
18.	Situational Awareness	--	**.04**			
19.	Sociability	**-.02**	**.13**			
20.	Team Orientation	**.06**	--	--	**.22**	
21.	Tolerance	**.10**	**.21**	--	**.14**	**.08**

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data (August 2011-March 2012), n = 14,201-71,742. Correlations in bold are statistically significant, $p < .01$ (two-tailed). Missing values reflect scales that do not appear on all TAPAS versions.

Table A.10. Correlations among TAPAS Scale Scores on the 15D-CAT Version 2 Forms by Education Tier (August 2011-March 2012)

TAPAS Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Achievement		**.09**	**.08**	**.05**	**.14**	**.08**	**.22**	**.26**	**.08**	**.23**	**.23**	**.09**	**.21**	**.17**	**.31**
2. Adjustment	**.10**		**.11**	**.11**	**.06**	**.04**	**.12**	**.07**	**.23**	**.16**	**.07**	**.24**	-.03	.01	**.15**
3. Adventure Seeking	**.10**	**.16**		**.19**	**.08**	**-.12**	--	**.13**	**-.08**	.05	**-.17**	.03	-.02	**.23**	--
4. Attention Seeking	**.04**	**.10**	**.17**		.05	**-.13**	**.14**	**.27**	**-.05**	**.08**	**-.11**	**.07**	.03	**.09**	-.04
5. Commitment to Serve	**.11**	**.04**	**.03**	-.01		.03	--	.05	**.06**	.03	**.10**	.01	.04	.05	--
6. Cooperation	**.08**	-.01	**-.14**	**-.14**	**.04**		--	**-.15**	**.31**	-.01	**.31**	**.12**	**.09**	**-.07**	--
7. Courage	**.21**	**.15**	--	**.10**	--	--		**.22**	.00	**.20**	**.08**	**.06**	--	**.12**	**.21**
8. Dominance	**.28**	**.09**	**.13**	**.25**	**.06**	**-.12**	**.22**		**-.05**	**.26**	-.01	**.09**	**.09**	**.15**	**.18**
9. Even Tempered	**.09**	**.23**	-.04	-.07	**.05**	**.32**	**.03**	**-.05**		**.09**	**.29**	**.19**	.02	**-.11**	**.18**
10. Intellectual Efficiency	**.24**	**.19**	**.07**	**.10**	.01	**-.05**	**.18**	**.27**	**.09**		**.08**	**.08**	**.06**	**.05**	**.18**
11. Non-Delinquency	**.19**	**.01**	**-.18**	**-.16**	**.12**	**.31**	**.02**	**-.03**	**.26**	**.03**		**.15**	**.15**	-.04	**.30**
12. Optimism	**.13**	**.23**	**.03**	**.04**	.01	**.11**	**.03**	**.09**	**.17**	**.10**	**.13**		-.02	-.01	**.14**
13. Order	**.18**	**-.06**	**-.10**	-.02	**.06**	**.13**	--	**.07**	.02	**.07**	**.15**	.02		**.13**	--
14. Physical Conditioning	**.19**	**.05**	**.25**	**.09**	.01	**-.09**	**.12**	**.17**	**-.09**	**.05**	**-.08**	.02	.03		**.06**
15. Responsibility	**.32**	**.12**	--	**-.05**	--	--	**.14**	**.16**	**.15**	**.17**	**.22**	**.15**	--	**.05**	
16. Self-Control	**.25**	**.10**	--	**-.08**	--	**.21**	**.09**	**.06**	**.24**	**.18**	**.29**	**.13**	**.23**	-.01	**.22**
17. Selflessness	**.14**	**-.07**	-.03	**-.05**	.04	**.22**	--	**.07**	**.10**	.00	**.18**	**.08**	**.08**	-.01	--
18. Situational Awareness	**.19**	**.15**	**.10**	**.05**	**.07**	.00	--	**.12**	**.10**	**.26**	**.11**	**.09**	**.15**	**.05**	--
19. Sociability	**.09**	**.07**	--	**.31**	--	**.02**	**.08**	**.21**	.01	**.06**	-.01	**.10**	-.01	.02	**.04**
20. Team Orientation	**.07**	**.03**	--	**.12**	--	--	**.03**	**.10**	**.05**	-.04	**.05**	**.06**	--	**.06**	**.03**
21. Tolerance	**.07**	**.04**	--	**.03**	--	**.08**	**.04**	**.02**	**.13**	**.13**	**.06**	**.09**	.02	-.04	**.08**

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data (August 2011-March 2012). Correlations below the diagonal are for Education Tier 1 applicants, *n* = 13,134-66,316. Correlations above the diagonal are for Education Tier 2 applicants, *n* = 1,067-5,426. Correlations in bold are statistically significant, *p* < .01 (two-tailed). Missing values reflect scales that do not appear on all TAPAS versions.

Table A.10. Correlations among TAPAS Scale Scores on the 15D-CAT Version 2 Forms by Education Tier (August 2011-March 2012) (cont'd)

	TAPAS Scale	16	17	18	19	20	21
1.	Achievement	**.23**	**.14**	**.21**	.06	**.10**	.05
2.	Adjustment	**.15**	**-.06**	**.14**	.07	.07	.06
3.	Adventure Seeking	--	-.03	**.11**	--	--	--
4.	Attention Seeking	**-.07**	-.02	.02	**.33**	**.12**	.02
5.	Commitment to Serve	--	.03	**.08**	--	--	--
6.	Cooperation	**.13**	**.24**	-.01	.04	--	.06
7.	Courage	**.12**	--	--	**.11**	.04	.00
8.	Dominance	.05	.01	**.14**	**.21**	**.14**	.02
9.	Even Tempered	**.26**	**.11**	**.11**	.02	**.06**	**.09**
10.	Intellectual Efficiency	**.19**	-.02	**.25**	**.05**	-.02	**.08**
11.	Non-Delinquency	**.33**	**.19**	**.10**	-.01	**.11**	**.09**
12.	Optimism	**.14**	**.09**	**.06**	**.12**	.03	**.09**
13.	Order	**.19**	**.11**	**.16**	-.06	--	.00
14.	Physical Conditioning	.00	-.01	**.06**	.02	.04	**-.05**
15.	Responsibility	**.24**	--	--	**.06**	**.06**	**.07**
16.	Self-Control		.06	--	**-.05**	**.09**	**.08**
17.	Selflessness	**.12**		.05	**.18**	--	**.20**
18.	Situational Awareness	--	**.04**		--	--	--
19.	Sociability	**-.02**	**.13**	--		**.22**	**.12**
20.	Team Orientation	**.05**	--	--	**.22**		**.10**
21.	Tolerance	**.10**	**.21**	--	**.15**	**.08**	

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data (August 2011-March 2012). Correlations below the diagonal are for Education Tier 1 applicants, n = 13,134-66,316. Correlations above the diagonal are for Education Tier 2 applicants, n = 1,067-5,426. Correlations in bold are statistically significant, $p < .01$ (two-tailed). Missing values reflect scales that do not appear on all TAPAS versions.

Table A.11. Correlations between TOPS Composites and TAPAS Scales with AFQT by Version

TOPS Composite/ TAPAS Scale	15D-Static, 15D-CAT Version 1 (June 2009-August 2011)			Version 2 Forms (August 2011-March 2012)		
	Tier 1	Tier 2	*Tier 1+ 2 (Combined)*	Tier 1	Tier 2	*Tier 1+ 2 (Combined)*
N	161,721	7,041	168,762	13,134-66,316	1,067-5,426	14,201-71,742
Individual TAPAS Scales						
Achievement	**.10**	**.05**	**.09**	**.04**	.00	**.04**
Adjustment	**.11**	**.11**	**.11**	**.12**	**.10**	**.12**
Adventure Seeking	--	--	--	**.12**	.03	**.11**
Attention Seeking	**.11**	**.08**	**.11**	.01	-.02	.01
Commitment to Serve	--	--	--	**-.15**	**-.10**	**-.14**
Cooperation	.00	.01	.00	**-.14**	**-.09**	**-.13**
Courage	--	--	--	**.06**	**.07**	**.06**
Dominance	**.09**	-.01	**.08**	**.13**	.02	**.12**
Even Tempered	**.09**	**.09**	**.09**	**.09**	**.10**	**.09**
Intellectual Efficiency	**.42**	**.37**	**.41**	**.32**	**.30**	**.32**
Non-Delinquency	-.01	.03	.00	**-.07**	-.01	**-.06**
Optimism	**.02**	.01	**.02**	**.09**	**.06**	**.09**
Order	**-.18**	**-.17**	**-.18**	**-.17**	**-.17**	**-.17**
Physical Conditioning	**.05**	-.03	**.04**	**.06**	-.02	**.06**
Responsibility	--	--	--	**.14**	**.07**	**.14**
Self-Control	-.01	**.04**	-.01	**-.05**	-.01	**-.05**
Selflessness	**-.07**	**-.05**	**-.07**	**-.08**	**-.09**	**-.08**
Situational Awareness	--	--	--	.01	.03	.01
Sociability	**-.08**	**-.07**	**-.08**	**-.12**	**-.10**	**-.12**
Team Orientation	--	--	--	**-.11**	**-.11**	**-.11**
Tolerance	**-.01**	.01	**-.01**	**.08**	**.05**	**.08**
TOPS Composites (Original)						
Can-Do	**.23**	**.19**	**.23**	**.15**	**.14**	**.15**
Will-Do	**.05**	.02	**.05**	**.04**	**.04**	**.04**
TOPS Composites (Revised)						
Can-Do	**.45**	**.40**	**.45**	**.38**	**.34**	**.38**
Will-Do	**.10**	.00	**.09**	**.12**	.01	**.11**
Adaptation	**.19**	**.12**	**.19**	**.21**	**.15**	**.21**

Note. Correlations in bold are statistically significant, *p* < .01 (two-tailed). Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS score data. Missing values reflect scales that do not appear on all TAPAS versions.

Table A.12. Descriptive Statistics for AFQT, ASVAB Subtests, and Aptitude Area (AA) Composites in the TOPS Applicant Sample by Version

Subtest/Composite	15D-Static, 15D-CAT Version 1 (June 2009-August 2011)					Version 2 Forms (August 2011-March 2012)				
	n	*M*	*SD*	*Min*	*Max*	*n*	*M*	*SD*	*Min*	*Max*
AFQT	168,762	57.09	23.06	10	99	71,742	53.87	21.98	10	99
ASVAB Subtests										
Arithmetic Reasoning (AR)	167,895	52.57	7.69	18	72	71,410	51.41	7.55	23	72
Assembling Objects (AO)	165,321	55.19	7.79	25	70	67,643	54.27	7.83	26	70
Auto & Shop Information (AS)	167,894	49.95	9.39	19	86	71,409	48.27	9.04	22	82
Electronics Information (EI)	167,894	51.99	9.07	16	84	71,409	50.66	8.83	16	84
General Science (GS)	167,895	51.70	8.40	19	76	71,410	50.81	8.17	20	76
Math Knowledge (MK)	167,895	53.45	6.97	24	73	71,410	52.99	6.71	26	73
Mechanical Comprehension (MC)	167,894	53.44	8.41	14	82	71,408	52.16	8.21	23	82
Paragraph Comprehension (PC)	167,895	52.86	7.05	21	69	71,410	52.06	6.84	23	69
Word Knowledge (WK)	167,895	51.33	8.05	16	76	71,410	50.22	7.68	16	76
Aptitude Area (AA) Composites										
Clerical (CL)	167,896	105.87	13.92	35	152	71,408	103.65	13.29	56	152
Combat (CO)	167,896	105.68	14.85	29	160	71,408	103.06	14.20	54	159
Electronics (EL)	167,896	105.47	14.85	29	160	71,408	102.77	14.19	54	159
Field Artillery (FA)	167,896	105.85	14.78	28	159	71,408	103.24	14.14	55	159
General Maintenance (GM)	167,896	105.17	15.32	28	161	71,408	102.36	14.68	54	160
General Technical (GT)	167,897	104.91	14.33	39	149	71,410	102.55	13.74	49	149
Mechanical Maintenance (MM)	167,896	104.28	16.31	25	165	71,408	101.10	15.65	51	163
Operators and Food Service (OF)	167,896	105.17	15.28	27	160	71,408	102.30	14.64	55	160
Signal Communications (SC)	167,896	105.87	14.48	29	159	71,408	103.35	13.81	54	158
Skilled Technical (ST)	167,896	105.73	14.49	32	157	71,408	103.23	13.83	55	156

Note. Results are limited to the TOPS Applicant Sample (non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS scores.

Table A.13. Descriptive Statistics for AFQT, ASVAB Subtests, and Aptitude Area (AA) Composites in the Pre-August 2011 TOPS Applicant Sample by Education Tier

Composite/Subtest	Tier 1					Tier 2				
	n	*M*	*SD*	*Min*	*Max*	*n*	*M*	*SD*	*Min*	*Max*
AFQT	161,721	57.21	23.21	10	99	7,041	54.27	19.04	10	99
ASVAB Subtests										
Arithmetic Reasoning (AR)	160,867	52.60	7.74	18	72	7,028	51.84	6.37	24	72
Assembling Objects (AO)	158,410	55.21	7.80	25	70	6,911	54.79	7.52	26	69
Auto & Shop Information (AS)	160,866	49.85	9.40	19	86	7,028	52.30	8.84	26	81
Electronics Information (EI)	160,866	51.95	9.12	16	84	7,028	53.07	7.81	18	82
General Science (GS)	160,867	51.70	8.45	19	76	7,028	51.67	6.99	23	74
Math Knowledge (MK)	160,867	53.61	6.98	24	73	7,028	49.75	5.69	28	73
Mechanical Comprehension (MC)	160,866	53.42	8.45	14	82	7,028	53.91	7.51	23	79
Paragraph Comprehension (PC)	160,867	52.84	7.09	23	69	7,028	53.17	6.18	21	69
Word Knowledge (WK)	160,867	51.31	8.11	16	76	7,028	51.84	6.67	22	76
Aptitude Area (AA) Composites										
Clerical (CL)	160,868	105.96	14.03	35	152	7,028	103.77	10.91	56	143
Combat (CO)	160,868	105.73	14.97	29	160	7,028	104.60	11.91	51	153
Electronics (EL)	160,868	105.49	14.97	29	160	7,028	104.88	11.86	52	151
Field Artillery (FA)	160,868	105.90	14.89	28	159	7,028	104.67	11.82	51	152
General Maintenance (GM)	160,868	105.18	15.43	28	161	7,028	104.90	12.42	48	154
General Technical (GT)	160,869	104.92	14.44	39	149	7,028	104.63	11.46	54	145
Mechanical Maintenance (MM)	160,868	104.22	16.41	25	165	7,028	105.65	13.83	46	155
Operators and Food Service (OF)	160,868	105.17	15.39	27	160	7,028	105.25	12.40	50	152
Signal Communications (SC)	160,868	105.93	14.59	29	159	7,028	104.48	11.44	54	150
Skilled Technical (ST)	160,868	105.78	14.61	32	157	7,028	104.70	11.45	56	148

Note. Results are limited to the TOPS Applicant Sample (non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid scores on the pre-August 2011 versions of the TAPAS (15D-Static or 15D-CAT).

