



**Research Note 2022-01**

**Validation of Accessions Screening Tools:  
Cycle 20 Evaluation**

**Deirdre J. Knapp**

Human Resources Research Organization

**Cristina D. Kirkendall**

U.S. Army Research Institute

**October 2021**

**United States Army Research Institute  
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited.

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

**Department of the Army  
Deputy Chief of Staff, G1**

**Authorized and approved:**

**MICHELLE L. ZBYLUT, Ph.D.  
Director**

---

Research accomplished under contract  
for the Department of the Army by:

Human Resources Research Organization

Technical Review by:

Colin Omori, U.S. Army Research Institute  
Michelle Kim, U.S. Army Research Institute

**DISPOSITION**

This Research Note has been submitted to the  
Defense Technical Information Center (DTIC).

## REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) October 2021		2. REPORT TYPE Interim		3. DATES COVERED (from. . . to) April 2019 to December 2020	
4. TITLE AND SUBTITLE  Validation of Accessions Screening Tools: Cycle 20 Evaluation			5a. CONTRACT OR GRANT NUMBER W911NF-19-C-0032		
			5b. PROGRAM ELEMENT NUMBER 633007		
6. AUTHOR(S) Deirdre J. Knapp & Cristina D. Kirkendall			5c. PROJECT NUMBER A792		
			5d. TASK NUMBER 1029		
			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 6 <sup>th</sup> Street (Bldg 1464/Mail Stop 5610) Ft. Belvoir, VA 22060-5610			10. MONITOR ACRONYM ARI		
			11. MONITOR REPORT NUMBER Research Note 2022-01		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES ARI Research POC: Dr. Cristina D. Kirkendall, Selection and Assignment Research Unit					
14. ABSTRACT ( <i>Maximum 200 words</i> ): 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 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), beginning administration to applicants in 2009. Since that time criterion data were compiled at 6-month intervals from administrative records, schools for selected military occupational specialties (MOS), and Soldiers in units. This is the 20th in a series of biannual evaluations of TAPAS and the second that has been done under a talent management support initiative known as <i>Validation of Accessions Screening Tools (VAST)</i> . The cumulative results suggest that TAPAS scores significantly predict a number of criteria of interest (e.g., attrition, Army fit), indicating that the measure holds promise for both selection and classification purposes.					
15. SUBJECT TERMS Behavioral and social science                      Personnel                      Manpower                      Selection and classification					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT Unlimited Unclassified	20. NUMBER OF PAGES  56	21. RESPONSIBLE PERSON  Tonia S. Heffner 703-545-4408
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified			

**Research Note 2022-01**

**Validation of Accessions Screening Tools:  
Cycle 20 Evaluation**

**Deirdre J. Knapp**  
Human Resources Research Organization

**Cristina D. Kirkendall**  
U.S. Army Research Institute

**Selection and Assignment Research Unit**  
**Tonia S. Heffner, Chief**

**October 2021**

---

Approved for public release; distribution is unlimited.

## ACKNOWLEDGEMENT

---

We would like to acknowledge the Army cadre who support criterion data collection efforts at the schoolhouses. These noncommissioned officers (NCOs) ensure that trainees are scheduled to take the research measures and provide ratings of their Soldiers' performance in training. Those Army personnel who support the in-unit data collections are also instrumental to this research program.

We also extend our appreciation to the Army Test Program Advisory Team (ATPAT), a group of senior NCOs who periodically meet with ARI researchers to help guide this work in a manner that ensures its relevance to the Army and who help enable the Army support required to implement the research. Members of the ATPAT are:

Mr. Thomas Baxter	SFC Ricardo Martinez
MSG Andre Betton	CSM Shawn McDonald
Mr. Shawn Bova	SFC Samuel Meadows
SGM Daniel Colonmendez	SGM Jeffrey Morris
SFC Randal Davenport	CW5 Anthony Moschella
MSG Christopher Evans	SFC Anthony Northcutt
SFC Merhil Figueroa	SGM Nicci Onugha
SFC Jose Flores	MSG Todd Pameticky
SGM Bryant Fowle	SGM Terry Pittman
Mr. Alphonsa Green	Mr. John Plotts
SGM Latoya Hammond	Mr. Gerald Purcell
COL James Hayes	MSG Joel Raglin
SGM Robert Haynie	CSM Latosha Ravenell
Mr. Lawrence Hayward	MSG Felicia Redd
1SG Clifford Hill	MSG Derrian Richardson
1SG Thomas Hill	MSG Timothy Robinson
MSG Kaulai Hollis	Mr. Mark Russo
MSG Richard Howe	SFC Fred Scaldaferrri
CSM TV Jackson	CSM Todd Shirley
SFC Saba James	SGM Thomas Skelton
SGM Louis Johnson	SFC Marvin Skinner
MSG Chad Kenney	1SG David Smith
SGM Gareth Kilpatrick	1SG Christopher Stanifer
SFC Clifford Kurten	SGM Michael Steward
SGM Alfredo Lassiseramirez	SFC Ronnie Vengua
Mr. Lamar Lauderdale	MSG Wesley Weaver
CSM Timothy Lawless	Ms. Natalya Weber

# VALIDATION OF ACCESSIONS SCREENING TOOLS: CYCLE 20 EVALUATION

## 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 AFQT has proven to be and will continue to serve as a useful metric for selecting new Soldiers, there is a growing recognition of the need to consider whole-person assessment that takes other personal attributes, in particular non-cognitive attributes (e.g., temperament, interests, and values) into consideration. Non-cognitive attributes 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)*. TAPAS capitalizes on the latest advances in testing technology to assess motivation through the measurement of personality characteristics. The TOPS IOT&E evolved into a long-term personnel database effort that supports empirically-based policy-making across a wide range of talent accession decisions, including those related to the operational use of TAPAS. This broader-based effort is now known as the *Validation of Accessions Screening Tools (VAST)* research program.

### Procedure:

In May 2009, the U.S. Army began administering TAPAS at Military Entrance Processing Stations (MEPS). To validate TAPAS, ARI began collecting outcome (criterion) data at multiple points in time from Soldiers who took TAPAS at entry. At present, initial military training (IMT) criterion data are being collected at schools for Soldiers in 71 military occupational specialties (MOS). Project teams also collect criterion data from in-unit Soldiers (regardless of MOS) in annual waves of site visits to Army installations.

Starting with an abstract of the Total Army Personnel Database (TAPDB), researchers biannually match ASVAB and TAPAS scores with criterion data collected specifically for ARI as well as data from administrative sources (e.g., course grades). The ARI measures include job knowledge tests, an attitudinal assessment (the Army Life Questionnaire), and performance rating scales completed by the Soldiers' supervisors (in units) and their peers (in both IMT and in units).

The data presented in this report come from TAPAS data collected through October 2019 and criterion data collected through December 2019. The TAPAS database consists of over 1.2M applicants. The validation sample sizes are considerably smaller, with the IMT Validation Sample comprising 94,049 Soldiers, the In-Unit Validation Sample comprising 9,207 Soldiers, and the Administrative Validation Sample (which includes Soldiers with criterion data [e.g., attrition] from at least one administrative source) comprising 731,133 Soldiers.

Data from job knowledge tests, attitudinal assessment, performance ratings provided by supervisors and/or peers, and administrative sources were combined to yield an array of scores representing important Soldier outcomes. In general, the criterion scores exhibited acceptable and theoretically consistent psychometric properties. The exception to this conclusion is the performance rating scales which exhibit relatively low inter-rater reliability. As a result, findings involving the rating scale scores may underestimate relationships with other variables.

Our approach to analyzing the incremental predictive validity of TAPAS over AFQT was consistent with previous evaluations of this measure and similar experimental non-cognitive predictors (e.g., Ingerick et al., 2009; Knapp & Heffner, 2009, 2010, 2011). In brief, this approach involved testing a series of hierarchical regression models, regressing scores for each criterion measure onto Soldiers' AFQT scores, followed by their TAPAS composite or TAPAS scale scores in the second step. The resulting increment in the multiple correlation value ( $\Delta R$ ) when TAPAS composite or scale scores were added to the baseline regression models served as our index of incremental validity. Analyses used the TAPAS Will-Do (motivation), Can-Do (knowledge), and Adaptation composite scores.

#### Findings:

Results presented here align with the findings from prior analysis cycles (e.g., Hughes, 2018). Overall, the TAPAS composites demonstrated incremental validity over AFQT in predicting several important first-term Soldier outcomes. In regard to prediction of continuous criterion measures, the TAPAS Will-Do composite demonstrated the greatest incremental validity overall, with  $\Delta R$  estimates exceeding .05 for predicting peer performance ratings, Army Physical Fitness Test (APFT) scores, peer leadership, and resilience for Soldiers in IMT and in units, as well as commitment/fit and resilience for Soldiers in IMT. The TAPAS Adaptation composite also yielded incremental validity estimates of .05 or greater for peer performance ratings, APFT scores, and resilience for Soldiers in IMT and APFT scores for Soldiers in units. Both the Will-Do and Adaptation composites also demonstrated significant incremental relationships beyond AFQT with all of the dichotomous criteria (i.e., attrition, disciplinary incidents, and IMT restarts). Higher scores on these composites were associated with a reduction in the likelihood of these outcomes. Conversely, the TAPAS Can-Do composite did not provide any notable incremental validity for the criteria examined in this research. This finding is not surprising given the established strength of AFQT to predict the same criteria as the Can-Do composite was designed to predict.

#### Utilization and Dissemination of Findings:

The VAST research findings will be used by the Assistant Secretary of the Army (Manpower and Reserve Affairs); Army Deputy Chief of Staff, G-1; U.S. Army Training and Doctrine Command; U.S. Army Recruiting Command; and to evaluate the effectiveness of tools used for Army applicant selection and assignment.

**VALIDATION OF ACCESSIONS SCREENING TOOLS:  
CYCLE 20 EVALUATION**

**CONTENTS**

---

	Page
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>1</b>
Background.....	1
Evaluating TAPAS.....	3
Overview of Report.....	3
<b>CHAPTER 2: RESEARCH METHOD .....</b>	<b>4</b>
Overview.....	4
Predictor Measures.....	4
Criterion Measures.....	7
Data Collection .....	18
Biannual Longitudinal Database Updates.....	19
Analysis Plan Overview.....	20
<b>CHAPTER 3: CYCLE 20 SAMPLE CHARACTERISTICS AND DESCRIPTIVE     STATISTICS.....</b>	<b>21</b>
Description of Analysis Samples.....	21
Descriptive Statistics for Predictor Measures .....	23
Summary .....	26
<b>CHAPTER 4: EVIDENCE FOR THE PREDICTIVE VALIDITY OF THE TAPAS .....</b>	<b>27</b>
Incremental Validity Analyses.....	27
Summary .....	38
<b>REFERENCES.....</b>	<b>39</b>

**APPENDICES**

<b>APPENDIX A - LIST OF FREQUENTLY USED ACRONYMS.....</b>	<b>A-1</b>
---	------------

<b>APPENDIX B - BIVARIATE CORRELATIONS AMONG STUDY VARIABLES .....</b>	<b>B-1</b>
--	------------

**CONTENTS (CONTINUED)**

---

Page

**LIST OF TABLES**

Table 2.1. TAPAS Facet Names and Definitions .....5  
Table 2.2. IMT Criterion Scores .....7  
Table 2.2. IMT Criterion Scores .....8  
Table 2.3. In-Unit Criterion Scores.....8  
Table 2.4. Attrition Criterion Scores.....9  
Table 2.5. Army Life Questionnaire Scale and Composite Scores.....11  
Table 2.6. IMT Peer Performance Rating Scale Dimensions and Associated Example Behaviors .....12  
Table 2.7. In-Unit Peer Performance Rating Scale Dimensions and Associated Example Behaviors .....14  
Table 2.8. Peer Performance Rating Scale Interrater Reliability Estimates .....15  
Table 2.9. In-Unit PRS-S Army-Wide Rating Scale Dimensions and Associated Definitions.....16  
  
Table 3.1. Background and Demographic Characteristics of the Analysis Samples .....22  
Table 3.2. Means and Standard Deviations for the TAPAS Composites by TAPAS Version .....24  
Table 3.3. Descriptive Statistics for Criterion Scores in the IMT Validation Sample.....24  
Table 3.4. Descriptive Statistics for Criterion Scores in the In-Unit Validation Sample .....25  
Table 3.5. Base Rates for Regular Army Attrition in the Administrative Validation Sample.....25  
  