Table A.14. Descriptive Statistics for AFQT, ASVAB Subtests, and Aptitude Area (AA) Composites in the Post-August 2011 TOPS Applicant Sample by Education Tier

Composite/Subtest	Tier 1					Tier 2				
	n	*M*	*SD*	*Min*	*Max*	*n*	*M*	*SD*	*Min*	*Max*
AFQT	66,316	53.81	22.23	10	99	5,426	54.65	18.63	10	99
ASVAB Subtests										
Arithmetic Reasoning (AR)	65,995	51.38	7.63	23	72	5,415	51.82	6.44	27	72
Assembling Objects (AO)	62,465	54.25	7.85	26	70	5,178	54.57	7.52	27	69
Auto & Shop Information (AS)	65,994	48.02	9.01	22	82	5,415	51.27	8.91	23	82
Electronics Information (EI)	65,994	50.49	8.89	16	84	5,415	52.65	7.89	21	80
General Science (GS)	65,995	50.74	8.25	20	76	5,415	51.59	7.09	26	76
Math Knowledge (MK)	65,995	53.22	6.74	26	73	5,415	50.13	5.57	30	73
Mechanical Comprehension (MC)	65,993	52.05	8.25	23	82	5,415	53.55	7.58	24	78
Paragraph Comprehension (PC)	65,995	51.97	6.89	23	69	5,415	53.27	6.06	27	69
Word Knowledge (WK)	65,995	50.10	7.74	16	76	5,415	51.70	6.70	21	76
Aptitude Area (AA) Composites										
Clerical (CL)	65,993	103.63	13.47	56	152	5,415	103.82	10.79	71	148
Combat (CO)	65,993	102.97	14.37	54	159	5,415	104.15	11.97	63	155
Electronics (EL)	65,993	102.64	14.35	54	159	5,415	104.42	11.90	65	156
Field Artillery (FA)	65,993	103.16	14.30	55	159	5,415	104.28	11.87	64	155
General Maintenance (GM)	65,993	102.21	14.83	54	160	5,415	104.29	12.53	61	157
General Technical (GT)	65,995	102.39	13.91	49	149	5,415	104.56	11.37	65	145
Mechanical Maintenance (MM)	65,993	100.82	15.75	51	163	5,415	104.57	14.02	59	161
Operators and Food Service (OF)	65,993	102.11	14.79	55	160	5,415	104.61	12.52	64	157
Signal Communications (SC)	65,993	103.28	13.99	54	158	5,415	104.23	11.41	66	154
Skilled Technical (ST)	65,993	103.13	14.00	55	156	5,415	104.45	11.43	68	152

Note. Results are limited to the TOPS Applicant Sample (non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid scores on the post-August 2011 forms of the TAPAS (15D-CAT Version 2).

Table A.15. Correlations among AFQT, ASVAB Subtests, and Aptitude Area (AA) Composite Scores in Pre-August 2011 TOPS Applicant Sample

Subtest/Composite	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. AFQT																		
ASVAB Subtests																		
2. Arithmetic Reasoning (AR)	.82																	
3. Assembling Objects (AO)	.44	.47																
4. Auto & Shop Info (AS)	.37	.31	.26															
5. Electronics Information (EI)	.60	.48	.36	.69														
6. General Science (GS)	.74	.55	.36	.51	.69													
7. Math Knowledge (MK)	.72	.69	.38	.06	.29	.43												
8. Mechanical Comp (MC)	.65	.60	.52	.62	.69	.67	.41											
9. Paragraph Comp (PC)	.80	.56	.35	.36	.53	.65	.42	.55										
10. Word Knowledge (WK)	.81	.48	.28	.41	.59	.73	.35	.55	.71									
Aptitude Area (AA) Composites																		
11. Clerical (CL)	.96	.90	.50	.43	.64	.74	.78	.72	.76	.74								
12. Combat (CO)	.88	.79	.51	.67	.79	.81	.67	.86	.71	.71	.94							
13. Electronics (EL)	.90	.80	.50	.67	.81	.81	.64	.83	.74	.75	.95	.99						
14. Field Artillery (FA)	.89	.82	.52	.65	.77	.80	.68	.86	.71	.71	.95	1.00	.99					
15. General Maintenance (GM)	.85	.78	.50	.73	.83	.81	.60	.85	.69	.70	.91	.99	.99	.99				
16. Mechanical Maintenance (MM)	.96	.88	.46	.41	.62	.74	.63	.68	.81	.82	.97	.88	.91	.90	.87			
17. Operators & Food (OF)	.74	.67	.46	.86	.85	.75	.45	.86	.63	.65	.81	.95	.95	.94	.97	.77		
18. Signal Communications (SC)	.86	.79	.50	.72	.81	.80	.59	.86	.71	.72	.92	.99	.99	.99	1.00	.89	.97	
19. Skilled Technical (ST)	.92	.83	.51	.60	.78	.79	.71	.81	.74	.75	.97	.99	.99	.99	.98	.93	.92	.98

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid scores on the pre-August 2011 versions of the TAPAS (15D-Static or 15D-CAT Version 1), n = 165,732 – 168,762. All correlations are statistically significant, $p < .01$ (one-tailed). Because scores on the AA composites tend to be highly correlated (.95 and above), correlations of 1.00 are computationally possible when the correlations are rounded to two decimal places.

Table A.16. Correlations among AFQT, ASVAB Subtests, and Aptitude Area (AA) Composite Scores in Post-August 2011 TOPS Applicant Sample

Subtest/Composite	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. AFQT																		
ASVAB Subtests																		
2. Arithmetic Reasoning (AR)	.82																	
3. Assembling Objects (AO)	.45	.47																
4. Auto & Shop Info (AS)	.37	.31	.27															
5. Electronics Information (EI)	.58	.46	.35	.66														
6. General Science (GS)	.72	.52	.36	.50	.67													
7. Math Knowledge (MK)	.71	.67	.38	.06	.27	.40												
8. Mechanical Comp (MC)	.64	.58	.51	.61	.67	.65	.39											
9. Paragraph Comp (PC)	.78	.53	.34	.34	.51	.62	.39	.52										
10. Word Knowledge (WK)	.80	.46	.28	.40	.58	.70	.32	.53	.67									
Aptitude Area (AA) Composites																		
11. Clerical (CL)	.97	.90	.50	.43	.63	.71	.76	.71	.73	.73								
12. Combat (CO)	.88	.78	.51	.67	.78	.80	.66	.85	.68	.70	.93							
13. Electronics (EL)	.90	.79	.50	.67	.80	.79	.63	.82	.71	.74	.95	.99						
14. Field Artillery (FA)	.89	.81	.52	.65	.76	.78	.67	.85	.68	.69	.95	1.00	.99					
15. General Maintenance (GM)	.85	.77	.50	.73	.81	.80	.59	.84	.66	.68	.91	.99	.99	.99				
16. Mechanical Maintenance (MM)	.96	.88	.46	.41	.60	.71	.60	.66	.78	.80	.97	.88	.91	.89	.86			
17. Operators & Food (OF)	.74	.66	.47	.86	.83	.74	.44	.85	.60	.64	.81	.95	.95	.94	.97	.77		
18. Signal Communications (SC)	.86	.79	.51	.72	.79	.79	.57	.85	.68	.71	.92	.99	.99	.99	1.00	.88	.97	
19. Skilled Technical (ST)	.92	.82	.51	.60	.77	.77	.69	.80	.72	.73	.97	.99	.99	.99	.98	.92	.91	.98

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid scores on the post-August 2011 forms of the TAPAS (15D-CAT Version 2), $n = 67,643 – 71,742$. All correlations are statistically significant, $p < .01$ (one-tailed). Because scores on the AA composites tend to be highly correlated (.95 and above), correlations of 1.00 are computationally possible when the correlations are rounded to two decimal places.

Table A.17. Standardized Group Means on the AFQT, TAPAS Scales, and TOPS Composites by Gender and Version

Measure/Scale	Female		Male	
	15D-Static, 15D-CAT Version 1	15D-CAT Version 2	15D-Static, 15D-CAT Version 1	15D-CAT Version 2
N	33,652	2,916 - 14,565	134,467	11,013 - 55,784
AFQT	52.40	49.16	58.29	55.23
Individual TAPAS Scales				
Achievement	.02	.05	-.01	-.04
Adjustment	-.26	-.27	.04	.05
Adventure Seeking	--	-.25	--	.07
Attention Seeking	-.12	-.08	.00	-.03
Commitment to Serve	--	.07	--	-.02
Cooperation	-.02	.20	.01	-.05
Courage	--	-.27	--	.03
Dominance	-.11	.02	.01	-.02
Even Tempered	-.09	.01	.01	.00
Intellectual Efficiency	-.17	-.09	.03	-.02
Non-Delinquency	.15	.17	.00	-.02
Optimism	-.02	.08	.01	-.02
Order	.13	.20	-.03	-.10
Physical Conditioning	-.28	-.27	.12	.07
Responsibility	--	.07	--	-.05
Self-Control	.01	.05	-.01	-.04
Selflessness	.30	.41	-.08	-.10
Situational Awareness	--	-.23	--	.01
Sociability	.00	.12	-.01	-.02
Team Orientation	--	-.13	--	.01
Tolerance	.26	.22	-.08	-.07
TOPS Composites (Original)				
Can-Do	109.83	109.64	110.79	107.17
Will-Do	104.39	108.31	105.78	107.69
TOPS Composites (Revised)				
Can-Do	95.01	94.45	101.06	100.42
Will-Do	95.83	96.19	101.60	100.35
Adaptation	93.57	92.31	102.38	101.81

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS scores. Missing values reflect scales that do not appear on all TAPAS versions.

Table A.18. Standardized Group Means on the AFQT, TAPAS Scales, and TOPS Composites by Race-Ethnicity and Version

Measure/Scale	Black, Non-Hispanic		Hispanic		White, Non-Hispanic	
	15D-Static, 15D-CAT Version 1	*15D-CAT Version 2*	*15D-Static, 15D-CAT Version 1*	*15D-CAT Version 2*	*15D-Static, 15D-CAT Version 1*	*15D-CAT Version 2*
N	26,282	3,049-15,266	22,009	2,097-10,278	96,931	8,124-41,261
AFQT	45.30	43.28	47.32	47.65	62.65	59.42
Individual TAPAS Scales						
Achievement	-.05	-.05	-.07	-.10	.04	.02
Adjustment	-.09	-.12	-.13	-.10	.04	.07
Adventure Seeking	--	-.48	--	-.02	--	.19
Attention Seeking	-.08	.13	-.05	-.04	.00	-.10
Commitment to Serve	--	.01	--	-.01	--	.00
Cooperation	.02	.18	-.04	.02	.00	-.06
Courage	--	-.19	--	-.09	--	.06
Dominance	.02	.04	-.01	-.06	.00	.00
Even Tempered	-.01	-.01	-.09	-.04	.02	.03
Intellectual Efficiency	-.11	.00	-.13	-.10	.05	-.02
Non-Delinquency	.09	.10	-.02	.00	.04	.01
Optimism	.04	.01	.00	-.02	.02	.03
Order	.19	.21	.13	.05	-.09	-.16
Physical Conditioning	-.12	-.13	-.01	-.02	.11	.05
Responsibility	--	-.05	--	-.19	--	.05
Self-Control	.17	.17	.06	.04	-.06	-.10
Selflessness	.10	.19	.03	-.02	-.05	-.06
Situational Awareness	--	-.03	--	-.09	--	-.03
Sociability	-.05	.07	-.01	-.02	.01	.01
Team Orientation	--	-.03	--	.04	--	-.05
Tolerance	.16	.07	.20	.09	-.13	-.09
TOPS Composites (Original)						
Can-Do	110.28	108.41	108.56	106.08	111.63	108.34
Will-Do	104.83	105.84	104.00	106.56	106.38	109.20
TOPS Composites (Revised)						
Can-Do	96.35	96.91	95.68	96.52	102.23	101.01
Will-Do	98.36	97.40	99.27	98.27	101.82	100.94
Adaptation	96.65	95.15	97.88	97.56	102.79	102.27

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS scores. Missing values reflect scales that do not appear on all TAPAS versions.

Table A.19. Standardized Group Means on the AFQT and TAPAS by Education Tier and Version in the TOPS Applicant Sample

Measure/ Scale	Tier 1		Tier 2	
	15D-Static, 15D-CAT Version 1	*15D-CAT Version 2*	*15D-Static, 15D-CAT Version 1*	*15D-CAT Version 2*
N	161,721	13,134 - 66,316	7,041	1,067 - 5,426
AFQT	57.21	53.81	54.27	54.65
Individual TAPAS Scales				
Achievement	-.01	-.03	.06	.07
Adjustment	-.02	-.02	.10	.14
Adventure Seeking	--	-.01	--	.05
Attention Seeking	-.02	-.04	-.07	-.02
Commitment to Serve	--	-.02	--	.18
Cooperation	.00	.01	-.02	-.04
Courage	--	-.04	--	.11
Dominance	-.01	-.01	-.08	-.07
Even Tempered	-.01	-.01	.11	.17
Intellectual Efficiency	-.01	-.04	.04	.10
Non-Delinquency	.04	.02	-.12	.00
Optimism	.01	.01	.03	-.02
Order	.00	-.04	-.01	-.07
Physical Conditioning	.05	.01	-.17	-.15
Responsibility	--	-.03	--	.06
Self-Control	-.01	-.03	.11	.11
Selflessness	-.01	.01	.02	-.08
Situational Awareness	--	-.06	--	.11
Sociability	-.01	.01	-.07	.03
Team Orientation	--	-.02	--	.01
Tolerance	-.01	-.01	.02	.01
TOPS Composites (Original)				
Can-Do	110.57	107.56	111.40	110.04
Will-Do	105.54	107.80	104.67	108.46
TOPS Composites (Revised)				
Can-Do	99.77	98.83	101.67	102.87
Will-Do	100.55	99.55	98.05	98.52
Adaptation	100.69	99.90	98.93	98.68

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS scores. Missing values reflect scales that do not appear on all TAPAS versions.