Table 4.1. Incremental Validity Estimates for TAPAS over AFQT for Predicting IMT Criteria .....28  
Table 4.2. Incremental Validity Estimates for TAPAS over AFQT for Predicting In-Unit Adjustment, Commitment and Fit, Fitness, and Retention Criteria.....31  
Table 4.3. Incremental Validity Estimates for TAPAS over AFQT for Predicting IMT Restarts and Disciplinary Incidents .....34  
Table 4.4. Incremental Validity Estimates for TAPAS over AFQT for Predicting Cumulative Attrition through 36 Months of Service .....35

**CONTENTS (CONTINUED)**

---

Page

**LIST OF FIGURES**

Figure 2.1. Sample IMT peer rater performance rating scales. ....13

Figure 2.2. Sample in-unit supervisor rater performance rating scale. ....17

Figure 2.3. Overview of VAST data file merging and nested sample generation process. ....19

Figure 4.1. Increase in prediction of IMT criteria using TAPAS Will-Do scales. ....30

Figure 4.2. Increase in prediction of in-unit criteria using the TAPAS Will-Do scales. ....32

Figure 4.3. Predictive accuracy of AFQT and TAPAS Will-Do composite in the  
discrimination of both IMT and in-unit Disciplinary Incidents and IMT Restarts. ....36

Figure 4.4. Predictive accuracy of AFQT and TAPAS Adaptation composite in the  
discrimination of attrition outcomes for Regular Army Soldiers. ....37

# VALIDATION OF ACCESSIONS SCREENING TOOLS: CYCLE 20 EVALUATION

## CHAPTER 1: INTRODUCTION

### Background

The Selection and Assignment Research Unit (SARU) 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 SARU'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.<sup>1</sup> Although 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 their 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 was a six-year effort conducted with contract support from the Human Resources Research Organization ([HumRRO]; Allen, Knapp, & Owens, 2013; Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009). Experimental predictors were administered to new Soldiers in 2007 and early 2008. Army Class collected school-based criterion data on a subset of the Soldier sample as they completed job training. Job performance criterion data were collected from Soldiers in the Army Class longitudinal validation sample in 2009 with a second round of data collection in Soldiers' units completed in April 2011 (Knapp, Owens, & Allen, 2012). Final analysis and reporting of this program of research can be found in Allen et al. (2013).

After the Army Class research began, 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

---

<sup>1</sup> A list of defined acronyms commonly used in this report is provided in Appendix A.

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

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), and
- Minimal administration time requirements.

The Tailored Adaptive Personality Assessment System ([TAPAS]; Stark, Chernyshenko, & Drasgow, 2010) surfaced as the top choice.<sup>2</sup> 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.

In May 2009, the Army began administering TAPAS at Military Entrance Processing Stations (MEPS) on the computer adaptive platform for the ASVAB (CAT-ASVAB) as part of the Tier One Performance Screen (TOPS) initial operational test and evaluation (IOT&E). Initially, TAPAS was to be administered only to Education Tier 1, non-prior service applicants.<sup>3</sup> This limitation to Education Tier 1 was removed early in CY2011 to allow the Army to evaluate TAPAS across all types of applicants.

The TOPS IOT&E investigated the potential for non-cognitive measures to identify applicants who would likely perform differently (higher or lower) than would be predicted by their ASVAB scores. The initial conceptualization for the IOT&E was to use TAPAS as a tool for “screening in” Education Tier 1 applicants with lower AFQT scores, allowing more of these applicants to access into the Army, hence the research program title (Tier One Performance Screen). However, changing economic conditions spurred a reconceptualization that led to using TAPAS as a tool to screen out low-motivated applicants, thus making the selection criteria to enter the Army more stringent. Subsequent decision rules regarding the derivation and use of TAPAS composite scores have been adjusted in myriad ways to reflect evolving market conditions.

In practice, as part of the TOPS IOT&E, TAPAS scores were originally used to screen out a small number of AFQT Category IIIB/IV applicants.<sup>4</sup> In 2014, Tier 2 applicants who scored below the 30<sup>th</sup> percentile on TAPAS were also screened out. In 2015, the TOPS IOT&E was

---

<sup>2</sup> Other promising assessments included the Work Preferences Assessment ([WPA]; Putka & Van Iddekinge, 2007) and the Information/Communications Technology Literacy test ([ICTL]; Russell & Sellman, 2009).

<sup>3</sup> Applicant educational credentials are classified as Tier 1 (primarily high school diploma), Tier 2 (primarily non-diploma graduate), and Tier 3 (not a high school graduate).

<sup>4</sup> 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).

modified to allow all Tier 1 Category IIIB applicants to enlist regardless of TAPAS score. Most recently (at the beginning of 2017), the Army temporarily suspended use of TAPAS for operational applicant screening, though TAPAS continues to be administered to most Army applicants.

## **Evaluating TAPAS**

Evaluation of TAPAS, which initially began as part of the TOPS IOT&E project described above, is being continued in the current research effort, *Validation of Accessions Screening Tools* (VAST). As part of the TAPAS evaluation, ARI collects training criterion data on Soldiers in multiple military occupational specialties (MOS) as they complete initial military training (IMT). This originally involved a target group of eight large, highly critical MOS as well as MOS chosen to represent the diversity of job requirements across the Army. Over time, additional MOS were added, and the current target group includes 71 MOS. The criterion measures include job knowledge tests (JKTs); an attitudinal assessment, the Army Life Questionnaire (ALQ); and performance rating scales (PRS). These measures are computer-administered at the end of IMT for each of the target MOS. The process is overseen by Army personnel with guidance and support from both ARI and HumRRO. Administrative records are obtained for all Soldiers who took TAPAS, regardless of MOS. Although data are collected from many MOS, in most cases sample sizes are too small for MOS-specific analysis.

Criterion data are also collected from in-unit Soldiers and their supervisors during data collection at major Army installations throughout the course of the TAPAS validation effort. These proctored “in-unit” data collections began in January 2011 and target Soldiers who took TAPAS prior to enlistment. The in-unit criterion measures include JKTs, the ALQ, supervisor ratings of performance, and peer ratings of performance. Separation status of all Soldiers who took TAPAS prior to enlistment is tracked throughout the course of the research.

This report is the second evaluation cycle of the VAST research but serves as the 20th iteration of the TAPAS criterion-related validation that began with the TOPS IOT&E. Prior evaluations are described in a series of technical reports (Bynum & Mullins, 2017; Knapp & Heffner, 2011, 2012; Knapp, Heffner, & White, 2011; Knapp & Kirkendall, 2018, 2020; Knapp & LaPort, 2013a, 2013b, 2014; Knapp & Wolters; 2017) and internal memoranda.

## **Overview of Report**

Chapter 2 describes the VAST research method, including the predictor and criterion measures, data collection procedures, database construction strategy, and the basic data analysis plan. Chapter 3 describes characteristics of the most recent bi-annual evaluation sample and provides descriptive statistics of the scores used in the validation analyses reported in Chapter 4.

## CHAPTER 2: RESEARCH METHOD<sup>5</sup>

### Overview

The VAST research uses a longitudinal criterion-related validation evidence design. Army applicants take the ASVAB and TAPAS measures as part of the usual pre-enlistment accessioning process. We track all those who enlist by, at a minimum, retrieving available administrative outcome data (e.g., separation status) from Army records. We also seek to collect for-research-only criterion data from many of these Soldiers as they complete IMT by having ongoing data collection activity at a large number of MOS schoolhouses. Criterion data are collected from Soldiers who have successfully moved on to their assigned units through annual visits to various Army installations. We refer to these as the IMT and in-unit data collections.

Twice per year, we update the VAST longitudinal database with newly collected IMT and in-unit data, as well as updated administrative data. After each new iteration of the database becomes available, we analyze the data and document the results. Carrying forward the nomenclature from the TOPS research, validation analyses described here are based on the 20<sup>th</sup> cycle of biannual database updates.

### Predictor Measures

#### *Armed Forces Qualification Test*

The ASVAB is a multiple aptitude battery of tests administered by the Military Entrance Processing Command (MEPCOM). Most military applicants take the computer adaptive version of ASVAB. Scores on the ASVAB tests are combined to create composite scores for use in selecting applicants into the Army and qualifying them for various MOS. AFQT, the composite used for selecting applicants into the Army, comprises Verbal Expression<sup>6</sup> (VE), Arithmetic Reasoning (AR), and Math Knowledge (MK) scores ( $AFQT = 2*VE + AR + MK$ ). It is designed to be a measure of general cognitive ability (Maier, 1993) and has strong psychometric properties in terms of reliability and criterion-related validity (e.g., Campbell & Knapp, 2001; Carretta, Ree, & Teachout, 2016; DMDC, 2008; Hunter, 1986; Thacker, Russell, Knapp, & Gribben, 2020).

AFQT percentile scores are divided into the following categories:<sup>7</sup>

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

---

<sup>5</sup> The report authors would like to acknowledge the efforts of Doug Brown (IMT data collection support), Chris Graves (in-unit data collection manager), Martin Yu (administrative data manager), and Arielle Rogers (VAST database construction manager) and their respective teams in conducting the work described in this chapter.

<sup>6</sup> Verbal Expression is a scaled combination of the Word Knowledge (WK) and Paragraph Comprehension (PC) tests.

<sup>7</sup> For more information on ASVAB scoring, see the official website of the ASVAB, [www.officialasvab.com](http://www.officialasvab.com).

Per DoD regulations, AFQT Category V Soldiers are not eligible for enlistment. Category IV accessions are greatly restricted, some restriction is placed on Category IIIB accessions, and priority is given to Category I-III A accessions.

***Tailored Adaptive Personality Assessment System (TAPAS)***

TAPAS consists of pools of statements descriptive of various facets of temperament (Drasgow, Stark, Chernyshenko, Nye, Hulin, & White, 2012). Items are created by pairing statements of comparable social desirability and extremity from different facet pools. Statements are selected and paired adaptively to triangulate on estimates of an examinee’s standing on the various facets.

Of the 26 facets described in Table 2.1, the three versions of TAPAS administered to Army applicants from 2013 to 2019 each assessed 13. Of those 13, 10 facets were common to all three forms.

***Table 2.1. TAPAS Facet 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."

**Table 2.1. (Continued)**

Facet Name	Brief Description
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.
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.
Tolerance	High scoring individuals are interested in other cultures and opinions that may differ from their own. They are willing to adapt to novel environments and situations.
Virtue	High scoring individuals strive to adhere to standards of honesty, morality, and “good Samaritan” behavior.

*Note.* Table reprinted from Stark, Chernyshenko, Nye, and Drasgow (2018).

Three regression-weighted TAPAS composites designed to predict various outcomes are computed for purposes of the present research: (a) Can-Do, (b) Will-Do, and (c) Adaptation. TAPAS versions 7 and 8 did not include all the facets comprising the Can-Do and Adaptation composites. This is reflected in generally smaller sample sizes for analyses involving the Can-Do and Adaptation composites.

Given the forced-choice nature of the TAPAS response format, traditional methods for estimating reliability of measurement are not suitable. Therefore, alternative estimates of reliability have been examined for the TAPAS. For example, Zhang et al. (2020) examined the IRT-based marginal reliability estimates for a static version of the TAPAS. Across 10 TAPAS facets, these estimates ranged from .66 to .86. Alternatively, test-retest reliability estimates

ranged from .69 to .83 in this same study. Similarly, additional research (Wolters et al., 2017) confirmed these results with several additional TAPAS facets. In two separate samples, the test-retest reliability estimates ranged from .69 to .83. Importantly, this research also examined the test-retest reliabilities of the TAPAS composites and found estimates ranging from .76 to .78. Although these studies suggest that the TAPAS has moderate reliability estimates, more research is needed. All of the studies described above were with civilian samples, using static versions of the TAPAS (i.e., rather than a computer adaptive version), and conducted in low-stakes (i.e., non-operational) settings. In contrast, studies with military samples have found somewhat lower test-retest reliabilities with estimates ranging from .49 to .74 (average of .61) across eight TAPAS facets and .56, .74, and .70, respectively, for the Can-Do, Will-Do, and Adaptation composites (Wolters et al., 2017). Therefore, more research is needed to evaluate the reliability of the TAPAS in military samples and under operational conditions.

### Criterion Measures

Although TAPAS was designed primarily to predict what might be characterized as motivational aspects of performance in the Army, this research uses a set of criterion measures that provides a more comprehensive picture of outcomes valued by the Army. Here we describe the criterion scores used in the present report (and summarized in Tables 2.2 – 2.4), which are based on tests of job knowledge, self-report attitudes, performance ratings provided by peers and supervisors, and relevant administrative indicators.

**Table 2.2. IMT Criterion Scores**

Criterion	Associated Scales	Description
Peer Performance Ratings: Overall Performance Composite	PRS-P Effort and Discipline PRS-P Physical Fitness PRS-P MOS Qualification Knowledge & Skill PRS-P Resilience and Adjustment PRS-P Working Effectively with Others PRS-P Going Above and Beyond PRS-P Overall Performance Rating	General effort/motivation criterion as rated by peers. Scales are averaged to form the composite.
Commitment and Fit	ALQ Affective Commitment ALQ MOS Fit ALQ Army Fit	General commitment to and fit with the Army. Scales are averaged to form the composite.
Retention Cognitions	ALQ Career Intentions ALQ Reenlistment Intentions	General intentions of continuance in the Army. Scales are averaged to form the composite.
Peer Leadership	ALQ Motivation to Lead-Affective ALQ Motivation to Lead-Noncalculative ALQ Motivation to Lead-Socionormative ALQ Organizational Citizenship Behavior/Leadership	General aptitude for effective peer leadership in the Army. Scales are averaged to form the composite.
Counterproductive Soldier Behavior	---	ALQ: Intentional behaviors that harm or are intended to harm another Soldier or the legitimate interests of the unit

**Table 2.2. IMT Criterion Scores**

Criterion	Associated Scales	Description
Resilience	---	ALQ: Capacity to overcome difficult life events with minimal disruption or long-term negative impacts
Knowledge & Skill	WTBD JKT MOS JKT AIT Grade	WTBD JKT and MOS JKT are averaged to form an overall knowledge/skill composite. For those that do not have an MOS JKT score, AIT grade is substituted when available.
WTBD JKT	---	Warrior Tasks and Battle Drills (Army-wide) job knowledge test
APFT Score	—	ALQ: Most recent Army Physical Fitness Test Score.
Disciplinary Incidents	—	ALQ: Dichotomous indicator of whether there is at least one “yes” response to a list of possible disciplinary incidents
IMT Restarts	—	Administrative: Indicates whether a Soldier restarted IMT at least once.

**Table 2.3. In-Unit Criterion Scores**

Criterion	Associated Scales	Description
Supervisor Performance Ratings: Overall Performance Composite	PRS-S Performing Core Warrior Tasks PRS-S Performing MOS-Specific Tasks PRS-S Communicating with Others PRS-S Processing Information PRS-S Solving Problems PRS-S Exhibiting Effort PRS-S Exhibiting Personal Discipline PRS-S Contributing to the Team PRS-S Exhibiting Fitness & Bearing PRS-S Adjusting to the Army PRS-S Following Safety Procedures PRS-S Developing Own Skills PRS-S Managing Personal Matters	General effort/motivation criterion as rated by supervisors. Scales are averaged to form the composite.
Peer Performance Ratings Overall Performance	PRS-P Effort and Discipline PRS-P Physical Fitness PRS-P MOS Qualification Knowledge & Skill PRS-P Processing Information & Solving Problems PRS-P Resilience and Adjustment PRS-P Working Effectively with Others PRS-P Going Above and Beyond PRS-P Emergent Leadership PRS-P Overall Performance Rating	General effort/motivation criterion as rated by peers. Scales are averaged to form the composite.

**Table 2.3. (Continued)**

Criterion	Associated Scales	Description
Commitment and Fit	ALQ Affective Commitment ALQ MOS Fit ALQ Army Fit	General commitment to and fit with the Army. Scales are averaged to form the composite.
Retention Cognitions	ALQ Career Intentions ALQ Reenlistment Intentions	General intentions of continuance in the Army. Scales are averaged to form the composite.
Peer Leadership	ALQ Motivation to Lead-Affective ALQ Motivation to Lead-Noncalculative ALQ Motivation to Lead-Socionormative ALQ Organizational Citizenship Behavior/Leadership	General aptitude for effective peer leadership in the Army. Scales are averaged to form the composite.
Counterproductive Soldier Behavior	---	ALQ: Intentional behaviors that harm or are intended to harm another Soldier or the legitimate interests of the unit.
Resilience	---	ALQ: Capacity to overcome difficult life events with minimal disruption or long-term negative impacts.
WTBD JKT	---	Warrior Tasks and Battle Drills (Army-wide) job knowledge test
APFT Score	—	ALQ: Most recent Army Physical Fitness Test Score.
Disciplinary Incidents	—	ALQ: Dichotomous indicator of whether there is at least one “yes” response to a list of possible disciplinary incidents.

**Table 2.4. Attrition Criterion Scores**

Criterion	Description
6-month	Attrition at or before 6 months of service
12-month	Attrition at or before 12 months of service
24-month	Attrition at or before 24 months of service
36-month	Attrition at or before 36 months of service

### ***Job Knowledge Tests (JKTs)***

All Soldiers graduating from schools participating in this research and Soldiers in the in-unit data collections are administered a selected response test of general Soldier knowledge (i.e., Warrior Tasks and Battle Drills [WTBD]). To support classification research, IMT Soldiers in 12 MOS also take an MOS-specific JKT. The JKT development followed the process established and documented over the course of several ARI-sponsored criterion development efforts, including Project A (Campbell & Knapp, 2001) and Army Class (Moriarty, Campbell, Heffner, & Knapp, 2009).

The JKTs include roughly 30 scored items. Items can be worth multiple points, depending on item type (e.g., multiple-choice, multiple response). We compute a single, overall raw score for each JKT by summing the total number of points earned across all items and computing a percent correct score based on the maximum number of points that could be obtained on each test. For subsequent analyses, we convert the total raw scores to standardized scores (or z-scores) by standardizing the scores *within* each MOS.

As estimated using the Cycle 17 evaluation database, the MOS-specific JKT reliability estimates are reasonably high (mean estimated reliability = 80.33; Graves & Taylor, 2020). The WTBD reliability estimates are lower (.68 for both IMT and in-unit samples). It is possible that this could reflect greater heterogeneity in content for the WTBD compared to the MOS-specific tests.

Final Advanced Individual Training (AIT) Grade is the cumulative grade across all courses administratively recorded for the Soldier in the RITMS (Resident Individual Training Management System) database. We compute a standardized version of Final AIT Grade by standardizing each course grade for courses with 15 or more Soldiers. We have not been able to access RITMS data for some time, so this variable is only available for older cases in the database.

The IMT validation analyses use a Knowledge and Skill composite score that is an average of the WTBD JKT and MOS-specific JKT scores. If there is no MOS-specific JKT score, AIT Grade is used instead. The composite is not computed if there are not two scores to combine.

### ***Army Life Questionnaire (ALQ)***

The ALQ asks Soldiers to report on their attitudes toward the Army and their MOS, indicate their intentions for their future with the Army, and provide information related to their Army service (e.g., awards, proficiency test scores). Because the ALQ is used for multiple ARI research efforts, additional content is included to support unrelated ARI research and is not described further here. The self-reported administrative information is discussed in a later section.

The ALQ was originally developed under the Select21 project (Knapp, Sager, & Tremble, 2005), though it had its origins in Project A (Campbell & Knapp, 2001). Earlier forms of the ALQ (Van Iddekinge, Putka, & Sager, 2005) were modified slightly for use in the TOPS IOT&E, then additional modifications were made during the course of the research to broaden the scope of the criterion space covered by the ALQ (Nesbitt, Salmon, & Kirkendall, 2020). There are separate IMT and in-unit versions of the ALQ, though both versions are highly similar.

Table 2.5 lists the individual ALQ scales (11 on the IMT ALQ and 13 on the in-unit ALQ), along with the number of items comprising each and the internal consistency estimates based on Cycle 17 evaluation data. Validation analyses reported in Chapter 4 use the three composite scores shown at the top of Table 2.5, as well as the Counterproductive Soldier Behavior and Resilience scale scores. See Nesbitt et al. (2020) for more detailed information about ALQ content and derivation of the three composite scores.

**Table 2.5. Army Life Questionnaire Scale and Composite Scores**

Variable	IMT		In-Unit	
	Alpha	Number Scales/Items	Alpha	Number Scales/Items
<b>Composites</b>				
Commitment and Fit <sup>a</sup>	.81	3	.82	3
Peer Leadership <sup>a</sup>	.82	4	.84	4
Retention Cognitions	.92	2	.94	2
<b>Likert Scales</b>				
Affective Commitment	.80	3	.80	3
Army Fit	.77	4	.79	4
Army Life Adjustment	.79	4	—	—
Counterproductive Soldier Behavior	.85	9	.79	9
Deployment Satisfaction	—	—	.81	4
MOS Fit	.86	5	.88	5
MOS Satisfaction	—	—	.84	5
Motivation to Lead - Affective	.84	3	.84	3
Motivation to Lead - Noncalculative	.78	3	.78	3
Motivation to Lead - Social Normative	.77	3	.78	3
Organizational Citizenship Behavior/Leadership	.87	8	.87	8
Resilience	.81	7	.79	7
Work/Life Balance	.87	5	.86	5

*Note.* Army Career Intentions and Reenlistment Intentions are single-item measures, and therefore alpha is not reported for these variables. Table reprinted from Nesbitt, Salmon, & Kirkendall (2020).

<sup>a</sup> Alphas reflect the average of the composite scale alphas.

### **Peer Performance Rating Scales (PRS-P)**

Along with administration of the JKT(s) and ALQ, Soldiers also rate each other's performance. The rating scales include a *not-observed* option on each rating scale and a single *familiarity* rating that allow raters to indicate their opportunity to observe each Soldier they are being asked to rate using a 4-point scale.

Indicate your opportunity to observe each Soldier you will be rating.

1. Not enough to judge the Soldier's performance.
2. Enough to judge some aspects of the Soldier's performance.
3. Enough to judge many aspects of the Soldier's performance.
4. Enough to judge most aspects of the Soldier's performance.

The online screen lists the behavioral dimensions to be rated and encourages raters to carefully read the name and definition of each dimension, rate typical performance, and distinguish performance across scales. Moreover, the following reminders appear on each rating screen:

- Read the whole description
- Base your rating on what you have seen. Use Not Observed/Cannot Rate when necessary
- Reflect on Soldiers’ typical performance including strengths and weaknesses

Raters evaluating multiple Soldiers rate each Soldier on a given performance dimension before moving to the next dimension.

### ***IMT PRS-P***

The IMT PRS-P scales are presented in Table 2.6. The rating scale anchors are all on a 1-5 scale but vary by dimension. Figure 2.1 presents two examples.

***Table 2.6. IMT Peer Performance Rating Scale Dimensions and Associated Example Behaviors***

Scale	Example Behaviors
Effort and Discipline	<ul style="list-style-type: none"> <li>• Puts forth individual effort in study, practice, and other training activities</li> <li>• Shows respect in word and action towards superiors</li> <li>• Obeys instructions, rules, and regulations</li> <li>• Maintains self, uniforms, living areas, and barracks to Army standards</li> </ul>
Working Effectively with other Soldiers	<ul style="list-style-type: none"> <li>• Respects, assists, and cooperates with fellow Soldiers</li> <li>• Treats all Soldiers with courtesy and respect, regardless of gender, race, ethnicity, ability, or background differences</li> <li>• Communicates ideas and information clearly and directly</li> </ul>
Physical Fitness	<ul style="list-style-type: none"> <li>• Maintains and enhances physical readiness</li> <li>• Performs required physical tasks</li> </ul>
MOS Qualification Knowledge and Skill	<ul style="list-style-type: none"> <li>• Learns and demonstrates AIT/OSUT knowledge and skills required for MOS qualification</li> <li>• Effectively applies training, rules, and strategies to solve problems</li> </ul>
Resilience and Adjustment	<ul style="list-style-type: none"> <li>• Persists in carrying out difficult tasks, even in stressful circumstances</li> <li>• Remains calm during high-pressure situations</li> <li>• Bounces back after facing a difficult situation or failure</li> <li>• Adjusts well to life in the Army</li> </ul>
Counterproductive Soldier Behaviors	<ul style="list-style-type: none"> <li>• Takes shortcuts that may be harmful</li> <li>• Takes unauthorized breaks</li> <li>• Leaves a mess for someone else to clean up</li> </ul>

**Table 2.6. (Continued)**

Scale	Example Behaviors
Going Above and Beyond	<ul style="list-style-type: none"> <li>• Notices when a fellow Soldier is falling behind and offers assistance</li> <li>• Offers to help fellow Soldiers struggling to learn training material and skills</li> <li>• Goes out of the way to give another Soldier encouragement or express appreciation</li> <li>• Spends free time learning about procedures, equipment, etc.</li> </ul>
<i>Overall Performance</i>	Instructions: Considering your evaluation of your peers on the dimensions important to being a successful Soldier, please rate the overall performance of each Soldier.