Table A.20. Standardized Group Mean Score Differences on the AFQT, TAPAS Scales, and TOPS Composites by Version

Measure/Scale	Female-Male		Black, Non-Hispanic-White, Non-Hispanic		Hispanic-White, Non-Hispanic		Tier 1-Tier2	
	15D-Static, 15D-CAT	*15D-CAT Version 2*	*15D-Static, 15D-CAT*	*15D-CAT Version 2*	*15D-Static, 15D-CAT*	*15D-CAT Version 2*	*15D-Static, 15D-CAT*	*15D-CAT Version 2*
N	168,119	13,929 - 70,349	123,213	11,173 - 56,527	118,940	10,221 - 51,539	168,762	14,201 - 71,742
AFQT	**-.25**	**-.27**	**-.79**	**-.74**	**-.70**	**-.54**	**.13**	**-.04**
Individual TAPAS Scales								
Achievement	**.03**	**.09**	**-.09**	**-.07**	**-.10**	**-.12**	**-.07**	**-.10**
Adjustment	**-.31**	**-.33**	**-.13**	**-.19**	**-.17**	**-.17**	**-.12**	**-.17**
Adventure Seeking	--	**-.33**	--	**-.69**	--	**-.22**	--	**-.06**
Attention Seeking	**-.12**	**-.05**	**-.09**	**.23**	**-.05**	**.06**	.05	-.02
Commitment to Serve	--	**.09**	--	.01	--	-.01	--	**-.20**
Cooperation	**-.02**	**.25**	**.02**	**.25**	**-.04**	**.09**	.03	**.06**
Courage	--	**-.31**	--	**-.25**	--	**-.15**	--	**-.16**
Dominance	**-.12**	**.04**	**.02**	**.04**	.00	**-.05**	.07	.07
Even Tempered	**-.10**	.00	**-.03**	**-.04**	**-.11**	**-.07**	**-.13**	**-.18**
Intellectual Efficiency	**-.20**	**-.08**	**-.16**	.02	**-.18**	**-.08**	**-.05**	**-.15**
Non-Delinquency	**.16**	**.20**	**.05**	**.10**	**-.07**	.00	.16	.03
Optimism	**-.04**	**.10**	**.02**	-.01	-.02	**-.04**	-.03	.03
Order	**.16**	**.31**	**.29**	**.38**	.22	.22	.01	.04
Physical Conditioning	**-.41**	**-.35**	**-.23**	**-.18**	**-.12**	**-.07**	.22	**.16**
Responsibility	--	**.12**	--	**-.10**	--	**-.25**	--	**-.08**
Self-Control	.01	**.09**	**.23**	**.27**	**.12**	**.14**	**-.13**	**-.14**
Selflessness	**.40**	**.54**	**.15**	**.25**	**.07**	**.04**	-.03	**.09**
Situational Awareness	--	**-.24**	--	.00	--	**-.05**	--	**-.17**

A-23

Sociability	.01	**.14**	**-.06**	**.06**	-.01	-.02	**.07**	-.01
Team Orientation	--	**-.14**	--	.01	--	**.09**	--	-.04
Tolerance	**.35**	**.30**	**.29**	**.16**	**.33**	**.18**	**-.04**	-.02
TOPS Composites (Original)								
Can-Do	**-.06**	**.12**	**-.08**	.00	**-.18**	**-.11**	**-.05**	**-.12**
Will-Do	**-.09**	**.03**	**-.10**	**-.16**	**-.16**	**-.12**	**.06**	-.03
TOPS Composites (Revised)								
Can-Do	**-.30**	**-.31**	**-.29**	**-.20**	**-.32**	**-.22**	**-.09**	**-.21**
Will-Do	**-.29**	**-.21**	**-.17**	**-.17**	**-.12**	**-.13**	**.12**	.05
Adaptation	**-.45**	**-.47**	**-.31**	**-.35**	**-.25**	**-.23**	**.09**	.06

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS scores. The reported standardized group mean score differences (Cohen's *d*) reflect the difference in the mean scores between the first group listed (i.e., Female, Black, Hispanic-White, Tier 1) and the second group (i.e., Male, White, Non-Hispanic, Tier 2). A negative value indicates that the first group's mean is lower than the second group's mean, on average. Bolded differences are statistically significant, $p < .01$ (two-tailed), based on independent samples *t*-test analyses of mean differences. Missing values reflect scales that do not appear on all TAPAS versions.

CRITERION PSYCHOMETRIC PROPERTIES IN THE FULL IMT AND IN-UNIT SAMPLES

Table B.1. Descriptive Statistics and Reliability Estimates for the Job Knowledge Tests (JKTs) in the Full IMT Sample

Domain / JKT	n	M	SD	Min	Max	r_{WTBD}	α
MOS-Specific							
11B/C/X + 18X	12,578	61.4	10.1	20.9	88.4	.54	.78
19K	287	61.0	11.2	20.3	85.7	.56	.67
31B	6,365	68.5	8.5	35.0	93.2	.50	.77
68W	7,362	73.8	10.4	25.0	96.7	.50	.87
88M	5,005	63.6	10.8	30.6	94.4	.54	.77
91B	1,187	57.3	13.5	23.7	90.7	.47	.90
All MOS Combined	32,784	65.8	11.4	20.3	96.7	.54	.80
WTBD (Army-Wide)	39,882	64.9	12.7	6.0	100.0	--	.65

Note. Means, *SD*s, Min, and Max are based on percent correct. α = coefficient alpha. WTBD = Warrior Tasks and Battle Drills. r_{WTBD} = correlation with WTBD JKT scores. All correlations are statistically significant ($p < .01$, one-tailed). The coefficient alpha reported for All MOS Combined represents a sample-size weighted average of the MOS-specific estimates.

Table B.2. Descriptive Statistics and Reliability Estimates for the Performance Rating Scales (PRS) in the Full IMT Sample

Domain / PRS	n	M	SD	Min	Max	α	IRR
Army-Wide							
Adjustment to the Army	13,953	3.19	1.04	1.00	5.00	--	.14
Effort & Discipline	13,980	2.98	.99	1.00	5.00	--	.18
MOS Qualification Knowledge & Skill	12,761	3.15	.98	1.00	5.00	--	.10
Physical Fitness & Bearing	13,881	3.03	1.02	1.00	5.00	--	.09
Working with Others	13,925	2.93	.99	1.00	5.00	--	.16
Overall Performance	13,782	3.50	.85	1.00	5.00	--	.32
Overall Army-Wide [a]	13,982	3.05	.88	1.00	5.00	.94	.17
MOS-Specific							
11B/C/X + 18X	4,434	2.95	.81	1.00	5.00	.95	.17
19K	194	3.28	.56	1.00	4.86	.85	.41
31B	2,297	3.18	.81	1.00	5.00	.95	.10
68W	3,209	2.64	.72	1.00	5.00	.94	.01
88M	702	2.87	.74	1.00	5.00	.93	.00
91B	264	2.94	1.17	1.00	5.00	.97	.20
All MOS Combined	11,100	2.91	.81	1.00	5.00	.94	.10

Note. Ratings on PRS range from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. α = coefficient alpha. IRR = Single-rater interrater reliability, estimated using G(q,k) (Putka, Le, McCloy, & Diaz, 2008). IRR estimates were not estimated if 30 or fewer Soldiers were rated by more than one supervisor. The coefficient alpha and IRR reported for All MOS Combined represents a sample-size weighted average of the MOS-specific estimates.
[a] Composite of all scales except "Overall Performance."

Table B.3. Correlations among Performance Rating Scales (PRS) in the Full IMT Sample

Domain / PRS	1	2	3	4	5	6	7
Army-Wide							
1. Adjustment to the Army							
2. Effort & Discipline	.77						
3. MOS Qualification Knowledge & Skill	.72	.68					
4. Physical Fitness & Bearing	.68	.71	.63				
5. Working with Others	.74	.75	.71	.66			
6. Overall Performance	.55	.57	.52	.53	.54		
7. Overall Army-Wide [a]	.90	.90	.85	.85	.88	.62	
MOS-Specific							
8. 11B/C/X + 18X	.65	.65	.70	.61	.66	.53	.75
9. 19K	.66	.69	76	.62	.43	.58	.77
10. 31B	.64	.63	.69	.52	.64	.55	.73
11. 68W	.48	.44	.55	.38	.48	.30	.53
12. 88M	.58	.53	.63	.53	.56	.49	.66
13. 91B	.72	.67	.80	.67	.73	.57	.80
14. All MOS Combined	.62	.60	.68	.54	.62	.47	.70

Note. Army-wide PRS, $n = 12,724 – 13,953$. MOS-specific PRS, 11B, $n = 4,058 – 4,060$; 19K, $n = 185$; 31B, $n = 2,149 – 2,163$; 68W, $n = 1,826 –2,363$; 88M, $n = 632 – 649$; 91B, $n = 237 – 256$. Ratings on PRS range from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. All correlations are statistically significant ($p < .01$, one-tailed).
[a] Composite of all scales except "Overall Performance."

Table B.4. Descriptive Statistics and Reliability Estimates for the Army Life Questionnaire (ALQ) in the Full IMT Sample

Domain / Scale	*n*	*M*	*SD*	*Min*	*Max*	*α*
Retention						
Affective Commitment	41,611	3.88	.68	1.00	5.00	.86
Army Life Adjustment	41,611	4.08	.66	1.00	5.00	.86
Army Fit	41,611	4.07	.60	1.00	5.00	.86
Attrition Cognitions	41,611	1.53	.61	1.00	5.00	.77
MOS Fit	41,611	3.78	.84	1.00	5.00	.92
Achievement/Performance						
Disciplinary Incidents (#)	29,366	.27	.61	0.00	7.00	--
Disciplinary Incidents (Y/N)	29,366	.20	.40	0.00	1.00	--
Last APFT Score	41,120	250.28	31.55	100.00	300.00	--
Training Achievements (#)	41,572	.41	.61	0.00	2.00	--
Training Restarts (#)	41,609	.40	.64	0.00	4.00	--
Training Restarts (Y/N)	41,609	.33	.47	0.00	1.00	--

Note. α = coefficient alpha. APFT = Army Physical Fitness Test

Table B.5. Correlations among Army Life Questionnaire (ALQ) Scales in the Full IMT Sample

Domain / Scale	1	2	3	4	5	6	7	8	9	10
Retention										
1. Affective Commitment										
2. Army Life Adjustment	.45									
3. Army Fit	.84	.61								
4. Attrition Cognitions	-.63	-.53	-.68							
5. MOS Fit	.48	.36	.49	-.42						
Achievement/Performance										
6. Disciplinary Incidents (#)	-.07	-.17	-.11	.12	-.08					
7. Disciplinary Incidents (Y/N)	-.05	-.17	-.08	.09	-.07	.86				
8. Last APFT Score	.04	.24	.10	-.12	.08	-.14	-.16			
9. Training Achievement (#)	.06	.13	.08	-.04	.05	-.07	-.09	.23		
10. Training Restarts (#)	-.06	-.20	-.10	.12	-.09	.21	.20	-.28	-.12	
11. Training Restarts (Y/N)	-.06	-.19	-.08	.10	-.08	.18	.19	-.26	-.12	.90

Note. n = 29,335 – 41,611. All correlations are statistically significant (*p* < .01, two-tailed). APFT = Army Physical Fitness Test

Table B.6. Descriptive Statistics and Reliability Estimates for the Job Knowledge Tests (JKTs) in the Full In-Unit Sample

Domain / JKT	*n*	*M*	*SD*	*Min*	*Max*	r_{WTBD}	α
MOS-Specific							
11B/C/X + 18X	466	63.4	10.4	26.8	84.5	.59	.74
19K	--	--	--	--	--	--	--
31B	--	--	--	--	--	--	--
68W	94	74.2	8.3	50.9	90.6	.42	.57
88M	116	62.9	9.9	40.4	87.2	.65	.82
91B	116	58.9	11.6	35.1	85.3	.29	.75
All MOS Combined	876	64.1	11.3	26.8	90.6	.55	.73
WTBD (Army-Wide)	2,485	66.5	13.1	15.0	100.0	--	.62

Note. Means, *SD*s, Min, and Max are based on percent correct. α = coefficient alpha. WTBD = Warrior Tasks and Battle Drills. r_{WTBD} = correlation with WTBD JKT scores. All correlations are statistically significant (*p* < .05, one-tailed). Statistics based on fewer than 50 cases are not reported. The coefficient alpha reported for All MOS Combined represents a sample-size weighted average of the MOS-specific estimates, excluding MOS with insufficient sample size.

Table B.7. Descriptive Statistics and Reliability Estimates for the Performance Rating Scales (PRS) in the Full In-Unit Sample

Army-wide PRS	*n*	*M*	*SD*	*Min*	*Max*	α
Can Do [a]	1,370	4.83	1.22	1.00	7.00	.83
Effort & Personal Discipline [a]	1,370	5.19	1.35	1.00	7.00	.77
Physical Fitness & Bearing	1,367	5.22	1.52	1.00	7.00	--
Self-Management [a]	1,368	5.28	1.09	1.00	7.00	.78
Working with Others [a]	1,370	5.24	1.19	1.00	7.00	.74
Overall Leadership Potential	1,357	4.75	1.64	1.00	7.00	--
Army-Wide [a]	1,371	5.10	1.09	1.00	7.00	.94

Note. PRS ratings range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. α = coefficient alpha. Interrater reliability (IRR) estimates were not computed because a limited number of Soldiers were rated by more than one supervisor. Statistics were not reported on the new PRS and Adjustment to Army Life because of insufficient sample size as of May 2012.
[a] Ratings composite comprised of two or more Army-wide PRS.

Table B.8. Correlations among Performance Rating Scales (PRS) in the Full In-Unit Sample

Army-Wide PRS	1	2	3	4	5	6
1. Can Do [a]						
2. Effort & Personal Disc [a]	.75					
3. Physical Fitness & Bearing	.57	.61				
4. Self-Management [a]	.76	.75	.60			
5. Working with Others [a]	.76	.77	.59	.75		
6. Overall Leadership Potential	.67	.69	.58	.66	.64	
7. Army-Wide [a]	.93	.89	.71	.90	.88	.74

Note. Army-wide PRS, *n* = 1,353 - 1,371. PRS ratings range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. Statistics were not reported on the new PRS, Adjustment to Army Life, because of insufficient sample size, as of May 2012. All correlations are statistically significant (*p* < .05, one-tailed).
[a] Ratings composite comprised of two or more Army-wide PRS.

Table B.9. Descriptive Statistics and Reliability Estimates for the Army Life Questionnaire (ALQ) in the Full In-Unit Sample

Domain / Scale	*n*	*M*	*SD*	*Min*	*Max*	*α*
Retention						
Affective Commitment	2,509	3.66	.78	1.00	5.00	.88
Army Career Intentions	2,509	2.70	1.17	1.00	5.00	.92
Army Fit	2,509	3.94	.68	1.17	5.00	.80
Attrition Cognitions	2,509	1.65	.69	1.00	5.00	.76
MOS Fit	2,509	3.29	.93	1.00	5.00	.93
MOS Satisfaction	2,509	3.58	.89	1.00	5.00	.92
Reenlistment Intentions	2,509	3.05	1.14	1.00	5.00	.79
Achievement/Performance						
Accelerated Development	2,490	.29	.57	0.00	3.00	--
Award Pts Earned (Weighted)[a]	2,509	4.85	9.90	0.00	80.00	--
Disciplinary Incidents (#)	2,509	.34	.84	0.00	7.00	--
Disciplinary Incidents (Y/N)	2,509	.21	.41	0.00	1.00	--
Last APFT Score	2,434	244.99	32.34	105.00	300.00	--

Note. α = coefficient alpha. APFT = Army Physical Fitness Test.
[a] Award points earned were weighted by the number of promotion points assigned to each award according to current Army Enlisted promotion policy.