A. Effort and Discipline				
Example behaviors: <ul style="list-style-type: none"> <li>• Puts forth individual effort in study, practice, and other training activities</li> <li>• Shows respect in word and action towards superiors</li> <li>• Obeys instructions, rules, and regulations</li> <li>• Maintains self, uniforms, living areas, and barracks to Army standards</li> </ul>				
1	2	3	4	5
Much less than other Soldiers in my unit	Somewhat less than other Soldiers in my unit	About as much as other Soldiers in my unit	Somewhat more than other Soldiers in my unit	Much more than other Soldiers in my unit
B. Counterproductive Soldier Behaviors				
Example behaviors: <ul style="list-style-type: none"> <li>• Takes shortcuts that may be harmful</li> <li>• Takes unauthorized breaks</li> <li>• Leaves a mess for someone else to clean up</li> </ul>				
1	2	3	4	5
Much less often than other Soldiers in my unit	Somewhat less often than other Soldiers in my unit	About as often as other Soldiers in my unit	Somewhat more often than other Soldiers in my unit	Much more often than other Soldiers in my unit

**Figure 2.1. Sample IMT peer rater performance rating scales.**

## *In-Unit PRS-P*

The in-unit PRS-P rating dimensions and example behaviors are presented in Table 2.7. As with the IMT scales, the rating scale anchors are all on a 1-5 scale, and the response options vary somewhat by dimension. The in-unit PRS-P also includes two dimensions that are not included in the IMT version, due to less relevance with that population. With minor exceptions, the dimensions included on both the IMT and in-unit measures share the same behavioral examples.

**Table 2.7. In-Unit Peer Performance Rating Scale Dimensions and Associated Example Behaviors**

Scale	Example Behaviors
Effort and Discipline	<ul style="list-style-type: none"> <li>• Puts forth individual effort in completing tasks and handling responsibilities</li> <li>• Shows respect in word and action towards superiors</li> <li>• Obeys instructions, rules, and regulations</li> <li>• Maintains self, uniforms, and work space to Army standards</li> </ul>
Working Effectively with other Soldiers	<ul style="list-style-type: none"> <li>• Respects, assists, and cooperates with fellow Soldiers</li> <li>• Treats all Soldiers with courtesy and respect, regardless of gender, race, ethnicity, ability, or background differences</li> <li>• Communicates ideas and information clearly and directly</li> </ul>
Physical Fitness	<ul style="list-style-type: none"> <li>• Maintains and enhances physical readiness</li> <li>• Performs required physical tasks</li> </ul>
MOS Qualification Knowledge and Skill	<ul style="list-style-type: none"> <li>• Demonstrates knowledge and skills required for MOS</li> <li>• Handles most MOS task-related problems effectively</li> <li>• Effectively applies training, previous experience, rules, and strategies to complete job tasks</li> </ul>
Processing Information and Solving Problems	<ul style="list-style-type: none"> <li>• Efficiently handles day-to-day information load</li> <li>• Develops original and creative approaches to dealing with problems</li> <li>• Can integrate information from multiple sources</li> <li>• Considers relevant costs and benefits of alternative solutions</li> </ul>
Resilience and Adjustment	<ul style="list-style-type: none"> <li>• Persists in carrying out difficult tasks, even in stressful circumstances</li> <li>• Remains calm during high-pressure situations</li> <li>• Bounces back after facing a difficult situation or failure</li> <li>• Adjusts well to life in the Army</li> </ul>
Counterproductive Soldier Behaviors	<ul style="list-style-type: none"> <li>• Takes shortcuts that may be harmful</li> <li>• Takes unauthorized breaks</li> <li>• Leaves a mess for someone else to clean up</li> <li>• Speaks poorly about the Army to others</li> </ul>
Going Above and Beyond	<ul style="list-style-type: none"> <li>• Volunteers for additional duty</li> <li>• Goes out of the way to give another Soldier encouragement or express appreciation</li> <li>• Demonstrates concern about the image of the unit</li> </ul>
Emergent Leadership	<ul style="list-style-type: none"> <li>• Notices when a fellow Soldier is falling behind</li> <li>• Keeps the unit focused on goals and objectives even when supervisors are not present</li> <li>• Gains the cooperation and trust of other Soldiers in the unit</li> </ul>
<i>Overall Performance</i>	<p>Instructions: Considering your evaluation of your peers on the dimensions important to being a successful Soldier, please rate the overall performance of each Soldier.</p>

### ***PRS-P Interrater Reliability Estimates***

Table 2.8 presents interrater reliability (IRR) estimates for both the IMT and in-unit PRS-P based on Cycle 17 data. The estimates are higher for the IMT ratings than the in-unit ratings at least in part because there were generally more peer raters in IMT. This finding may also reflect differential opportunity to observe across the two environments – peers going through intensive training together compared to peers performing their jobs in the field.

Of note when reviewing these results is the Counterproductive Soldier Behaviors scale. This scale is reverse-scored and, although raters were alerted to the fact that the anchors were not the same across all dimensions, it is likely that this was a factor in the especially low interrater reliability estimates for that scale.

***Table 2.8. Peer Performance Rating Scale Interrater Reliability Estimates***

IMT	$G(q,k)$	$K$	$n_{\text{Ratees}}$	$n_{\text{Raters}}$
Effort and Discipline	.46	2.16	8,969	8,238
Working Effectively with Other Soldiers	.41	2.18	8,987	8,268
Physical Fitness	.60	2.14	8,959	8,199
MOS Qualification Knowledge & Skill	.45	2.15	8,971	8,240
Resilience and Adjustment	.42	2.14	8,972	8,209
Counterproductive Soldier Behaviors	.25	2.10	8,909	7,990
Going Above and Beyond	.39	2.14	8,954	8,178
Overall Performance Rating	.50	2.18	8,984	8,255
In-Unit	$G(q,k)$	$K$	$n_{\text{Ratees}}$	$n_{\text{Raters}}$
Effort and Discipline	.31	1.65	409	357
Working Effectively with Other Soldiers	.20	1.67	410	361
Physical Fitness	.49	1.66	406	354
MOS Qualification Knowledge & Skill	.28	1.62	402	351
Processing Information & Solving Problems	.27	1.62	401	347
Resilience and Adjustment	.29	1.62	404	347
Counterproductive Soldier Behaviors	.09	1.63	406	350
Going Above and Beyond	.28	1.64	403	352
Emergent Leadership	.34	1.61	409	355
Overall Performance Rating	.37	1.66	409	357

*Note.*  $k$  = harmonic mean number of raters per ratee. Reliability for the PRS-P scales reflect  $G(q,k)$  estimates (Putka, Le, McCloy, & Diaz, 2008). Reliability for the PRS-P composite reflects coefficient Alpha. PRS ratings from peers with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses. Table reprinted from Knapp, Salmon, Kirkendall, & Burgoyne (2020).

### ***PRS-P Scoring***

Validation analyses use a single Peer Performance Rating composite score. Specifically, we average scores across the scales (including Overall Performance and excluding Counterproductive Soldier Behaviors) to create the PRS-P Overall Performance criterion composite for both IMT and in-unit Soldiers.

### ***Supervisor Performance Rating Scales (PRS-S)***

The in-unit PRS-S, like the JKTs, were adapted from or based on previous research dating back to Project A (Moriarty et. al., 2009). Supervisor ratings are only collected on Soldiers in units. Table 2.9 lists the dimensions rated on the PRS-S.

***Table 2.9. In-Unit PRS-S Army-Wide Rating Scale Dimensions and Associated Definitions***

---

Performing Core Warrior Tasks	Performs most Core Warrior Tasks (e.g., navigation, first aid, weaponry, maintenance) competently and safely
Performing MOS-Specific Tasks	Performs MOS-specific work assignments; keeps informed of MOS and assignment changes
Communicating with Others	Speaks clearly and concisely; conveys intended message verbally and in writing
Processing Information	Monitors, interprets, and organizes information
Solving Problems	Adapts to new problem situations; applies prior training, rules, and strategies correctly; weighs alternatives when making decisions; develops novel solutions to problems; completes tasks despite major changes
Exhibiting Effort	Completes work in a timely manner; puts extra effort into completing work; seeks challenging assignments; persists in carrying out difficult assignments and responsibilities even under adverse and stressful conditions
Exhibiting Personal Discipline	Exhibits selfless service orientation; exhibits integrity and discipline both on and off the battlefield; follows instructions, rules, and regulations
Contributing to the Team	Treats team members courteously and respectfully; provides help and assistance to others; contributes to achieving team goals
Exhibiting Fitness and Bearing	Meets Army standards for physical fitness, strength, and weight; displays military bearing; meets Army standards for AR 670-1
Adjusting to the Army	Soldier fits within the Army; demonstrates adjustment to Army life, shows potential for promotion.
Following Safety Procedures	Follows safety procedures; recognizes and responds to possible dangerous or hazardous situations
Developing Own Skills	Stays up to date with his or her professional skills by seeking out additional education and training opportunities; commits to learning new things required by technology, mission, or situation
Managing Personal Matters	Manages personal life, including personal finances, family, and personal well-being
Leadership Potential	Potential effectiveness as an NCO

---

The in-unit PRS-S employ a 7-point behaviorally-anchored rating scale format (see Figure 2.2). We average ratings on the individual scales to form a single Supervisor Performance Rating composite score.

A. Performing Core Warrior Tasks						
Performs most Core Warrior Tasks (e.g., navigation, first aid, weaponry, maintenance) competently and safely.						
1	2	3	4	5	6	7
<ul style="list-style-type: none"> <li>• Is not able to perform most Core Warrior Tasks.</li> <li>• Requires constant supervision.</li> </ul>		<ul style="list-style-type: none"> <li>• Performs most Core Warrior Tasks competently.</li> <li>• Requires some supervision under difficult conditions.</li> </ul>		<ul style="list-style-type: none"> <li>• Performs almost all Core Warrior Tasks extremely effectively.</li> <li>• Requires little or no supervision, even under difficult conditions.</li> </ul>		

**Figure 2.2. Sample in-unit supervisor rater performance rating scale.**

The vast majority of Soldiers in units are rated by only one supervisor, so we cannot compute interrater reliability estimates. The internal consistency (alpha) estimate for the overall composite score computed using the Cycle 17 database was .92 (Knapp, Salmon, Kirkendall, & Burgoyne, 2020).

### ***Administrative Criteria***

Administrative criteria used in the VAST research include IMT Restarts, Army Physical Fitness Test (APFT) scores, disciplinary actions, and attrition.

#### ***IMT Restarts***

Soldiers' IMT completion status and whether they graduated from IMT with training recycles or restarts are extracted from Army Training Requirements and Resources System (ATRRS). IMT Restarts identify Soldiers who had to begin Basic Combat Training (BCT) or AIT again. Soldiers who had not had an opportunity to complete their training at the time data were extracted are excluded from analyses.

#### ***Attrition***

Attrition is a broad category that encompasses involuntary and voluntary separations for a variety of reasons (e.g., family concerns, drug or alcohol use, performance, physical standards or weight, conduct). The reason for separation is determined by the Soldiers' Separation Program Designator (SPD) code. Soldiers who left the Army for reasons outside of their or the Army's control (e.g., death or serious injury incurred while performing one's duties, secretarial authority) are excluded from our analyses. Additionally, Soldiers with SPD codes indicating re-enlistment or completion of service are not considered as having separated from the Army for the purposes of our analyses. Separation data are reported for Regular Army Soldiers only because reliable data for the Reserve and National Guard components are not available.