Table B.10. Correlations among the Army Life Questionnaire (ALQ) Scales in the Full In-Unit Sample

Domain/Scale	1	2	3	4	5	6	7	8	9	10	11
Retention											
1. Affective Commitment											
2. Army Career Intentions	**.59**										
3. Army Fit	**.77**	**.57**									
4. Attrition Cognitions	**-.57**	**-.47**	**-.66**								
5. MOS Fit	**.38**	**.21**	**.41**	**-.31**							
6. MOS Satisfaction	**.51**	**.33**	**.52**	**-.38**	**.56**						
7. Reenlistment Intentions	**.52**	**.81**	**.54**	**-.44**	**.18**	**.25**					
Achievement/Performance											
8. Accelerated Development	.00	-.01	.01	-.02	.01	.01	.00				
9. Award Pts Earned (Weighted)[a]	**-.04**	**-.05**	-.01	-.02	**.04**	.02	-.04	**.18**			
10. Disciplinary Incidents (#)	**-.13**	**-.09**	**-.19**	**.23**	**-.09**	**-.09**	**-.05**	-.04	.02		
11. Disciplinary Incidents (Y/N)	**-.11**	**-.08**	**-.17**	**.18**	**-.09**	**-.09**	**-.05**	-.03	**.02**	**.80**	
12. Last APFT Score	**.05**	.03	**.09**	**-.12**	.03	.02	.03	**.13**	**.08**	**-.05**	**-.08**

Note. n = 2,419 - 2,509. Correlations in bold are statistically significant (*p* < .05, two-tailed). APFT = Army Physical Fitness Test.
[a] Award points earned are weighted by the number of promotion points assigned to each award according to current Army Enlisted promotion policy.

CORRELATIONS AMONG CRITERION MEASURES IN THE IMT AND IN-UNIT VALIDATION SAMPLES

Table C.1. Correlations among the Performance Rating Scales (PRS) in the IMT Validation Sample

Domain/PRS	1	2	3	4	5	6	7
Army-Wide							
1. Adjustment to the Army							
2. Effort & Discipline	.77						
3. MOS Qualification Knowledge & Skill	.73	.69					
4. Physical Fitness & Bearing	.68	.72	.61				
5. Working with Others	.74	.77	.71	.66			
6. Overall Performance	.56	.59	.54	.53	.56		
7. Overall Army-Wide [a]	.90	.90	.85	.84	.89	.63	
MOS-Specific							
8. 11B/C/X + 18X	.63	.64	.69	.60	.64	.55	.74
9. 19K	.64	.65	.76	.62	.46	.60	.77
10. 31B	.67	.64	.73	.52	.68	.55	.74
11. 68W	.65	.60	.65	.46	.61	.32	.70
12. 88M	.64	.59	.61	.60	.59	.57	.70
13. 91B	.75	.79	.83	.78	.82	.58	.84
14. All MOS Combined	.66	.65	.70	.56	.65	.50	.74

Note. Army-wide PRS, n = 3,733 – 3,932. MOS-specific PRS, 11B, *n* = 1,138 – 1,141; 19K, *n* = 72; 31B, *n* = 793 – 796; 68W, *n* = 490 – 516; 88M, *n* = 116 – 117; 91B, *n* = 51 – 54; All MOS Combined, *n* = 2,667 – 2,696. Ratings on PRS range from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. All correlations are statistically significant ($p < .01$, one-tailed).
[a] Composite of all scales except "Overall Performance."

Table C.2. Correlations among the Army Life Questionnaire (ALQ) Scales in the IMT Validation Sample

Domain / Scale	1	2	3	4	5	6	7	8	9	10
Retention										
1. Affective Commitment										
2. Army Life Adjustment	.46									
3. Army Fit	.84	.61								
4. Attrition Cognitions	-.62	-.54	-.69							
5. MOS Fit	.47	.36	.48	-.41						
Achievement/Performance										
6. Disciplinary Incidents (#)	-.07	-.17	-.10	.11	-.08					
7. Disciplinary Incidents (Y/N)	-.04	-.17	-.07	.08	-.07	.86				
8. Last APFT Score	.03	.22	.09	-.11	.08	-.14	-.16			
9. Training Achievement (#)	.06	.13	.06	-.04	.06	-.07	-.09	.23		
10. Training Restarts (#)	-.05	-.20	-.08	.10	-.09	.21	.20	-.27	-.12	
11. Training Restarts (Y/N)	-.04	-.18	-.07	.08	-.08	.18	.19	-.25	-.12	.91

Note. n = 10,553-12,401. All correlations are statistically significant ($p < .01$, two-tailed). APFT = Army Physical Fitness Test

Table C.3. Correlations between the IMT Job Knowledge Tests (JKTs) and Performance Rating Scales (PRS) in the IMT Validation Sample

	JKT							
Domain / PRS	All MOS Combined	11B	19K	31B	68W	88M	91B	WTBD
Army-Wide								
Adjustment to the Army	.01	.03	**.29**	.04	-.01	**.22**	--	**.06**
Effort & Discipline	**.05**	.04	**.34**	.05	.01	.19	--	**.08**
MOS Qualification Knowledge & Skill	.04	.04	**.23**	.04	-.03	.09	--	**.07**
Physical Fitness & Bearing	.02	.03	**.32**	.04	.00	.17	--	**.07**
Working with Others	**.04**	.04	.16	.03	.00	**.22**	--	**.07**
Overall Performance	**.07**	**.08**	**.30**	.06	.01	.11	--	**.08**
Overall Army-Wide [a]	.03	.04	**.33**	.04	.00	**.21**	--	**.08**
MOS-Specific								
11B	.04	.04	--	--	--	--	--	**.11**
19K	.25	--	.25	--	--	--	--	**.32**
31B	-.01	--	--	-.01	--	--	--	**.09**
68W	.04	--	--	--	.04	--	--	.07
88M	.03	--	--	--	--	.03	--	.02
91B	--	--	--	--	--	--	--	.13
All MOS Combined	-.02	.04	.25	-.01	.04	.03	--	**.11**

Note. WTBD = Warrior Tasks and Battle Drills. Army-wide PRS, All MOS Combined, *n* = 3,066-3,227; 11B, *n* = 1,136-1,139; 19K, *n* = 74-77; 31B, *n* = 860-862; 68W, *n* = 850-1,000; 88M, *n* = 101-105; WTBD, *n* = 3,507-3,684. MOS-specific PRS, All MOS Combined, *n* = 58-2,791; 11B, *n* = 945-1,118; 19K, *n* = 58-67; 31B, *n*= 767-833; 68W, *n* = 554-613; 88M, *n* = 95-110; WTDB, *n* = 50-2,791. Correlations in bold are statistically significant ($p < .01$, two-tailed). Cells are blank if *n* < 50 or estimating the correlation is not feasible.
[a] Composite of all scales except "Overall Performance."

Table C.4. Correlations between the Job Knowledge Tests (JKTs) and the Army Life Questionnaire (ALQ) Scales in the IMT Validation Sample

Domain / Scale	All MOS Combined	11B	19K	31B	68W	88M	91B	WTBD
Retention								
Affective Commitment	**.03**	**.10**	.12	.05	**.06**	.03	**.19**	**.09**
Army Life Adjustment	**.09**	**.12**	.03	**.13**	**.14**	**.09**	.14	**.13**
Army Fit	**.08**	**.16**	.15	**.07**	**.14**	**.07**	**.20**	**.14**
Attrition Cognitions	**-.12**	**-.18**	-.02	**-.11**	**-.18**	**-.11**	**-.18**	**-.18**
MOS Fit	**.13**	**.10**	.08	.04	**.17**	.02	**.30**	**.13**
Achievement/Performance								
Disciplinary Incidents (#)	.00	-.02	.15	**-.10**	**-.06**	.01	-.02	**-.04**
Disciplinary Incidents (Y/N)	.02	-.01	.12	**-.08**	-.04	.01	.00	-.02
Last APFT Score	**.05**	**.05**	.06	-.01	.02	.00	-.01	**.08**
Training Achievement (#)	**-.12**	**-.14**	-.15	-.05	.02	**-.14**	**-.19**	**-.10**
Training Restarts (#)	-.02	**-.07**	.01	**-.12**	-.05	**-.10**	-.01	**-.08**
Training Restarts (Y/N)	-.02	**-.06**	.06	**-.11**	-.03	**-.11**	-.01	**-.07**

Note. WTBD = Warrior Tasks and Battle Drills. APFT = Army Physical Fitness Test. All MOS Combined, *n* = 8,540-9,926; 11B, *n* = 3,992-4,021; 19K, *n* = 109-112; 31B, *n* = 1,707-2105; 68W, *n* = 1,418-1,957; 88M, *n* = 1,120-1,463; 91B, *n* = 162-268; WTBD, *n* = 10,058-11,740. Correlations in bold are statistically significant (*p* < .01, two-tailed).

Table C.5. Correlations between the Army Life Questionnaire (ALQ) Scales and the Performance Ratings Scales (PRS) in the IMT Validation Sample

	ALQ Scale								
Domain / PRS	AFF COM	LIFE ADJ	Army Fit	ATT COG	MOS Fit	DSC INC	LAST APFT	TRN ACH	TRN RST
Army-Wide									
Adjustment to the Army	**.05**	**.06**	**.07**	**-.04**	**.07**	**-.10**	**.13**	**.11**	**-.05**
Effort & Discipline	**.05**	**.06**	**.07**	**-.04**	**.07**	**-.11**	**.13**	**.10**	**-.03**
MOS Qualification Knowledge & Skill	.03	**.06**	**.04**	-.02	**.06**	**-.07**	**.11**	**.07**	-.03
Physical Fitness & Bearing	.03	**.09**	**.05**	**-.05**	**.07**	**-.10**	**.26**	**.14**	**-.09**
Working with Others	.03	**.04**	**.04**	-.03	**.05**	**-.07**	**.11**	**.08**	-.03
Overall Performance	**.04**	**.13**	**.07**	**-.07**	**.07**	**-.14**	**.21**	**.16**	**-.12**
Overall Army-Wide [a]	**.04**	**.07**	**.06**	**-.04**	**.07**	**-.10**	**.17**	**.11**	**-.05**
MOS-Specific									
11B	.02	.05	.05	-.06	**.11**	-.03	**.12**	.06	-.02
19K	.05	.17	.08	-.16	.09	-.18	.07	-.07	-.21
31B	**.11**	**.16**	**.14**	**-.11**	.06	**-.16**	.05	**.13**	-.02
68W	.05	.08	.07	**-.10**	.01	-.02	.01	.03	.04
88M	-.01	-.15	-.03	.12	-.09	-.09	**.19**	-.03	.06
91B	-.31	-.04	-.28	.24	.18	--	.06	.14	-.07
All MOS Combined	**.06**	**.08**	**.09**	**-.07**	**.07**	**-.06**	**.09**	**.07**	-.02

Note. AFFCOM=Affective Commitment; LIFEADJ=Army Life Adjustment; ATTCOG=Attrition Cognitions; DSCINC=Disciplinary Incidents (# of); LASTAPFT=Last Army Physical Fitness Test Score; TRNACH=Training Achievements (# of); TRNRST=Training Restarts (# of). Army-wide PRS, n = 3,250-3,831. MOS-specific PRS, All MOS Combined, n = 2,514-2,900; 11B, n = 1,170-1,179; 19K, n = 71-72; 31B, n= 711-835; 68W, n = 433-636; 88M, n = 92-125; 91B, n = 52-53. Correlations in bold are statistically significant ($p < .01$, two-tailed). Cells are blank if $n < 50$.
[a] Composite of all scales except "Overall Performance."

Table C.6. Correlations between the Job Knowledge Tests (JKTs) and Administrative Criteria in the IMT Validation Sample

	JKT							
Domain/Measure	All MOS Combined	11B	19K	31B	68W	88M	91B	WTBD
Attrition								
6-Month Cumulative	-.01	-.01	--	.04	--	.00	--	.00
9-Month Cumulative	**-.04**	-.02	--	-.05	.01	-.07	-.01	-.03
12-Month Cumulative	**-.05**	-.02	--	-.05	-.05	-.06	-.01	-.02
15-Month Cumulative	**-.06**	-.02	--	-.08	-.07	-.01	-.01	-.01
18-Month Cumulative	**-.09**	-.04	--	-.14	**-.12**	-.05	-.04	**-.05**
21-Month Cumulative	**-.13**	-.08	--	**-.23**	**-.16**	-.06	-.20	**-.08**
24-Month Cumulative	**-.21**	**-.19**	--	**-.32**	**-.22**	--	--	**-.15**
IMT Restarts								
Restarted at Least Once During IMT	-.02	**.05**	--	.03	-.03	.01	-.05	.01
Academic or Other Pejorative Restart	.01	.04	--	.01	.04	.02	-.05	.02
Academic Restart	-.03	**.05**	--	.04	-.03	.01	-.05	.01
Final AIT School Grades								
Overall Average (Unstandardized)	.23	--	--	--	**.36**	--	--	**.29**
Overall Average (Standardized)	**.27**	--	--	--	**.34**	--	--	**.38**

Note. WTBD = Warrior Tasks and Battle Drills. Attrition, All MOS Combined, *n* = 503-9,206; 11B, *n* = 156-3,811; 31B, *n*= 67-1,832; 68W, *n* = 206-1,842; 88M, *n* = 138-1,369; 91B, *n* = 112-256; WTBD, *n* = 614-10,912. IMT Restarts, All MOS Combined, *n* = 6,816-6,930; 11B, *n* = 2,692-2,702; 31B, *n*= 1,283-1,291; 68W, *n* = 1,280-1,339; 88M, *n* = 1,248-1,292; 91B, *n* = 251-256; WTBD, *n* = 8,100-8,233. Final AIT School Grade, All MOS Combined, *n* = 86-87; 68W, *n* = 77-78; WTBD, *n* = 96-97. Correlations in bold are statistically significant (*p* < .01, two-tailed). Cells are blank if *n* < 50.

Table C.7. Correlations between the Army Life Questionnaire (ALQ) Scales and Administrative Criteria in the IMT Validation Sample

	ALQ Scale								
	AFF	LIFE	Army	ATT	MOS	DSC	LAST	TRN	TRN
Domain/Measure	COM	ADJ	Fit	COG	Fit	INC	APFT	ACH	RST
Attrition									
6-Month Cumulative	**-.04**	**-.09**	**-.06**	**.14**	**-.05**	**.07**	**-.09**	-.02	**.09**
9-Month Cumulative	**-.06**	**-.10**	**-.08**	**.18**	**-.06**	**.08**	**-.08**	-.01	**.08**
12-Month Cumulative	**-.07**	**-.11**	**-.10**	**.19**	**-.07**	**.08**	**-.08**	-.03	**.07**
15-Month Cumulative	**-.05**	**-.09**	**-.06**	**.15**	**-.06**	**.07**	**-.07**	**-.05**	.04
18-Month Cumulative	-.03	**-.06**	**-.05**	**.13**	**-.05**	**.07**	**-.06**	**-.05**	.03
21-Month Cumulative	-.05	**-.08**	**-.08**	**.16**	-.05	**.15**	**-.07**	**-.06**	.02
24-Month Cumulative	**-.10**	**-.13**	**-.14**	**.15**	-.06	**.20**	-.07	-.02	.05
IMT Restarts									
Restarted at Least Once During IMT	.01	**.03**	.02	**-.04**	.02	**-.09**	**.09**	-.01	**-.22**
Academic or Other Pejorative Restart	.00	**.05**	.02	**-.05**	**.03**	**-.10**	**.09**	.03	**-.24**
Academic Restart	.00	.01	.01	-.02	.01	**-.07**	**.07**	-.02	**-.21**
Final AIT School Grades									
Overall Average (Unstandardized)	.07	-.06	.11	-.10	.07	--	-.20	.12	.11
Overall Average (Standardized)	-.03	.05	.03	-.08	-.02	--	-.19	.06	-.01

Note. AFFCOM=Affective Commitment; LIFEADJ=Army Life Adjustment; ATTCOG=Attrition Cognitions; DSCINC=Disciplinary Incidents; LAST APFT=Last Army Physical Fitness Test Score; TRNACH=Training Achievements; TRNRST=Training Restarts. Correlations with attrition are limited to Regular Army Soldiers only. Attrition, n = 191 – 6,695. IMT Restarts, n = 7,287 – 8,660. Final AIT School Grade, n = 98 – 100. Correlations in bold are statistically significant ($p < .01$, two-tailed). Cells are blank if $n < 50$.

Table C.8. Correlations between the Performance Rating Scales (PRS) and the Administrative Criteria in the IMT Validation Sample

Domain / PRS	Attrition							IMT Restarts		
	6- Month	9- Month	12- Month	15- Month	18- Month	21- Month	24- Month	IMT Restart	PEJ Restart	ACAD Restart
Army-Wide										
Adjustment to the Army	-.04	-.05	-.04	-.01	-.03	-.04	-.02	**.06**	**.05**	.04
Effort & Discipline	-.04	-.02	.00	-.01	-.03	-.05	-.06	.01	.02	.00
MOS Qualification Knowledge & Skill	-.02	-.01	-.01	.02	.04	-.02	-.03	.01	.01	.01
Physical Fitness & Bearing	**-.07**	**-.06**	-.05	**-.08**	-.03	-.02	-.01	.03	.04	.01
Working with Others	-.02	-.01	.00	.02	.01	-.02	.01	.00	.00	-.01
Overall Performance	**-.06**	**-.08**	**-.07**	-.04	-.06	**-.13**	-.14	.03	**.05**	.01
Overall Army-Wide [a]	-.04	-.03	-.02	-.01	-.01	-.03	-.02	.03	.03	.01
MOS-Specific										
11B	.06	.05	.04	.03	-.05	-.07	-.05	-.04	-.03	-.04
19K	--	--	--	--	--	--	--	--	--	--
31B	-.08	-.10	-.03	-.04	.05	--	--	.05	.06	.03
68W	--	-.01	.03	.03	-.07	-.09	-.13	-.05	-.03	-.05
88M	--	--	--	--	--	--	--	**-.21**	.06	**-.21**
91B	--	--	--	--	--	--	--	.01	.01	--
All MOS Combined	.02	.02	.04	.04	-.01	-.05	-.02	-.01	.00	-.01

Note. IMT Restart -= Restarted at Least Once During IMT. PEJ Restart = Restarted at Least Once for Academic or Other Pejorative Reasons. ACAD Restart = Restarted at Least Once for Academic Reasons. Army-wide PRS, Attrition, n = 180-2,227; IMT Restarts, n = 2,687-2,839. MOS-specific PRS, Attrition, n = 62-1,718; IMT Restarts, n = 54-2,217. Correlations in bold are statistically significant ($p < .01$, two-tailed). Cells are blank if $n < 50$. Correlations with Final AIT Grade are not reported because $n < 50$.
[a] Composite of all scales except "Overall Performance."

Table C.9. Correlations among the Performance Rating Scales (PRS) in the In-Unit Validation Sample

Army-Wide PRS	1	2	3	4	5	6
1. Can Do [a]						
2. Effort & Personal Discipline [a]	.76					
3. Physical Fitness & Bearing	.56	.62				
4. Self-Management [a]	.77	.78	.63			
5. Working with Others [a]	.75	.80	.60	.77		
6. Overall Leadership Potential	.69	.70	.60	.69	.68	
7. Army-Wide [a]	.92	.90	.72	.91	.89	.77

Note. n = 300 - 304. Ratings on PRS range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 were excluded from analyses. All correlations are statistically significant ($p < .05$, one-tailed).
[a] Ratings composite comprises two or more Army-wide PRS.

Table C.10. Correlations among the Army Life Questionnaire (ALQ) Scales in the In-Unit Validation Sample

Domain/Scale	1	2	3	4	5	6	7	8	9	10	11
Retention											
1. Affective Commitment											
2. Army Career Intentions	**.57**										
3. Army Fit	**.76**	**.58**									
4. Attrition Cognitions	**-.50**	**-.45**	**-.60**								
5. MOS Fit	**.44**	**.27**	**.44**	**-.29**							
6. MOS Satisfaction	**.49**	**.29**	**.52**	**-.30**	**.55**						
7. Reenlistment Intentions	**.49**	**.80**	**.49**	**-.40**	**.21**	**.23**					
Achievement/Performance											
8. Accelerated Development	-.02	-.07	.00	.01	-.01	.03	-.06				
9. Award Pts Earned (Weighted) [a]	-.04	-.06	.01	.00	.06	.07	-.02	**.15**			
10. Disciplinary Incidents (#)	**-.10**	-.05	**-.11**	**.22**	**-.10**	-.05	-.05	-.02	.01		
11. Disciplinary Incidents (Y/N)	**-.10**	-.06	**-.10**	**.15**	**-.11**	-.07	-.04	-.01	.03	**.82**	
12. Last APFT Score	.03	.04	.04	-.08	.03	-.02	.03	**.11**	**.10**	-.03	-.04

Note. n = 579 - 600. Correlations in bold are statistically significant ($p < .05$, two-tailed). APFT = Army Physical Fitness Test.
[a] Awards earned are weighted by the number of promotion points associated with each award according to current Army Enlisted promotion policy.
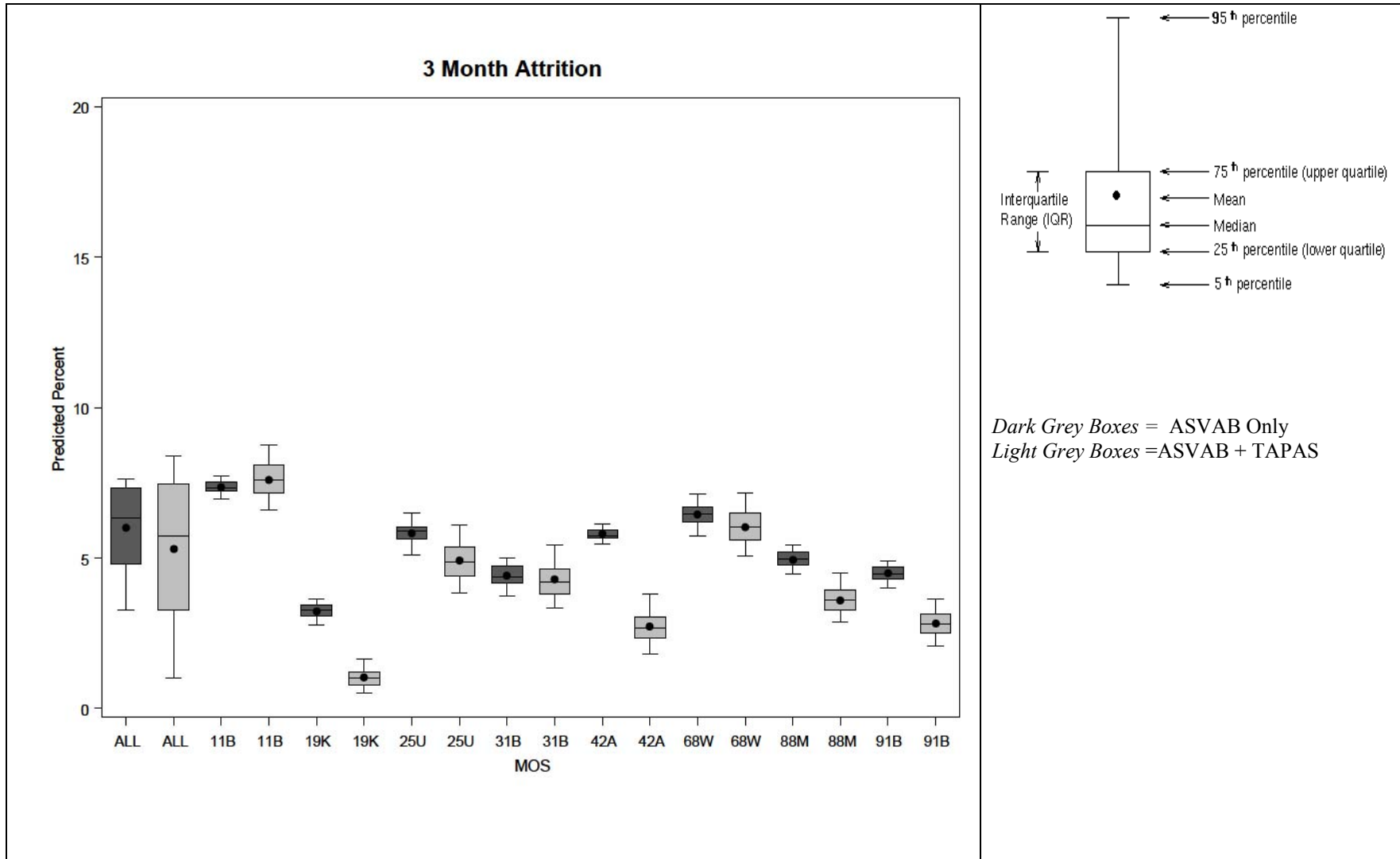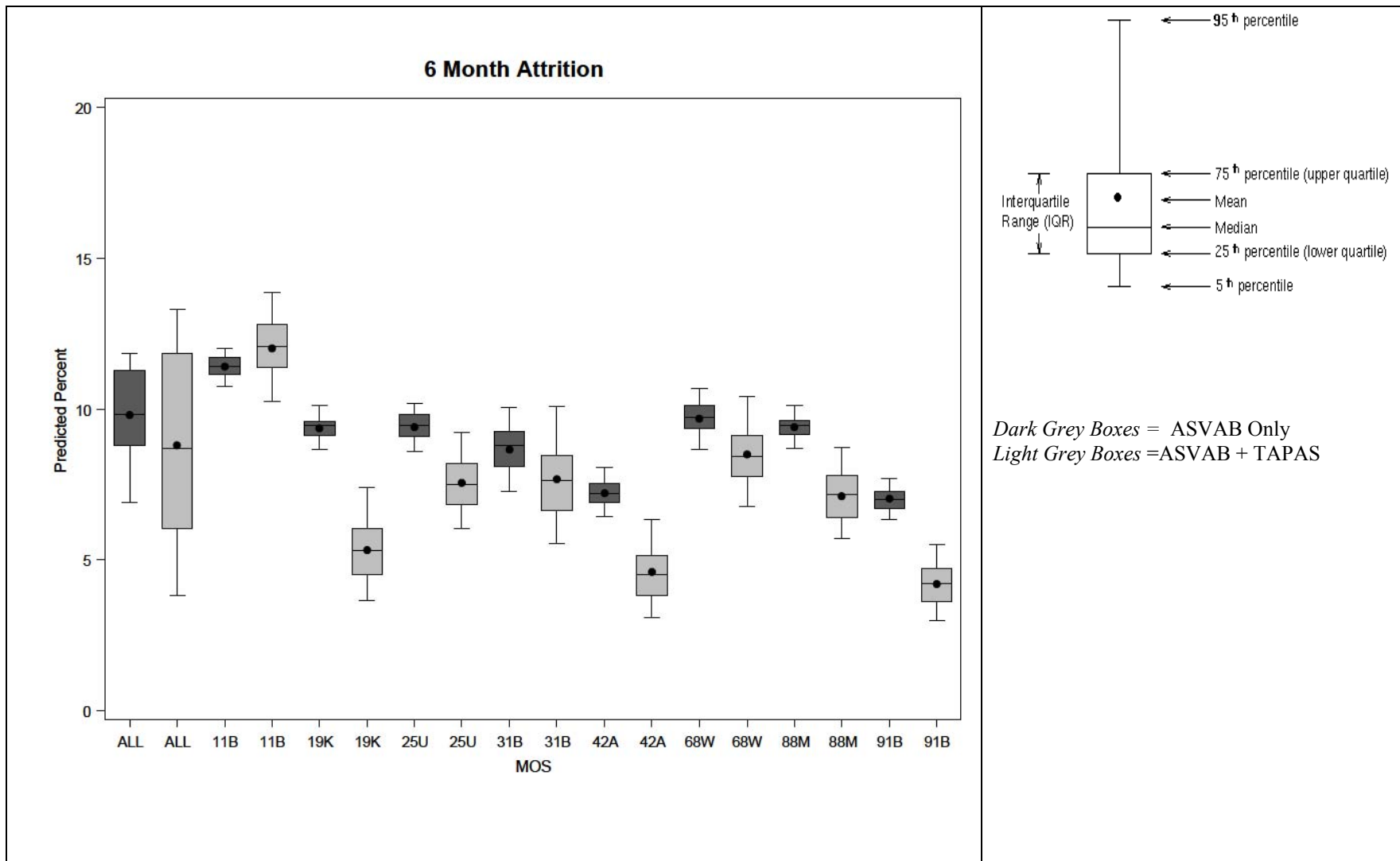
Figure D.1. Boxplot of the classification potential of the TAPAS relative to the ASVAB for minimizing 3-month attrition.