A Soldier's time in service is calculated as the difference between the date from which their TAPDB record was extracted and their "ship" date as indicated in the Military Integrated Resource System (MIRS). We create separate monthly attrition status variables through 36 months within a Soldier's time in service. For example, if a Soldier has a total of 12 months in service, they will be indicated as 0 (in service) or 1 (separated) for each month from first through 12<sup>th</sup> and indicated as missing data (cannot be determined) for each month from 13<sup>th</sup> through 36<sup>th</sup>. Their overall attrition status is taken as their most recent non-missing attrition status.

### ***Self-Reported Administrative Criteria***

As previously mentioned, the ALQ is also used to gather self-report administrative information, notably disciplinary actions (e.g., Article 15s) and Army Physical Fitness Test (APFT) scores. The APFT, which Soldiers take at approximate 6-month intervals, is a single continuous score. It is indexed by the ability to perform certain numbers of push-ups and sit-ups, and time taken to complete a 2-mile run, adjusted for age. Disciplinary Incidents is a count of the number of associated items (e.g., Article 15, formally counseled, or placed on restriction for problem behavior) for which there was an affirmative response.

## **Data Collection**

### ***IMT Data Collections***

Cadre at participating schoolhouses are generally responsible for ensuring that Soldiers take the online VAST assessment approximately 2 weeks prior to graduation.<sup>8</sup> The goal is to schedule the assessment toward the end of training, but before the intense period of activity right before Soldiers graduate.

After completing the other measures (e.g., JKT, ALQ), Soldiers in IMT are directed by instructional screens to find their name on a copy of the class roster, then rate the five Soldiers who follow them on the roster. Use of the rosters significantly reduces the burden on cadre, while also providing the Soldiers with the correct spelling of the rated peers' names.

### ***In-Unit Data Collections***

Most Army installations host an annual "umbrella week" which are time periods set aside to support research data collection activities. We schedule data collection visits during these umbrella weeks, resulting in rolling data collection throughout the year at various locations. Soldiers and their supervisors are instructed to report to a computer classroom at which time data collection staff (HumRRO and/or ARI researchers) guide them through the online data collection process.

To identify peer rater-ratee pairs in units, data collectors ask supervisors arriving to the test site to organize their Soldiers into groups according to who each supervisor can rate. As each group signs in, data collectors ask the Soldiers, as a group, if they have worked with each other long

---

<sup>8</sup> IMT and in-unit measures are administered online using an application called Verint which is hosted on an Army server.

enough to be familiar with each other’s performance. If the group includes more than six Soldiers, only five peers are assigned to each Soldier. After all Soldiers and supervisors are signed in, data collectors complete a peer rating instruction sheet for each Soldier. This sheet lists the names and project identification numbers of the Soldiers for whom they are to provide ratings. Soldiers who are determined to have no peers present are instructed to skip the peer ratings section of the assessment.

### Biannual Longitudinal Database Updates

Twice per year, the VAST longitudinal database is updated with new cases and updated data for Soldiers who were introduced into the database in earlier cycles.

An illustrative view of our data sources is provided in Figure 2.3 (see Appendix A for acronym definitions). The analysis data files are built on a foundational database drawn from the TAPDB onto which predictor and criterion data from other sources are added. The lighter boxes within the figure represent sources of data, and the darker boxes represent samples on which descriptive or inferential analyses are conducted. The leftmost column in the figure summarized the predictor and primary demographic data sources used to derive the *Applicant Sample*. The other columns summarize the research-only (i.e., non-administrative) and administrative criterion data that are also matched to TAPDB records. Soldiers with predictor and criterion data form the *IMT*, *In-Unit*, and *Administrative Validation Samples*.

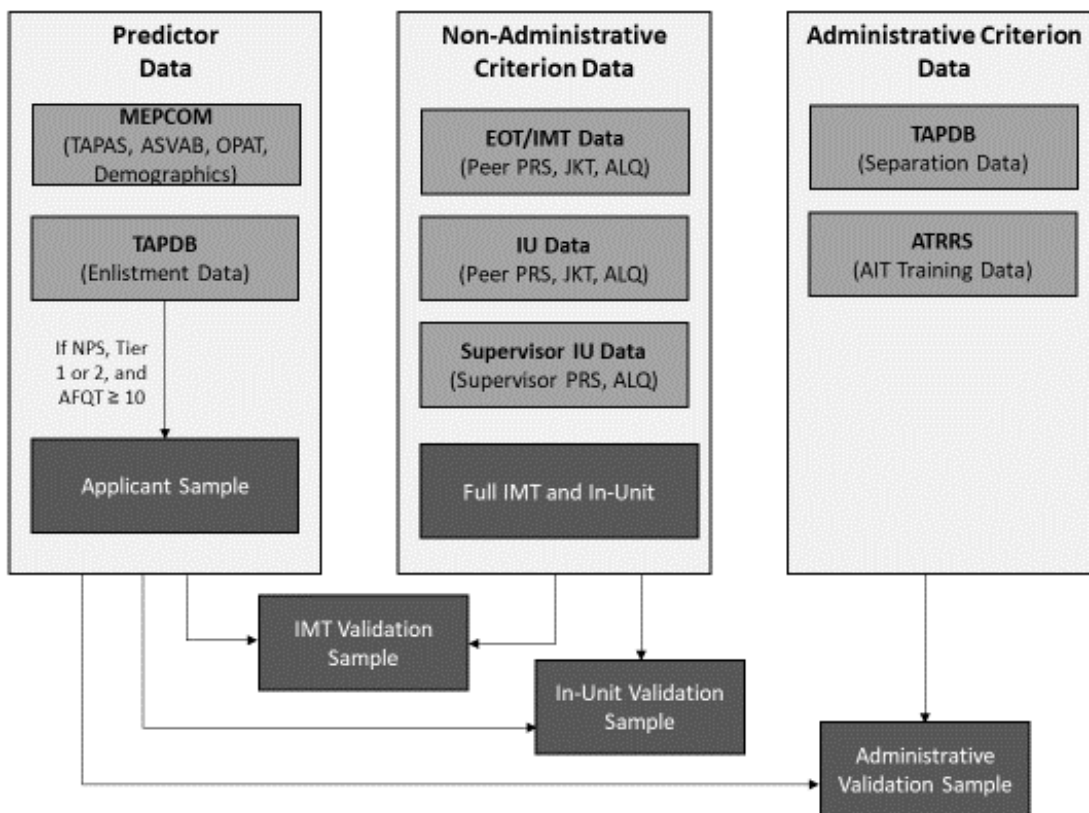


Figure 2.3. Overview of VAST data file merging and nested sample generation process.

For the purpose of evaluating the effectiveness of TAPAS, exclusions to the analysis samples were imposed based on prior service, AFQT score, and education level. Specifically, applicants who previously served in the military were dropped. AFQT Category V cases were deleted since they are not eligible for military service. Similarly, non-high school graduates (Education Tier 3 – individuals not currently attending high school and do not possess a high school diploma or alternate credential) were removed because they are ineligible for accession.

The Applicant Sample includes Soldiers from all Army components – Regular Army (RA), U.S. Army Reserve (USAR), and U.S. Army National Guard (ARNG) who took TAPAS. Soldiers from all components are included in all analyses except for those involving attrition. For those analyses, only Regular Army Soldiers are included.

It bears noting that the TOPS research databases (Cycles 1 through 18) used the TAPAS database provided by MEPCOM as the foundational database rather than TAPDB. Using the much larger TAPDB makes the foundational database more useful to ARI for a variety of personnel research needs.

### **Analysis Plan Overview**

With each biannual cycle, we routinely report descriptive statistics for the major analysis variables and conduct analyses examining the extent to which TAPAS scores increment prediction of key outcome measures over AFQT. Periodically we also report updated reliability estimates of our measures. In addition to these routine analyses, we periodically conduct analyses targeted at ancillary research questions of operational and scientific interest.

## CHAPTER 3: CYCLE 20 SAMPLE CHARACTERISTICS AND DESCRIPTIVE STATISTICS<sup>9</sup>

This chapter describes characteristics of the samples used in the TAPAS evaluation analyses and descriptive statistics for the predictor and criterion measures. We begin by describing how Soldier data were filtered for analysis, then describe multiple subsamples that were created to support various types of analyses, and, finally, provide descriptive statistics for the predictor and criterion variables included in the validation analyses.

### Description of Analysis Samples

Table 3.1 summarizes the key variables that were used to create the analysis samples and the demographic characteristics observed in those samples. The Applicant Sample includes non-prior service Soldiers

- in Education Tier 1 or 2 (high school diploma or non-diploma graduates),
- in AFQT Category I-IV (eligible for military service),
- with valid TAPAS score data (e.g., without flat responding, large amounts of missing data), and
- who were administered TAPAS versions other than the 13D-CAT.<sup>10</sup>

From the sample of TAPAS data, these filters excluded 16,461 Soldiers in Education Tier 3, 32,831 Soldiers in AFQT Category V, 1,654 Soldiers who completed TAPAS version 13D and 66,068 Soldiers with invalid TAPAS data. There were thus 1,227,032 Soldiers who met the screening criteria and were included in the Applicant Sample.

The Administrative Validation Sample includes 731,133 Soldiers who meet all of the inclusion criteria for the Applicant Sample and also have a signed contract and at least one record from an administrative criterion data source (e.g., ATRRS). There are 94,049 Soldiers with IMT criterion data linked to an administrative TAPAS record and included in the IMT Validation Sample. There are 9,207 Soldiers with in-unit data and a matching TAPAS record included in the In-Unit Validation Sample.

The Applicant Sample is much larger than the Administrative Validation Sample because not all applicants access into the Army. Similarly, the IMT and In-Unit Validation Samples are much smaller than the Administrative Validation Sample in part because not all Soldiers participate in IMT and in-unit data collections. Furthermore, we rely on self-reported name and date of birth to match TAPAS records to the IMT and in-unit criterion data, which result in a nontrivial percentage of unsuccessful matches.

---

<sup>9</sup> The report authors would like to acknowledge the efforts of Arielle Rogers, Colby Nesbitt, Alyssa Marshall, and their teams in contributing to the work described in this chapter.

<sup>10</sup> This was an early version of TAPAS, which had minor dissimilarities with subsequent TAPAS forms. Prior to Cycle 19, these cases were included in the various analysis samples and filtered out prior to validation work. This process change has resulted in lower sample sizes in the tables in this chapter than would be expected compared to corresponding tables in the final TOPS technical report (Knapp & Kirkendall, 2020).

**Table 3.1. Background and Demographic Characteristics of the Analysis Samples**

Characteristic	Applicant <sup>a</sup>		Administrative Validation <sup>b</sup>		IMT Validation <sup>c</sup>		In-Unit Validation <sup>d</sup>	
	<i>n</i>	%	<i>n</i>	%	<i>N</i>	%	<i>n</i>	%
<i>Component</i>								
Regular	705,499	57.5	440,722	60.3	57,125	60.7	9,135	99.2
ARNG	361,357	29.4	199,656	27.3	27,058	28.8	39	0.4
USAR	160,176	13.1	90,755	12.4	9,866	10.5	33	0.4
<i>Education Tier</i>								
Tier 1	1,165,726	95.0	701,088	95.9	90,418	96.1	8,918	96.9
Tier 2	61,306	5.0	30,045	4.1	3,631	3.9	289	3.1
<i>Military Occupational Specialty</i>								
Infantryman (11B/C/X + 18X)	122,320	10.0	108,490	14.8	19,293	20.5	2,124	23.1
Combat Engineer (12B)	23,348	1.9	20,101	2.7	6,578	7.0	359	3.9
Cannon Crewmember (13B)	15,304	1.2	13,371	1.8	483	0.5	204	2.2
Field Artillery Tactical Data Systems Specialist (13D)	4,529	0.4	4,413	0.6	982	1.0	65	0.7
Fire Support Specialist (13F)	8,705	0.7	7,695	1.1	1,182	1.3	74	0.8
Cavalry Scout (19D)	20,666	1.7	18,115	2.5	2,765	2.9	453	4.9
M1 Armor Crewmember (19K)	9,491	0.8	8,599	1.2	2,645	2.8	246	2.7
Military Police (31B)	33,973	2.8	30,696	4.2	11,322	12.0	169	1.8
Human Resources Specialist (42A)	20,744	1.7	18,690	2.6	4,095	4.4	235	2.6
Combat Medic Specialist (68W)	38,123	3.1	33,099	4.5	8,956	9.5	350	3.8
Motor Transport Operator Soldiers (88M)	38,680	3.2	34,207	4.7	10,990	11.7	649	7.0
Wheeled Vehicle Mechanic (91B)	40,855	3.3	35,796	4.9	3,315	3.5	504	5.5
Other	427,597	34.8	374,794	51.3	21,443	22.8	3,775	41.0
Unknown	422,697	34.4	23,067	3.2	--	--	--	--
<i>AFQT Category<sup>e</sup></i>								
I	69,239	5.6	42,635	5.8	5,697	6.1	393	4.3
II	348,827	28.4	227,413	31.1	32,494	34.6	2,403	26.1
IIIA	248,818	20.3	162,153	22.2	21,405	22.8	2,051	22.3
IIIB	382,862	31.2	255,337	34.9	29,034	30.9	3,720	40.4
IV	177,286	14.4	43,595	6.0	5,419	5.8	640	7.0

**Table 3.1. (Continued)**

Characteristic	Applicant <sup>a</sup>		Administrative Validation <sup>b</sup>		IMT Validation <sup>c</sup>		In-Unit Validation <sup>d</sup>	
	<i>n</i> = 1,227,032		<i>n</i> = 731,133		<i>n</i> = 94,049		<i>n</i> = 9,207	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>Contract Status</i>								
Signed	840,721	68.5	--	--	--	--	--	--
Not signed	371,970	30.3	--	--	--	--	--	--
<i>Gender</i>								
Female	283,626	23.1	145,119	19.8	14,811	15.7	1,152	12.5
Male	943,406	76.9	586,014	80.2	79,238	84.3	8,055	87.5
<i>Race</i>								
African American	290,614	23.7	158,965	21.7	16,470	17.5	2,208	24.0
American Indian	9,866	0.8	5,726	0.8	738	0.8	72	0.8
Asian	58,034	4.7	33,277	4.6	3,927	4.2	529	5.7
Hawaiian/Pacific Islander	2,642	0.2	1,365	0.2	147	0.2	23	0.2
Caucasian	848,451	69.1	524,296	71.7	71,796	76.3	6,286	68.3
Multiple	3,829	0.3	2,261	0.3	230	0.2	32	0.3
Declined to Answer/Missing	13,596	1.1	5,243	0.7	741	0.8	57	0.6
<i>Ethnicity</i>								
Hispanic/Latino	206,536	16.8	116,460	15.9	16,018	17	1,726	18.7
Not Hispanic	1,007,2	82.1	610,033	83.4	77,405	82.3	7,419	80.6
Declined to Answer/Missing	13,256	1.1	4,640	0.6	626	0.7	62	0.7

<sup>a</sup>Non-prior service Soldiers in Education Tier 1 and 2 and AFQT Category IV or above, with no prior service and with valid TAPAS score data on forms other than the 13D-CAT

<sup>b</sup>Soldiers in Applicant Sample with a signed contract and at least one criterion record (i.e., schoolhouse, in-unit, ATRRS, RITMS, or attrition).

<sup>c</sup>Soldiers in Applicant Sample with a signed contract and criterion data collected at schoolhouses.

<sup>d</sup>Soldiers in Applicant Sample with a signed contract and criterion data collected in units.

<sup>e</sup>AFQT Categories IIIB and IV are oversampled. Values presented are not representative of Army accessions.

Sample sizes reported in subsequent chapters and appendices are generally smaller than the figures reported here because of further data filtering or disaggregation that occurred. For example, criterion scores were determined to be valid if they passed multiple data quality screens intended to identify unmotivated responding. Additional screens were analysis-specific and thus not applied to the descriptive analysis of the samples described in this chapter.

### Descriptive Statistics for Predictor Measures

Table 3.2 presents the means and standard deviations for the three TAPAS composite scores by TAPAS version. We also computed descriptive statistics for AFQT for the entire Applicant Sample ( $n = 1,227,032$ ,  $M = 54.55$ ,  $SD = 22.49$ ). Correlations among these scores (as well as correlations with criterion measures) are reported in Appendix B.

**Table 3.2. Means and Standard Deviations for the TAPAS Composites by TAPAS Version**

TAPAS Composite	15D-Static/CAT v4/5/7/8 ( <i>n</i> = 219,761-421,818) <sup>a</sup>		13D-CAT v9 ( <i>n</i> = 267,837)		13D-CAT v10 ( <i>n</i> = 267,500)		13D-CAT v11 ( <i>n</i> = 269,877)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Can-Do	99.76	20.92	99.88	19.63	99.25	19.63	100.01	19.58
Will-Do	99.92	19.16	98.84	20.18	98.56	19.57	99.15	20.25
Adaptation	100.41	19.33	99.56	19.89	99.47	19.79	100.06	19.83

*Note.* Results are limited to the Applicant Sample (which excludes prior service, Education Tier 3 and AFQT Category V with valid TAPAS score data on forms other than the original 13D-CAT).

<sup>a</sup>Can-Do and Adaptation composites were not calculated for Soldiers taking TAPAS versions 7 and 8, resulting in reduced sample sizes.

## Descriptive Statistics for Criterion Measures

### IMT Criteria

Descriptive statistics for IMT criteria are reported in Table 3.3. The criteria for the IMT Validation Sample include measures from self-report, other-report, and administrative sources. As described in Chapter 2, the Knowledge and Skill composite represents the average of the Soldier’s WTBD JKT and MOS JKT.<sup>11</sup> The IMT Peer Performance Rating composite is an average of seven scale scores concerning a Soldier’s performance as rated by peers. IMT Restarts is a dichotomous administrative variable that identifies Soldiers who had to begin training again during IMT. Several self-reported variables, scales, and composites from the ALQ are reported. For more information about the development and psychometric properties of various IMT criteria, refer to Chapter 2. Correlations among the IMT criteria (as well as bivariate correlations with predictor scores) can be found in Appendix B.

**Table 3.3. Descriptive Statistics for Criterion Scores in the IMT Validation Sample**

Domain/Scale	<i>n</i>	<i>M</i>	<i>SD</i>	Min	Max
Knowledge & Skill Composite	44,312	0.04	0.85	-4.00	2.40
WTBD JKT	71,836	58.38	14.86	2.70	100.00
Peer Performance Rating Composite	35,093	3.47	0.61	1.00	5.00
IMT Restarts (Y/N)	93,950	0.05	0.21	0.00	1.00
<i>Army Life Questionnaire (ALQ)</i>					
Commitment & Fit	82,416	3.80	0.64	1.00	5.00
Retention Cognitions	82,103	3.60	1.08	1.00	5.00
Peer Leadership	38,616	3.31	0.44	1.00	5.00
Counterproductive Soldier Behavior	38,616	1.92	0.63	1.00	5.00
Resilience	38,616	4.01	0.60	1.00	5.00
Army Physical Fitness Test (APFT)	79,369	251.49	28.20	180.00	300.00
Disciplinary Incidents (Y/N)	81,032	0.23	0.42	0.00	1.00

*Note.* Ratings on PRS-P range from 1 and 5. PRS ratings from peers with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses.

<sup>11</sup> For IMT Soldiers who do not have an MOS JKT score, AIT grade is substituted when available.

### ***In-Unit Criteria***

The in-unit criteria presented in Table 3.4 are analogous to the IMT criteria, with the addition of the Supervisor Performance Rating composite, which is the average of 13 scale scores regarding a Soldier’s performance as rated by a supervisor. The in-unit Peer Performance Rating composite is the average of nine scale scores of peer ratings concerning a Soldier’s performance. Correlations among the in-unit criteria (as well as bivariate correlations with predictor scores) are presented in Appendix B.

***Table 3.4. Descriptive Statistics for Criterion Scores in the In-Unit Validation Sample***

Domain/Scale	<i>n</i>	<i>M</i>	<i>SD</i>	Min	Max
Supervisor Performance Rating Composite	6,102	4.87	1.13	1.00	7.00
Peer Performance Rating Composite	2,503	3.73	0.71	1.00	5.00
WTBD JKT	6,148	58.87	13.63	8.11	96.15
<i>Army Life Questionnaire (ALQ)</i>					
Commitment & Fit Composite	8,152	3.31	0.78	1.00	5.00
Retention Cognitions Composite	3,021	2.88	1.32	1.00	5.00
Peer Leadership Composite	3,023	3.25	0.49	1.00	5.00
Counterproductive Work Behavior	3,023	2.04	0.61	1.00	5.00
Resilience	3,023	3.87	0.66	1.00	5.00
Army Physical Fitness Test (APFT) Score	7,801	251.12	28.78	180.00	300.00
Disciplinary Incidents (Y/N)	8,203	0.26	0.44	0.00	1.00

*Note.* Ratings on PRS-P range from 1 and 5. Ratings on PRS-S range from 1 and 7. PRS ratings from raters with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses.

### ***Attrition***

Table 3.5 summarizes the base rates for attrition in Regular Army Soldiers at various time intervals.

***Table 3.5. Base Rates for Regular Army Attrition in the Administrative Validation Sample***

Criterion	<i>n<sup>a</sup></i>	<i>n<sub>Attrit</sub></i>	<i>%Attrit</i>
<i>Cumulative Attrition</i>			
6-Month	389,549	46,703	12.0
12-Month	377,192	57,139	15.1
24-Month	333,986	80,127	24.0
36-Month	289,366	100,832	34.8

*Note.* *n<sup>a</sup>* = number of Soldiers with attrition data at the time data were extracted. *n<sub>Attrit</sub>* = number of Soldiers who attrited at the specified months of service. *%Attrit* = percentage of Soldiers who attrited through the specified months of service [ $(n_{Attrit}/n) \times 100$ ].

## Summary

The VAST analysis samples represent a combination of administrative, IMT, and in-unit data obtained from Soldiers, their supervisors, and peers, and archival resources at multiple points in time using a variety of data collection methods. In this chapter, we have described the filters used to construct the analysis samples, with the Applicant Sample including 1,227,032 applicants who took TAPAS, and 731,133 with at least one type of criterion data in the Administrative Validation Sample. Lastly, we provided descriptive statistics for key predictor and criterion variables that are included in the validation analyses.

## CHAPTER 4: EVIDENCE FOR THE PREDICTIVE VALIDITY OF THE TAPAS<sup>12</sup>

This chapter describes our evaluation of TAPAS and its potential to predict Soldiers' performance and retention through their first enlistment term. For these analyses, we focus on the Can-Do, Will-Do, and Adaptation TAPAS composites, and their relationships with various cognitive, motivational, and attrition criteria. Consistent with the Army's personnel goals, we examined performance and retention-related outcomes that, as a group, provide representative coverage of the criterion space (Campbell, Hanson, & Oppler, 2001; Campbell, McHenry, & Wise, 1990; Knapp & Tremble, 2007; Strickland, 2005). We used regression analysis to examine the incremental validity of the TAPAS to predict outcomes beyond AFQT.

We evaluated TAPAS using a common set of criterion measures. Collectively, the measures represent a range of important Soldier outcomes and reflect both cognitive and non-cognitive indices, including attitudinal measures (i.e., ALQ) and performance ratings. Tables 2.2 through 2.4 summarize the IMT and in-unit criteria, including a brief description of each. For composite criteria, the component scales are provided. Intercorrelations among these criteria are provided in Appendix B.

### Incremental Validity Analyses

The purpose of the incremental validity analyses is to examine the potential of TAPAS composites to predict Soldier outcomes beyond AFQT. For each TAPAS composite and criterion of interest, we used hierarchical regression to examine the incremental validity of TAPAS after accounting for the effect of AFQT. This approach is generally consistent with previous evaluations of TAPAS and similar experimental non-cognitive predictors (e.g., Ingerick et al., 2009; Knapp & Heffner, 2009; 2010; Trippe, Caramagno, Allen, & Ingerick, 2011).