Table D.1. Classification Potential of the TAPAS Relative to the ASVAB for Minimizing 3-Month Attrition

		Predicted Percentage (%) of Soldiers Attriting	
		3-Month	
		ASVAB Only	*ASVAB + TAPAS*
	Obs	*(5%, 95%)*	*(5%, 95%)*
Overall	7.2	6.0	5.3%
	--	(3.3, 7.6)	(1.0, 8.4)
11B	7.8	7.4	7.6
	--	(7.0, 7.7)	(6.6, 8.7)
19K	5.0	3.2	**1.0**
	--	(2.8, 3.6)	**(.5, 1.7)**
25U	8.1	5.8	4.9
	--	(5.1, 6.5)	(3.9, 6.1)
31B	6.6	4.4	4.3
	--	(3.7, 5.0)	(3.4, 5.4)
42A	6.2	5.8	**2.7**
	--	(5.5, 6.1)	**(1.8, 3.8)**
68W	6.0	6.5	6.0
	--	(5.7, 7.1)	(5.1, 7.2)
88M	7.7	4.9	3.6
	--	(4.5, 5.4)	(2.9, 4.5)
91B	5.8	4.5	**2.8**
	--	(4.0, 4.9)	**(2.1, 3.7)**

Note. Obs = Observed aggregate attrition rate by MOS or for the target MOS overall. ASVAB Only = Predicted aggregate attrition rates when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Predicted aggregate attrition rates when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (11B = 42%, 19K = 2%, 25U = 4%, 31B = 10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, n = 23,058; 19K, n = 1,236; 25U, n = 2,297; 31B, n = 5,691; 42A, n = 2,979; 68W, n = 6,902; 88M, n = 6,731; 91B, n = 6,158).
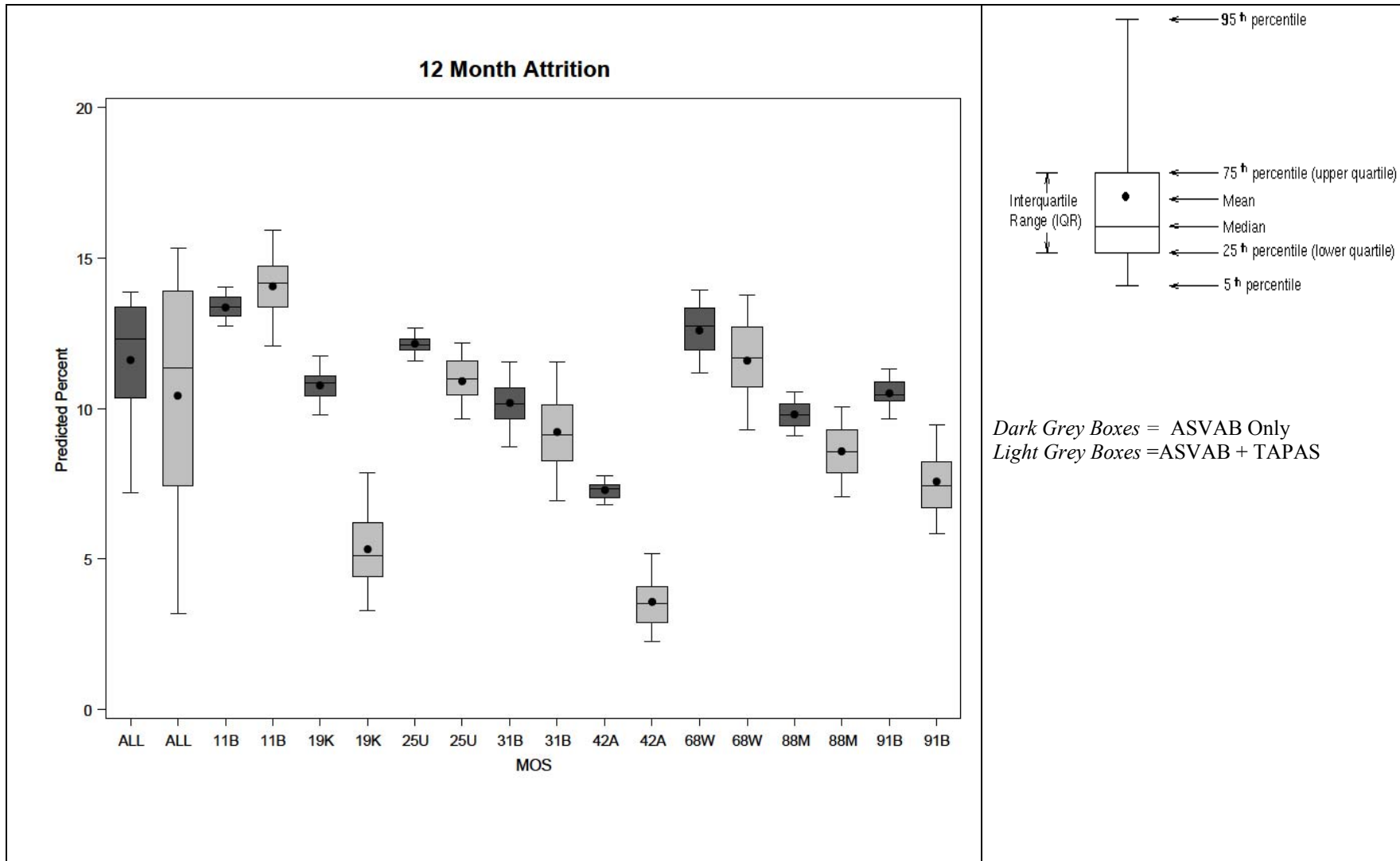
Figure D.2. Boxplot of the classification potential of the TAPAS relative to the ASVAB for minimizing 6-month attrition.

Table D.2. Classification Potential of the TAPAS Relative to the ASVAB for Minimizing 6-Month Attrition

		Predicted Percentage (%) of Soldiers Attriting	
		6-Month	
		ASVAB Only	*ASVAB + TAPAS*
	Obs	*(5%, 95%)*	*(5%, 95%)*
Overall	11.2	9.8	8.8
	--	(6.9, 11.9)	(3.8, 13.3)
11B	12.0	11.4	12.0
	--	(10.7, 12.0)	(10.3, 13.8)
19K	11.1	9.4	**5.3**
	--	(8.7, 10.1)	**(3.7, 7.4)**
25U	11.0	9.4	7.6
	--	(8.6, 10.2)	(6.0, 9.2)
31B	12.5	8.7	7.7
	--	(7.2, 10.0)	(5.5, 10.1)
42A	9.8	7.2	**4.6**
	--	(6.4, 8.0)	**(3.1, 6.3)**
68W	9.0	9.7	8.5
	--	(8.6, 10.7)	(6.8, 10.4)
88M	11.6	9.4	7.1
	--	(8.7, 10.1)	(5.7, 8.7)
91B	8.9	7.0	**4.2**
	--	(6.3, 7.7)	**(3.0, 5.5)**

Note. Obs = Observed aggregate attrition rate by MOS or for the target MOS overall. ASVAB Only = Predicted aggregate attrition rates when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Predicted aggregate attrition rates when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (11B = 42%, 19K = 2%, 25U = 4%, 31B = 10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, n = 23,058; 19K, n = 1,236; 25U, n = 2,297; 31B, n = 5,691; 42A, n = 2,979; 68W, n = 6,902; 88M, n = 6,731; 91B, n = 6,158).
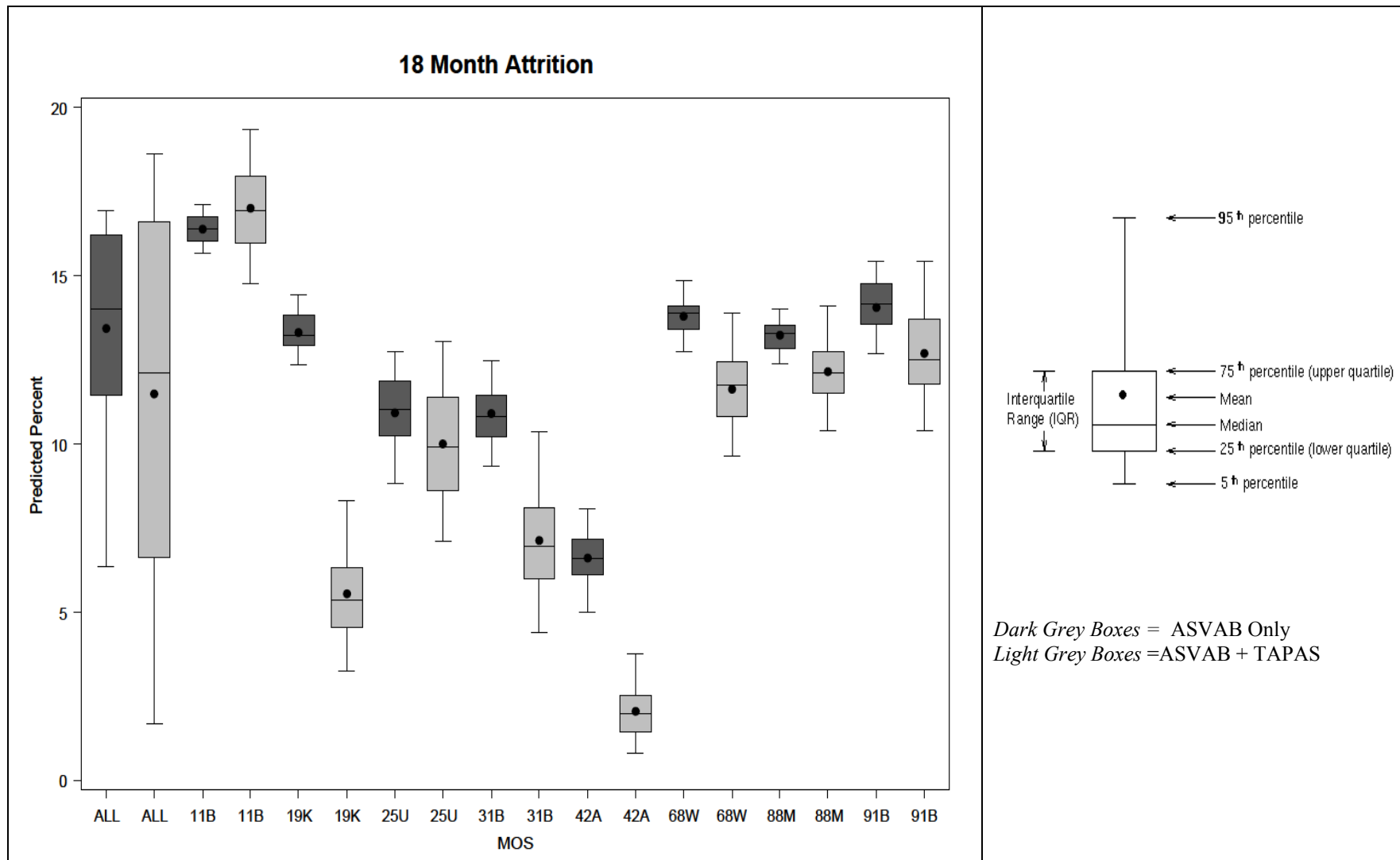
Figure D.3. Boxplot of the classification potential of the TAPAS relative to the ASVAB for minimizing 12-month attrition.

Table D.3. Classification Potential of the TAPAS Relative to the ASVAB for Minimizing 12-Month Attrition

		Predicted Percentage (%) of Soldiers Attriting	
		12-Month	
		ASVAB Only	*ASVAB + TAPAS*
	Obs	*(5%, 95%)*	*(5%, 95%)*
Overall	13.4	11.6	10.4
	--	(7.2, 13.9)	(3.2, 15.3)
11B	14.1	13.4	14.1
	--	(12.7, 14.0)	(12.1, 15.9)
19K	14.2	10.8	**5.3**
	--	(9.8, 11.8)	**(3.3, 7.9)**
25U	11.8	12.1	10.9
	--	(11.6, 12.7)	(9.6, 12.2)
31B	14.0	10.2	9.2
	--	(8.7, 11.5)	(6.9, 11.5)
42A	8.7	7.3	**3.6**
	--	(6.8, 7.7)	**(2.3, 5.2)**
68W	12.0	12.6	11.6
	--	(11.2, 13.9)	(9.3, 13.8)
88M	13.1	9.8	8.6
	--	(9.1, 10.5)	(7.1, 10.0)
91B	12.9	10.5	**7.6**
	--	(9.6, 11.3)	**(5.8, 9.4)**

Note. Obs = Observed aggregate attrition rate by MOS or for the target MOS overall. ASVAB Only = Predicted aggregate attrition rates when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Predicted aggregate attrition rates when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (11B = 42%, 19K = 2%, 25U = 4%, 31B = 10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, n = 23,058; 19K, n = 1,236; 25U, n = 2,297; 31B, n = 5,691; 42A, n = 2,979; 68W, n = 6,902; 88M, n = 6,731; 91B, n = 6,158).

Figure D.4. Boxplot of the classification potential of the TAPAS relative to the ASVAB for minimizing 18-month attrition.

Table D.4. Classification Potential of the TAPAS Relative to the ASVAB for Minimizing 3-Month, 6-Month, 12-Month and 18-Month Attrition

		Predicted Percentage (%) of Soldiers Attriting	
		18-Month	
		ASVAB Only	*ASVAB + TAPAS*
	Obs	*(5%, 95%)*	*(5%, 95%)*
Overall	16.3	13.4	11.5
	--	(6.4, 16.9)	(1.7, 18.6)
11B	17.1	16.4	17.0
	--	(15.7, 17.1)	(14.8, 19.4)
19K	17.7	13.3	**5.5**
	--	(12.3, 14.4)	**(3.2, 8.3)**
25U	16.4	10.9	10.0
	--	(8.8, 12.7)	(7.1, 13.0)
31B	15.3	10.9	7.1
	--	(9.3, 12.5)	(4.4, 10.4)
42A	11.5	6.6	**2.0**
	--	(5.0, 8.1)	**(.8, 3.8)**
68W	13.4	13.8	11.6
	--	(12.7, 14.9)	(9.6, 13.9)
88M	16.4	13.2	12.1
	--	(12.4, 14.0	(10.4, 14.1)
91B	16.7	14.1	12.7
	--	(12.7, 15.4)	(10.4, 15.4)

Note. Obs = Observed aggregate attrition rate by MOS or for the target MOS overall. ASVAB Only = Predicted aggregate attrition rates when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Predicted aggregate attrition rates when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (11B = 42%, 19K = 2%, 25U = 4%, 31B = 10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, n = 23,058; 19K, n = 1,236; 25U, n = 2,297; 31B, n = 5,691; 42A, n = 2,979; 68W, n = 6,902; 88M, n = 6,731; 91B, n = 6,158).
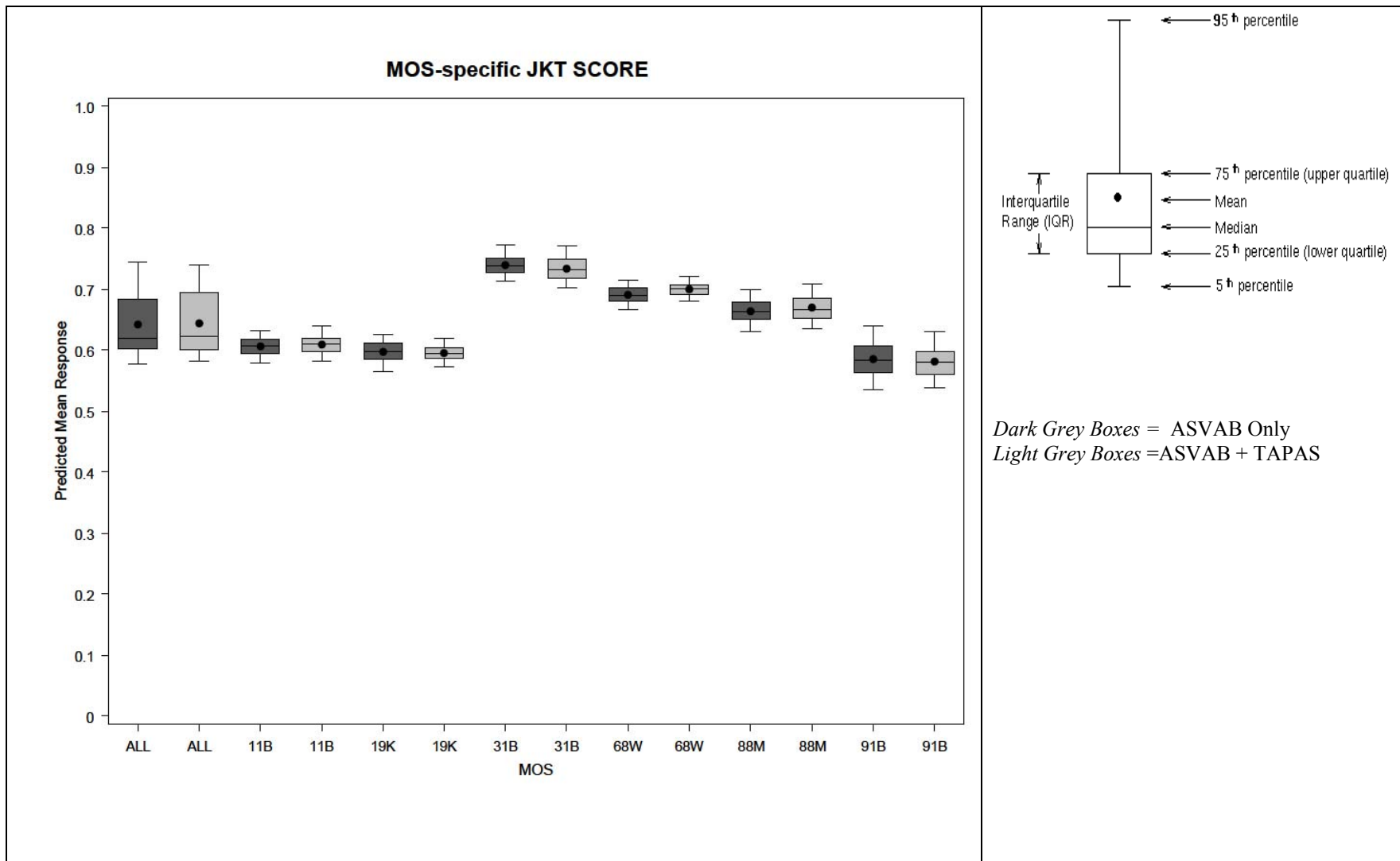
Figure D.5. Boxplot of the classification potential of the TAPAS relative to the ASVAB for maximizing MOS-specific Job Knowledge Test (JKT) score.

Table D.5. Classification Potential of the TAPAS Relative to the ASVAB for Maximizing MOS-Specific Job Knowledge Test (JKT)

		MOS-Specific JKT (% Correct)	
	Obs	*ASVAB Only (5%, 95%)*	*ASVAB + TAPAS (5%,95%)*
Overall	65.5	64.2	64.4
	--	(57.7, 74.5)	(58.2, 74.0)
11B	61.5	60.6	61.0
	--	(57.9, 63.2)	(58.2, 64.0)
19K	60.8	59.7	59.6
	--	(56.6, 62.5)	(57.2, 61.9)
25U	--	--	--
	--	--	--
31B	68.5	74.0	73.4
	--	(71.4, 77.2)	(70.1, 77.1)
42A	--	--	--
	--	--	--
68W	73.4	69.1	70.0
	--	(66.6, 71.5)	(68.0, 72.1)
88M	63.9	66.4	67.0
	--	(63.1, 69.8)	(63.5, 70.9)
91B	57.5	58.5	58.1
	--	(53.6, 64.0)	(53.9, 63.1)

Note. Obs = Observed mean predicted criterion score by MOS or for the target MOS overall. ASVAB Only = Mean predicted criterion score when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Mean predicted criterion score when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (MOS-Specific JKT, 11B = 46%, 19K = 2%, 31B=11%, 68W=14%, 88M = 14%, 91B = 12%; Restarted IMT, 11B = 42%, 19K=2%, 25U=4%, 31B=10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%; Last APFT Score, 11B = 44%, 19K=2%, 31B=11%, 42A =6%, 68W=13%, 88M = 13%, 91B =12%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, $n = 23,058$; 19K, $n = 1,236$; 25U, $n = 2,297$; 31B, $n = 5,691$; 42A, $n = 2,979$; 68W, $n = 6,902$; 88M, $n = 6,731$; 91B, $n = 6,158$).
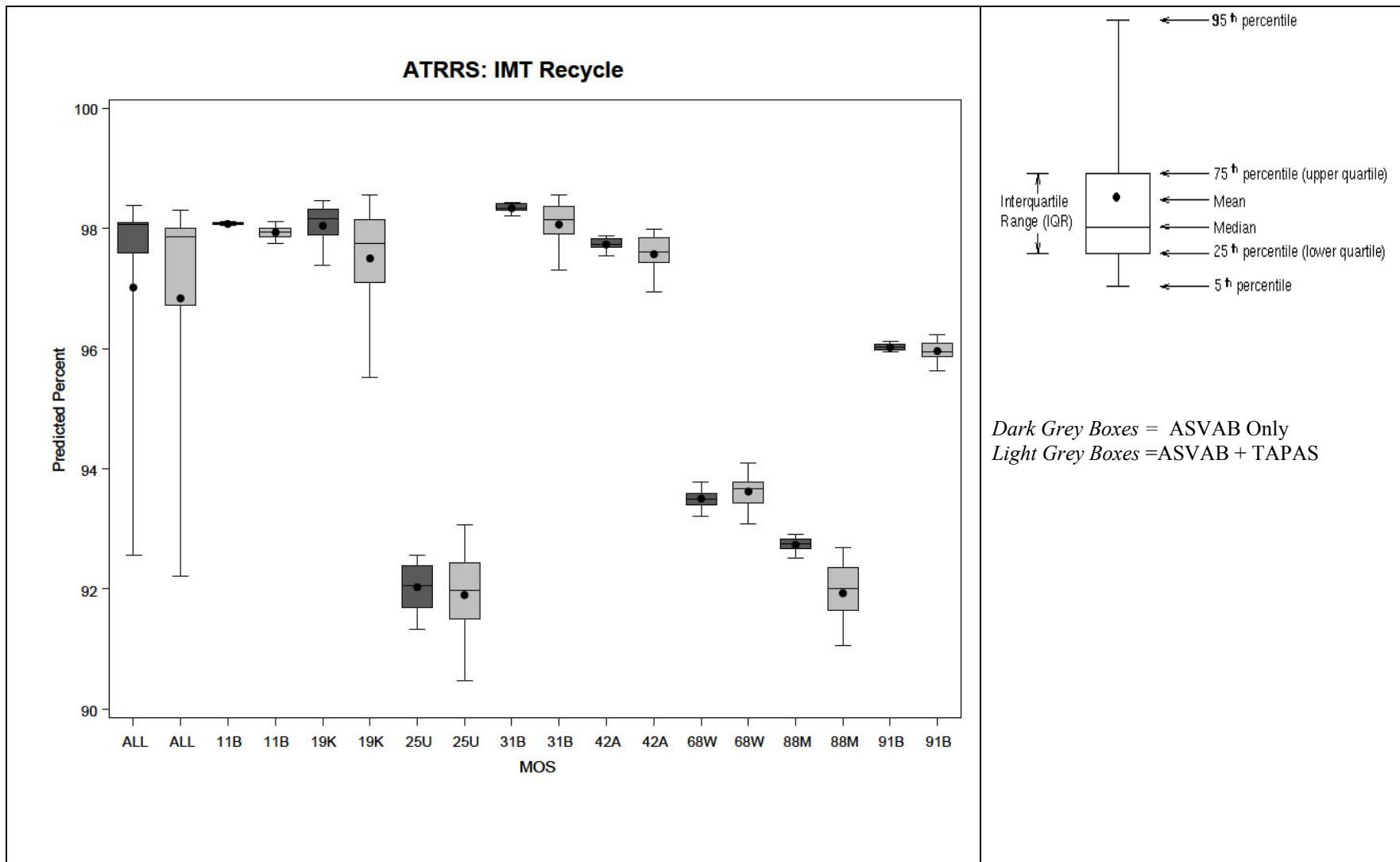
Figure D.6. Boxplot of the classification potential of the TAPAS relative to the ASVAB for minimizing IMT Recycle.

Table D.6. Classification Potential of the TAPAS Relative to the ASVAB for Minimizing IMT Recycle

		IMT Recycle (% with a Recycle)	
	Obs	*ASVAB Only* *(5%, 95%)*	*ASVAB + TAPAS* *(5%,95%)*
Overall	13.8	12.6	11.7
	--	(3.4, 17.8)	(2.2, 18.0)
11B	15.5	15.1	15.4
	--	(14.5, 15.6)	(13.5, 17.3)
19K	15.0	10.3	**5.5**
	--	(8.9, 11.5)	**(3.9, 7.8)**
25U	14.5	10.0	8.4
	--	(8.5, 11.1)	(6.3, 10.4)
31B	15.0	13.3	11.6
	--	(12.4, 14.2)	(9.4, 13.6)
42A	4.4	3.5	**2.2**
	--	(3.1, 3.8)	**(1.6, 2.8)**
68W	16.1	17.8	17.8
	--	(17.2, 18.4)	(17.1, 18.7)
88M	10.9	10.4	**8.9**
	--	(10.1, 10.7)	**(8.0, 9.8)**
91B	9.4	9.1	8.5
	--	(8.7, 9.7)	(7.6, 9.3)