To conduct the hierarchical regression analyses, we regressed the criterion onto Soldiers' AFQT scores in Step 1, followed by the given TAPAS composite scores in Step 2 for each regression model. By using the TAPAS composites in the analyses, these models aim to reflect potential operational use of the TAPAS. Results of these analyses are presented in Tables 4.1 to 4.4. While sample sizes for all of the IMT analyses are quite large, they are generally smaller for analyses involving the Can-Do and Adaptation TAPAS composites compared to the Will-Do composite because TAPAS versions 7 and 8 did not include all the facets required to compute these scores using the same formula. Sample sizes for the in-unit analyses are much smaller than for the IMT analyses, but the correspondingly smaller sample sizes for the Can-Do and Adaptation composites are sufficient to provide reasonably stable results.

For all hierarchical regression analyses, we used Ordinary Least Squares (OLS) regression to examine continuously-scaled criteria, and logistic regression for dichotomous criteria (i.e., attrition, Disciplinary Incidents, IMT Restarts). We discuss the results of the OLS regression analyses with respect to *R* values when interpreting the validity of TAPAS. We refer primarily to odds ratios (ORs) when interpreting the results of the logistic regression analyses. Additional details concerning the specific indices presented for the logistic regression analyses are provided with

---

<sup>12</sup> The report authors would like to acknowledge that the text in this chapter is largely derived from a comparable chapter authored by Michael Hughes in the *Tier One Performance Screen Initial Operational Test and Evaluation: Capstone Report* (Knapp & Kirkendall, 2020). Alyssa Marshall led the analysis work reported here.

the results of the dichotomous outcomes later in the chapter. Attrition criteria were examined for Regular Army Soldiers only.

### *IMT Incremental Validity Analyses*

Table 4.1 summarizes the incremental validity results of the TAPAS composites for predicting IMT performance criteria over and above AFQT. Overall, the results suggest that both the Will-Do and Adaptation TAPAS composites can enhance the Army’s ability to predict a number of important outcomes. Below, we describe the specific outcomes for which the TAPAS composites demonstrated notable predictive gains beyond AFQT alone.

With respect to the motivation-based performance criteria, the TAPAS Will-Do composite enhanced the prediction of the Resilience scale ( $\Delta R = .19$ ), Peer Leadership composite ( $\Delta R = .19$ ), and APFT score ( $\Delta R = .17$ ); and to a lesser extent, the prediction of Commitment and Fit ( $\Delta R = .10$ ), Counterproductive Soldier Behavior ( $\Delta R = .07$ ), and the Peer Performance Rating composite ( $\Delta R = .08$ ). This small incremental gain in the ratings composite is likely attenuated due to the low interrater reliability of that measure.

***Table 4.1. Incremental Validity Estimates for TAPAS over AFQT for Predicting IMT Criteria***

IMT Criterion Measure / Model	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	$\Delta R$
<i>Knowledge and Skill</i>			
<i>n = 38,024 - 44,312</i>			
Can-Do	<b>.43</b>	<b>.43</b>	<b>.00</b>
Will-Do	<b>.43</b>	<b>.43</b>	<b>.00</b>
Adaptation	<b>.43</b>	<b>.43</b>	<b>.00</b>
<i>WTBD JKT</i>			
<i>n = 63,132 - 71,836</i>			
Can-Do	<b>.36</b>	<b>.36</b>	<b>.00</b>
Will-Do	<b>.36</b>	<b>.36</b>	<b>.00</b>
Adaptation	<b>.36</b>	<b>.36</b>	<b>.00</b>
<i>Peer Performance Ratings</i>			
<i>n = 34,862 - 35,093</i>			
Can-Do	<b>.07</b>	<b>.07</b>	<b>.00</b>
Will-Do	<b>.07</b>	<b>.15</b>	<b>.08</b>
Adaptation	<b>.07</b>	<b>.12</b>	<b>.05</b>
<i>APFT</i>			
<i>n = 70,560 - 79,369</i>			
Can-Do	<b>.09</b>	<b>.09</b>	<b>.00</b>
Will-Do	<b>.09</b>	<b>.26</b>	<b>.17</b>
Adaptation	<b>.09</b>	<b>.18</b>	<b>.09</b>
<i>ALQ Commitment and Fit</i>			
<i>n = 73,301 - 82,416</i>			
Can-Do	<b>.04</b>	<b>.05</b>	<b>.01</b>
Will-Do	<b>.04</b>	<b>.14</b>	<b>.10</b>
Adaptation	<b>.04</b>	<b>.04</b>	<b>.01</b>

**Table 4.1. (Continued)**

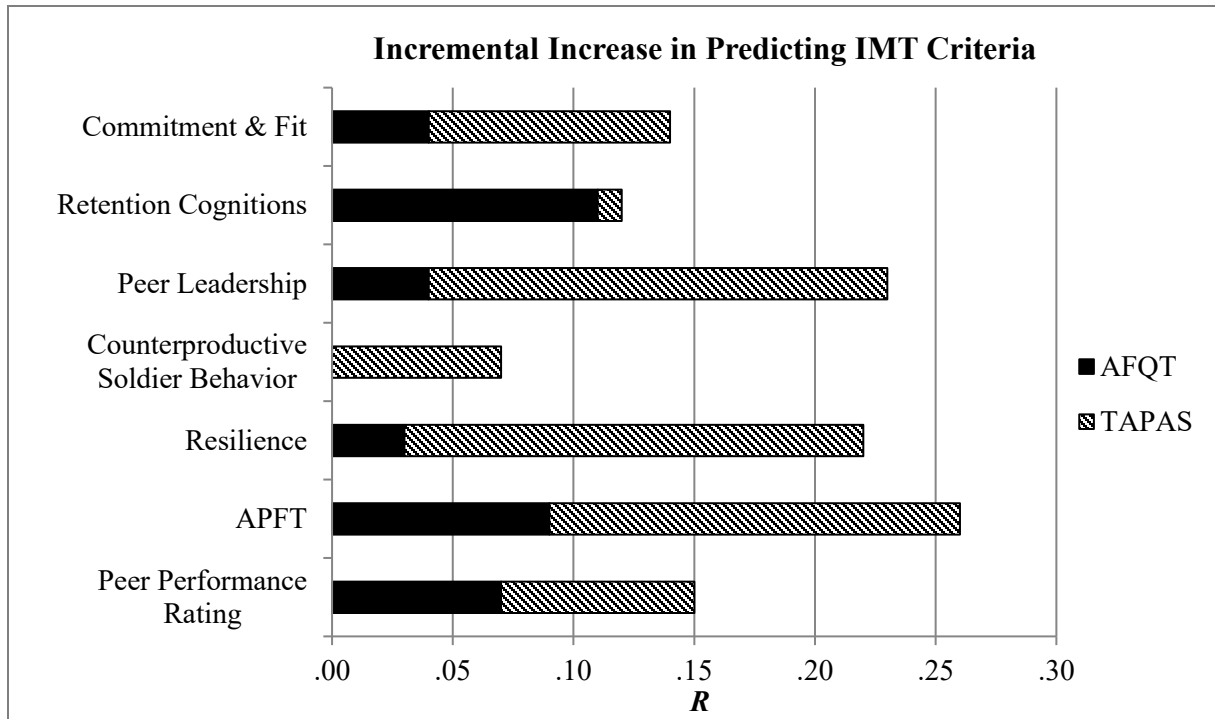
IMT Criterion Measure / Model	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	$\Delta R$
<i>ALQ Retention Cognitions</i> <span style="float: right;"><i>n</i> = 73,024 - 82,103</span>			
Can-Do	<b>.11</b>	<b>.11</b>	<b>.00</b>
Will-Do	<b>.11</b>	<b>.12</b>	<b>.01</b>
Adaptation	<b>.11</b>	<b>.11</b>	<b>.00</b>
<i>ALQ Peer Leadership</i> <span style="float: right;"><i>n</i> = 38,341 - 38,616</span>			
Can-Do	<b>.04</b>	<b>.06</b>	<b>.02</b>
Will-Do	<b>.04</b>	<b>.23</b>	<b>.19</b>
Adaptation	<b>.04</b>	<b>.04</b>	<b>.00</b>
<i>ALQ Counterproductive Soldier Behavior</i> <span style="float: right;"><i>n</i> = 38,341 - 38,616</span>			
Can-Do	.00	.01	<b>.01</b>
Will-Do	.00	<b>.07</b>	<b>.07</b>
Adaptation	.00	.01	.00
<i>ALQ Resilience</i> <span style="float: right;"><i>n</i> = 38,341 - 38,616</span>			
Can-Do	<b>.03</b>	<b>.04</b>	<b>.01</b>
Will-Do	<b>.03</b>	<b>.21</b>	<b>.19</b>
Adaptation	<b>.03</b>	<b>.08</b>	<b>.05</b>

Note. *R* = multiple correlations between AFQT and selected TAPAS composite scales with the targeted criterion measure.  $\Delta R$  = Increment in *R* from adding the selected TAPAS composite scale to the regression model [(AFQT + TAPAS) – AFQT Only]. Bolded values indicate  $p < .05$ . Zero values may be significant due to rounding error (i.e., deltas less than .005).

The TAPAS Adaptation composite also demonstrated a small degree of incremental validity in the prediction of APFT score ( $\Delta R = .09$ ), Resilience ( $\Delta R = .05$ ), and Peer Performance Rating composite ( $\Delta R = .05$ ). The Can-Do composite is not designed to predict Will Do criteria, and it did not demonstrate any meaningful incremental validity for the motivation-based outcomes.

Regarding the cognitive criteria, the TAPAS composites (Can-Do, Will-Do, and Adaptation) evidenced no notable increments over AFQT in predicting scores on the composite measure of Knowledge and Skill and the Army-wide job knowledge test ( $\Delta R_s = .00$ ). These results are not surprising given that AFQT is an established strong predictor of Can Do outcomes.

Figure 4.1 provides results of analyses examining the incremental validity of criteria-specific TAPAS scales over and above AFQT for IMT Will Do and Can Do criteria, respectively. Similar to the results of the TAPAS composite models discussed previously, the Will-Do scales provided the largest gains over AFQT. Increases in *R* were largest for the Peer Leadership, Resilience, and APFT score ( $\Delta R_s \geq .17$ ). Results of the Can-Do TAPAS scales showed no notable increments over AFQT in predicting the Can Do criteria.



**Figure 4.1.** Increase in prediction of IMT criteria using TAPAS Will-Do scales.

### ***In-Unit Incremental Validity Analyses***

The incremental validity results for predicting in-unit performance are presented in Table 4.2. Similar to the results for the IMT performance criteria, these results also suggest that the TAPAS composites are useful predictors of in-unit performance outcomes. For multiple outcomes, the Will-Do and Adaptation TAPAS composites both exhibited enhanced prediction beyond AFQT alone.