Note. Obs = Observed mean predicted criterion score by MOS or for the target MOS overall. ASVAB Only = Mean predicted criterion score when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Mean predicted criterion score when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (MOS-Specific JKT, 11B = 46%, 19K = 2%, 31B=11%, 68W=14%, 88M = 14%, 91B = 12%; Restarted IMT, 11B = 42%, 19K=2%, 25U=4%, 31B=10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%; Last APFT Score, 11B = 44%, 19K=2%, 31B=11%, 42A =6%, 68W=13%, 88M = 13%, 91B =12%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, *n* = 23,058; 19K, *n* = 1,236; 25U, *n* = 2,297; 31B, *n* = 5,691; 42A, *n* = 2,979; 68W, *n* = 6,902; 88M, *n* = 6,731; 91B, *n* = 6,158).
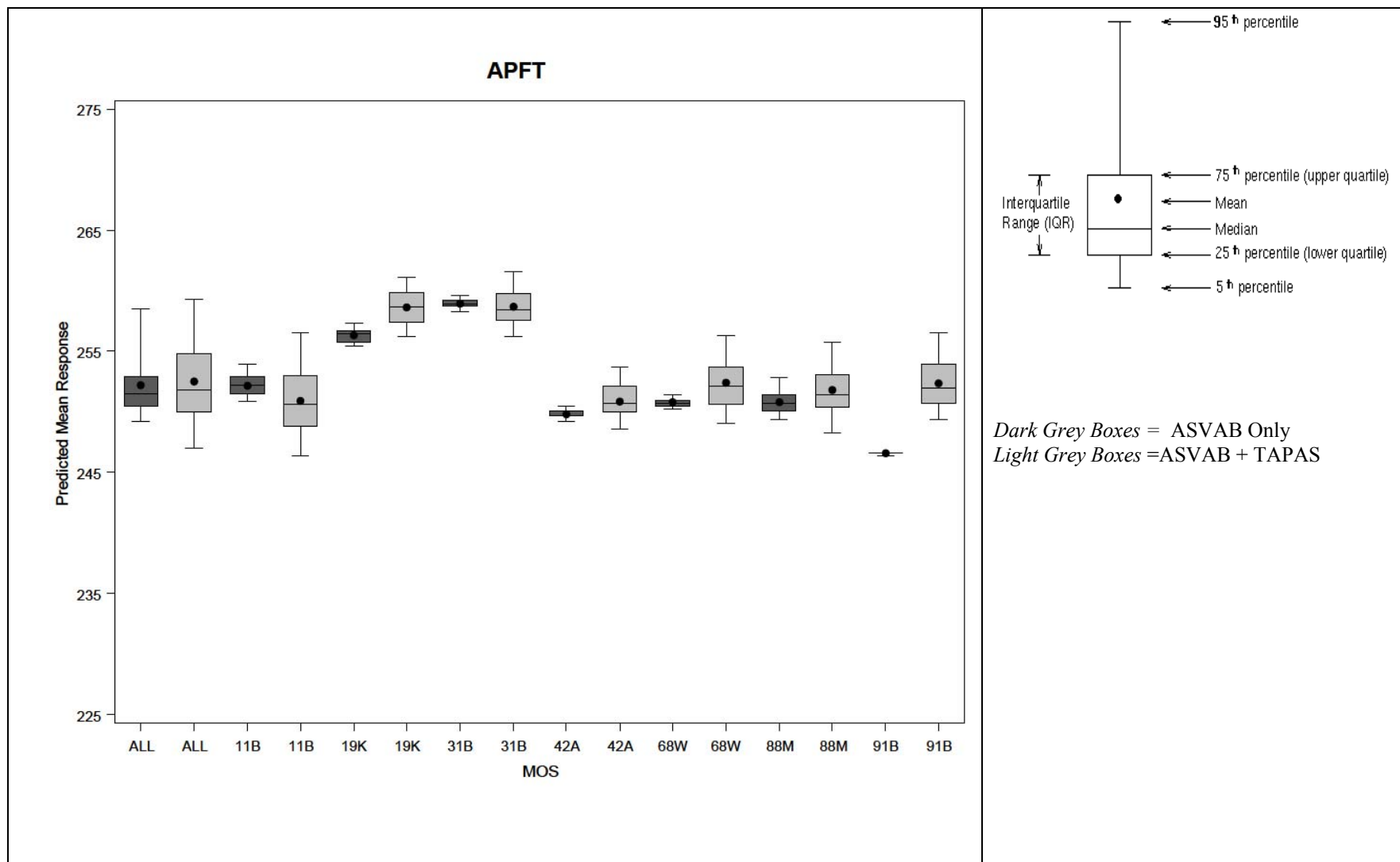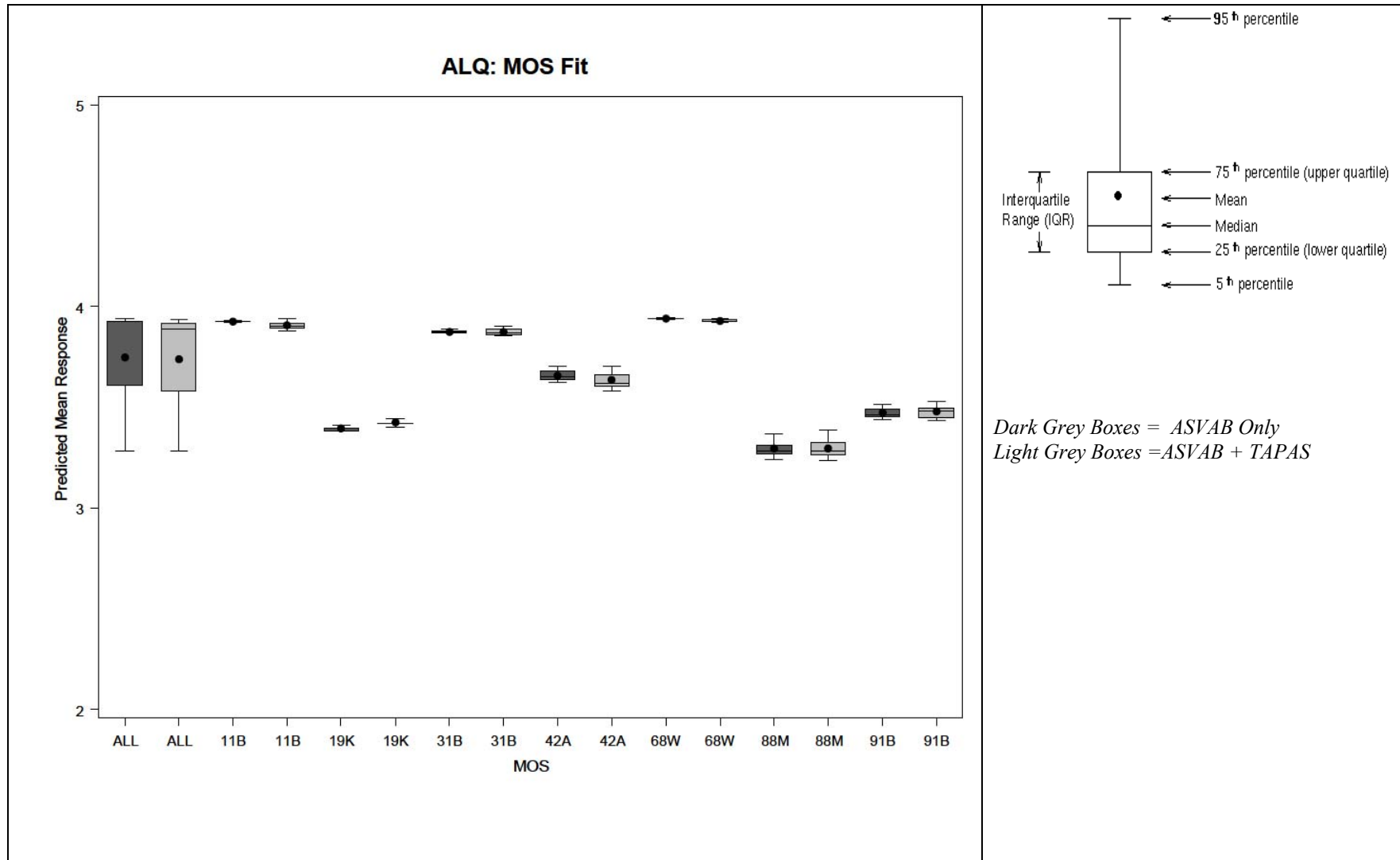
Figure D.7. Boxplot of the classification potential of the TAPAS relative to the ASVAB for maximizing Last Army Physical Fitness Test (APFT) score.

Table D.7. Classification Potential of the TAPAS Relative to the ASVAB for Maximizing Last Army Physical Fitness Test (APFT)

		Last APFT Score (100-300)	
	Obs	*ASVAB Only (5%, 95%)*	*ASVAB + TAPAS (5%,95%)*
Overall	251.6	252.2	252.5
	--	(249.2, 258.5)	(247.0, 259.3)
11B	250.5	252.2	250.9
	--	(250.8, 253.9)	(246.4, 256.6)
19K	253.9	256.3	258.6
		(255.5, 257.3)	(256.2, 261.1)
25U	--	--	--
	--	--	--
31B	258.4	258.9	258.7
	--	(258.3, 259.6)	(256.2, 261.6)
42A	247.6	249.8	250.9
	--	(249.2, 250.4)	(248.5, 253.6)
68W	251.4	250.8	252.4
	--	(250.2, 251.4)	(249.1, 256.2)
88M	248.2	250.8	251.8
	--	(249.3, 252.8)	(248.2, 255.7)
91B	245.0	246.6	252.4
	--	(246.4, 246.6)	(249.4, 256.5)

Note. Obs = Observed mean predicted criterion score by MOS or for the target MOS overall. ASVAB Only = Mean predicted criterion score when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Mean predicted criterion score when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (MOS-Specific JKT, 11B = 46%, 19K = 2%, 31B=11%, 68W=14%, 88M = 14%, 91B = 12%; Restarted IMT, 11B = 42%, 19K=2%, 25U=4%, 31B=10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%; Last APFT Score, 11B = 44%, 19K=2%, 31B=11%, 42A =6%, 68W=13%, 88M = 13%, 91B =12%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, n = 23,058; 19K, n = 1,236; 25U, n = 2,297; 31B, n = 5,691; 42A, n = 2,979; 68W, n = 6,902; 88M, n = 6,731; 91B, n = 6,158).

Figure D.8. Boxplot of the classification potential of the TAPAS relative to the ASVAB for minimizing ALQ MOS Fit.

Table D.8. Classification Potential of the TAPAS Relative to the ASVAB for Maximizing MOS Fit

		MOS Fit (1 = Low Fit-5 = High Fit)	
	Obs	*ASVAB Only (5%, 95%)*	*ASVAB + TAPAS (5%,95%)*
Overall	3.80	3.75	3.74
	--	(3.28, 3.94)	(3.28, 3.93)
11B	3.92	3.92	3.91
	--	(3.92, 3.93)	(3.88, 3.94)
19K	3.42	3.39	3.42
	--	(3.38, 3.41)	(3.40, 3.44)
25U	--	--	--
	--	--	--
31B	3.87	3.87	3.87
	--	(3.87, 3.89)	(3.85, 3.90)
42A	3.69	3.65	3.63
	--	(3.62, 3.70)	(3.58, 3.70)
68W	3.96	3.94	3.93
	--	(3.94, 3.94)	(3.92, 3.94)
88M	3.22	3.29	3.29
	--	(3.24, 3.37)	(3.23, 3.39)
91B	3.51	3.47	3.48
	--	(3.44, 3.51)	(3.43, 3.53)

Note. Obs = Observed mean predicted criterion score by MOS or for the target MOS overall. ASVAB Only = Mean predicted criterion score when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Mean predicted criterion score when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (IMT MOS-Specific JKT, 11B = 46%, 19K = 2%, 31B = 11%, 68W = 14%, 88M = 14%, 91B = 12%; Restarted IMT, 11B = 42%, 19K = 2%, 25U = 4%, 31B=10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%; Last APFT Score, 11B = 44%, 19K = 2%, 31B = 11%, 42A = 6%, 68W=13%, 88M = 13%, 91B =12%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, n = 23,058; 19K, n = 1,236; 31B, n = 5,691; 42A, n = 2,979; 68W, n = 6,902; 88M, n = 6,731; 91B, n = 6,158).
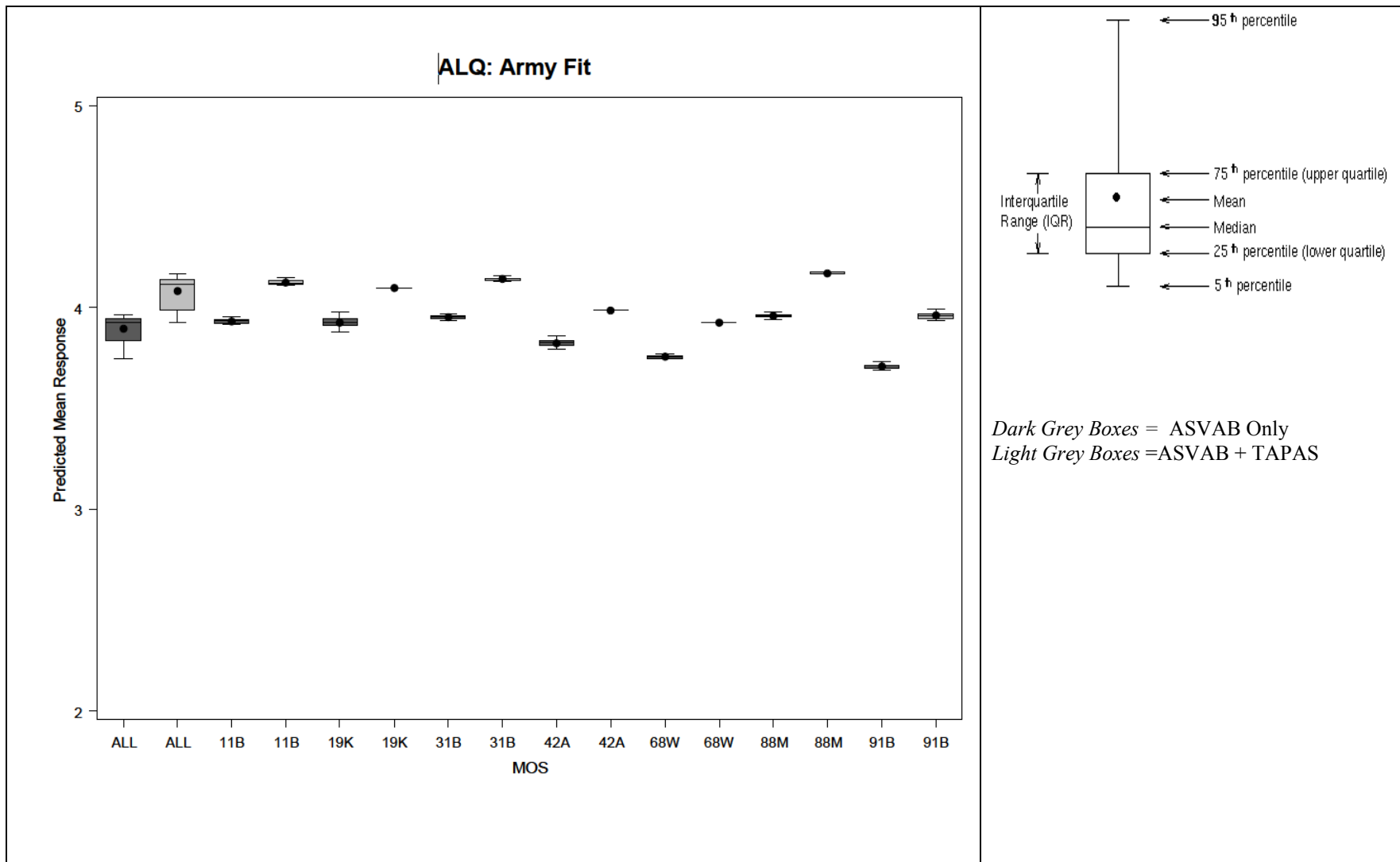
Figure D.9. Boxplot of the classification potential of the TAPAS relative to the ASVAB for minimizing Army Life Questionnaire (ALQ) Army Fit.

Table D.9. Classification Potential of the TAPAS Relative to the ASVAB for Maximizing Army Fit

		Army Fit (1 = Low Fit-5 = High Fit)	
	Obs	*ASVAB Only (5%, 95%)*	*ASVAB + TAPAS (5%,95%)*
Overall	4.08	3.89	4.08
	--	(3.75, 3.96)	(3.92, 4.17)
11B	4.11	3.93	**4.12**
	--	(3.91, 3.96)	**(4.11, 4.15)**
19K	4.07	3.92	**4.10**
	--	(3.88, 3.98)	**(4.10, 4.10)**
25U	--	--	--
	--	--	--
31B	4.13	3.95	**4.14**
	--	(3.93, 3.97)	**(4.13, 4.15)**
42A	4.04	3.82	**3.99**
	--	(3.79, 3.86)	**(3.99, 3.99)**
68W	3.92	3.75	**3.92**
	--	(3.74, 3.77)	**(3.92, 3.93)**
88M	4.16	3.96	**4.17**
	--	(3.94, 3.98)	**(4.16, 4.18)**
91B	3.87	3.71	**3.96**
	--	(3.69, 3.73)	**(3.93, 3.99)**

Note. Obs = Observed mean predicted criterion score by MOS or for the target MOS overall. ASVAB Only = Mean predicted criterion score when the ASVAB was used to classify Soldiers into the MOS listed. ASVAB + TAPAS = Mean predicted criterion score when the ASVAB and TAPAS were used to classify Soldiers into the MOS listed. Bolded values mean that the predicted attrition rates for the ASVAB + TAPAS do not overlap with the predicted rates for the ASVAB Only. Allocation percentages were based on the number of Soldiers in each MOS in the Accession Sample (IMT MOS-Specific JKT, 11B = 46%, 19K = 2%, 31B = 11%, 68W = 14%, 88M = 14%, 91B = 12%; Restarted IMT, 11B = 42%, 19K = 2%, 25U = 4%, 31B=10%, 42A =5%, 68W=13%, 88M = 12%, 91B =11%; Last APFT Score, 11B = 44%, 19K = 2%, 31B = 11%, 42A = 6%, 68W=13%, 88M = 13%, 91B =12%). Estimates based on Soldiers in the Validation Sample with non-missing predictor data (11B, n = 23,058; 19K, n = 1,236; 31B, n = 5,691; 42A, n = 2,979; 68W, n = 6,902; 88M, n = 6,731; 91B, n = 6,158).