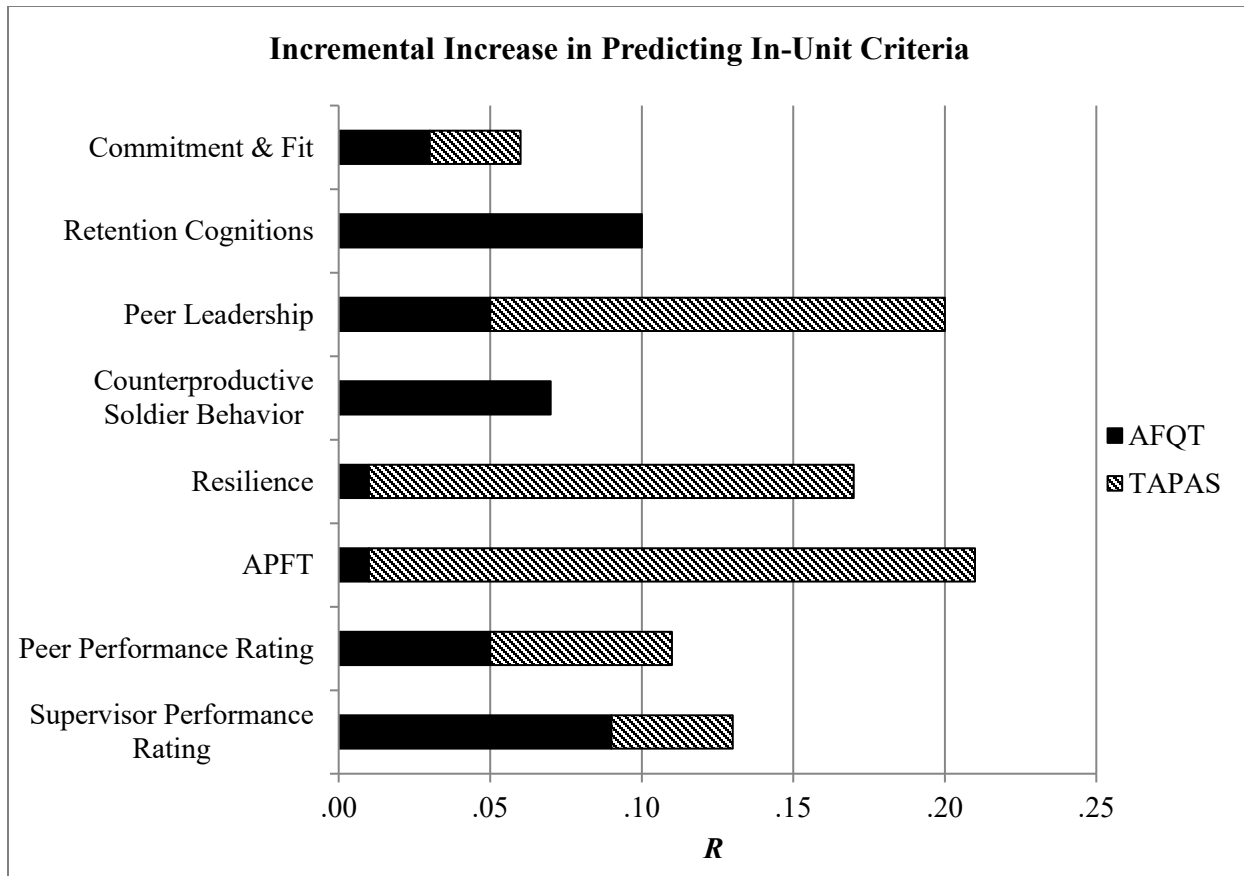
For motivation-based criteria, the Will-Do composite showed increases beyond AFQT in predicting APFT score ( $\Delta R = .20$ ), Peer Leadership ( $\Delta R = .15$ ), Resilience ( $\Delta R = .16$ ), and Peer Performance Rating composite ( $\Delta R = .06$ ) scores. The TAPAS Adaptation composite demonstrated small degrees of incremental validity for predicting APFT score ( $\Delta R = .11$ ) and Resilience ( $\Delta R = .04$ ). Similar to the analyses of the IMT criteria, the Can-Do composite did not provide incremental validity in the prediction of any in-unit criteria. However, this result is expected given that the Can-Do composite is not intended to be related to Will Do outcomes, and AFQT is an established strong predictor of Can Do outcomes.

**Table 4.2. Incremental Validity Estimates for TAPAS over AFQT for Predicting In-Unit Adjustment, Commitment and Fit, Fitness, and Retention Criteria**

In-Unit Criterion Measure / Model	AFQT R	AFQT + TAPAS R	$\Delta R$
<i>Peer Performance Ratings</i>			
	<i>n = 2,348 - 2,503</i>		
Can-Do	<b>.05</b>	<b>.05</b>	.00
Will-Do	<b>.05</b>	<b>.11</b>	<b>.06</b>
Adaptation	<b>.05</b>	.05	.00
<i>Supervisor Performance Ratings</i>			
	<i>n = 4,573 - 6,102</i>		
Can-Do	<b>.09</b>	<b>.09</b>	.00
Will-Do	<b>.09</b>	<b>.13</b>	<b>.04</b>
Adaptation	<b>.09</b>	<b>.09</b>	<b>.02</b>
<i>APFT</i>			
	<i>n = 5,873 - 7,801</i>		
Can-Do	.01	.01	.00
Will-Do	.01	<b>.21</b>	<b>.20</b>
Adaptation	.01	<b>.12</b>	<b>.11</b>
<i>ALQ Commitment and Fit</i>			
	<i>n = 6,154 - 8,152</i>		
Can-Do	<b>.04</b>	<b>.04</b>	.00
Will-Do	<b>.03</b>	<b>.07</b>	<b>.03</b>
Adaptation	<b>.04</b>	<b>.04</b>	.00
<i>ALQ Retention Cognitions</i>			
	<i>n = 2,811 - 3,021</i>		
Can-Do	<b>.10</b>	<b>.10</b>	.00
Will-Do	<b>.10</b>	<b>.10</b>	.00
Adaptation	<b>.10</b>	<b>.10</b>	.00
<i>ALQ Peer Leadership</i>			
	<i>n = 2,813 - 3,023</i>		
Can-Do	<b>.04</b>	.04	.00
Will-Do	<b>.05</b>	<b>.20</b>	<b>.15</b>
Adaptation	<b>.04</b>	<b>.05</b>	.00
<i>ALQ Counterproductive Soldier Behavior</i>			
	<i>n = 2,813 - 3,023</i>		
Can-Do	<b>.06</b>	<b>.08</b>	<b>.02</b>
Will-Do	<b>.07</b>	<b>.07</b>	.00
Adaptation	<b>.06</b>	<b>.07</b>	.00
<i>ALQ Resilience</i>			
	<i>n = 2,813 - 3,023</i>		
Can-Do	.01	.02	.01
Will-Do	.01	<b>.18</b>	<b>.16</b>
Adaptation	.01	<b>.05</b>	<b>.04</b>

Note. R = multiple correlations between AFQT and selected TAPAS composite scales with the targeted criterion measure.  $\Delta R$  = Increment in R from adding the selected TAPAS composite score to the regression model [(AFQT + TAPAS) – AFQT Only]. Bolded values indicate  $p < .05$ .

The results of the incremental validity analyses for Will-Do TAPAS scales are shown in Figure 4.2. Similar to the results of the TAPAS composite models predicting both IMT and in-unit criteria, the Will-Do scales provided large gains over AFQT. Increases in  $R$  were largest for APFT, Peer Leadership, and Resilience ( $\Delta R_s \geq .15$ ).



**Figure 4.2.** Increase in prediction of in-unit criteria using the TAPAS Will-Do scales.

### ***Predicting Dichotomous Outcomes***

In addition to the OLS regression analyses of IMT and in-unit criteria, we conducted logistic regression analyses of the dichotomous outcomes, including Disciplinary Incidents, IMT Restarts, and attrition. For these models, we estimated odds ratios (ORs) for the predictors as well as the corresponding confidence intervals (CIs). Additionally, we computed point biserial correlations ( $r_{pb}$ ) and conducted  $\chi^2$  tests of the change in model deviance (i.e., negative two log-likelihood; -2LL) from the AFQT-only to the AFQT + TAPAS composite models.

Odds ratios can be used to assess the likelihood (or odds) of a given outcome depending on change in a predictor. Specifically, for a given logistic regression model, a unique odds ratio is estimated for each predictor, and represents the amount of change in the odds of the outcome that is associated with change in the given predictor. For the present analyses, the ORs represent the amount of change in the likelihood of each outcome that can be attributed to every 1.0 change in

the predictor score. For ORs below 1.0, values closer to 0.0 indicate stronger negative relationships. Although values of ORs cannot fall below 0.0, there is no upper limit for ORs (Cohen, Cohen, West, & Aiken, 2003). In addition, we computed 95% CIs for the ORs, which can be interpreted as an index of statistical significance for each. That is, a CI that contains 1.0 suggests that the relationship between the associated predictor and outcome is not significant.

Point biserial correlations represent the correlation between a Soldier's predicted probability of exhibiting a selected behavior and his or her actual behavior (e.g., being involved in a disciplinary incident; Tabachnick & Fidell, 2013). As such, stronger point biserial correlations reflect stronger relationships between predicted and observed outcomes, and thus are indicative of better-fitting models. Model deviance (i.e., -2LL) also provides an index of model fit. Moreover, the difference in deviances obtained from nested logistic regression models can be tested using likelihood ratio  $\chi^2$  tests to determine the statistical significance of change in model fit. In the present application, statistically significant likelihood ratio  $\chi^2$  tests of the change in deviances suggest that the inclusion of a given TAPAS composite to a regression model provides significantly better prediction of the outcome than AFQT alone.

Results of the analyses examining Disciplinary Incidents (IMT and in-unit) and IMT Restarts are provided in Table 4.3. Both the Will-Do and Adaptation TAPAS composites enhanced the prediction of IMT Disciplinary Incidents and IMT Restarts beyond the AFQT-only models. Specifically, for both Disciplinary Incidents ( $OR_{Will-Do} = .988$ ;  $OR_{Adaptation} = .994$ ) and IMT Restarts ( $OR_{Will-Do} = .994$ ;  $OR_{Adaptation} = .994$ ), ORs associated with both composites were below 1.0, indicating that as scores on these composites increased, the likelihood of the outcome decreased. Similarly, for in-unit Disciplinary Incidents, both the Will-Do and Adaptation composites had significant relationships ( $OR_{Will-Do} = .993$ ;  $OR_{Adaptation} = .995$ ) and resulted in better model fit over AFQT alone. The Can-Do composite did not predict either IMT or in-unit Disciplinary Incidents (as evidenced by CIs that include 1.000 for the associated ORs).

Table 4.4 presents the results of the logistic regression analyses examining attrition through 6, 12, 24, and 36 months of service for Regular Army Soldiers. The Will-Do ( $OR_{Will-Do} = .989$ ,  $.989$ ,  $.991$ , and  $.991$ , respectively) and Adaptation ( $OR_{Adaptation} = .990$ ,  $.989$ ,  $.991$ , and  $.991$ , respectively) composites were negatively related to attrition at all four time points, and their respective inclusion in the models resulted in significantly better fit over AFQT alone. Conversely, the TAPAS Can-Do composite was positively related to attrition at each of the time points for Regular Army Soldiers ( $OR_{Can-Do} > 1.00$ ). However, the OR lower bounds of the CIs for the Can-Do composite were very near 1.0 in each of the attrition models, and these results may not represent a true effect.

**Table 4.3. Incremental Validity Estimates for TAPAS over AFQT for Predicting IMT Restarts and Disciplinary Incidents**

Criterion Measure / Model	OR <sub>AFQT</sub>	OR <sub>TAPAS</sub>	$r_{pb}$	$\Delta$ -2LL
<i>IMT Restarts</i>		<i>n = 569,312 - 697,384</i>		
Can-Do				
AFQT	.974 (.964-.985)		0.01	
AFQT + TAPAS	.979 (.968-.991)	.999 (.999-1.000)	0.01	<b>5.19</b>
Will-Do				
AFQT	.978 (.968-.988)		0.03	
AFQT + TAPAS	.988 (.978-.998)	.994 (.993-.994)	0.01	<b>587.69</b>
Adaptation				
AFQT	.974 (.964-.985)		0.03	
AFQT + TAPAS	.993 (.982-1.005)	.994 (.994-.995)	0.01	<b>391.91</b>
<i>IMT Disciplinary Incidents</i>		<i>n = 71,921 - 81,032</i>		
Can-Do				
AFQT	.976 (.960-.993)		0.01	
AFQT + TAPAS	.983 (.965-1.001)	.999 (.998-1.000)	0.01	<b>4.41</b>
Will-Do				
AFQT	.974 (.958-.989)		0.01	
AFQT + TAPAS	.989 (.974-1.005)	.988 (.988-.989)	0.10	<b>762.23</b>
Adaptation				
AFQT	.976 (.960-.993)		0.01	
AFQT + TAPAS	.997 (.980-1.014)	.994 (.993-.995)	0.05	<b>185.89</b>
<i>In-Unit Disciplinary Incidents</i>		<i>n = 6,205 - 8,203</i>		
Can-Do				
AFQT	.958 (.903-1.016)		0.02	
AFQT + TAPAS	.956 (.898-1.018)	1.000 (.997-1.003)	0.02	0.02
Will-Do				
AFQT	.945 (.898-.993)		0.02	
AFQT + TAPAS	.947 (.900-.996)	.993 (.990-.995)	0.07	<b>31.72</b>
Adaptation				
AFQT	.958 (.903 -1.016)		0.02	
AFQT + TAPAS	.972 (.915-1.031)	.995 (.992-.998)	0.04	<b>8.67</b>

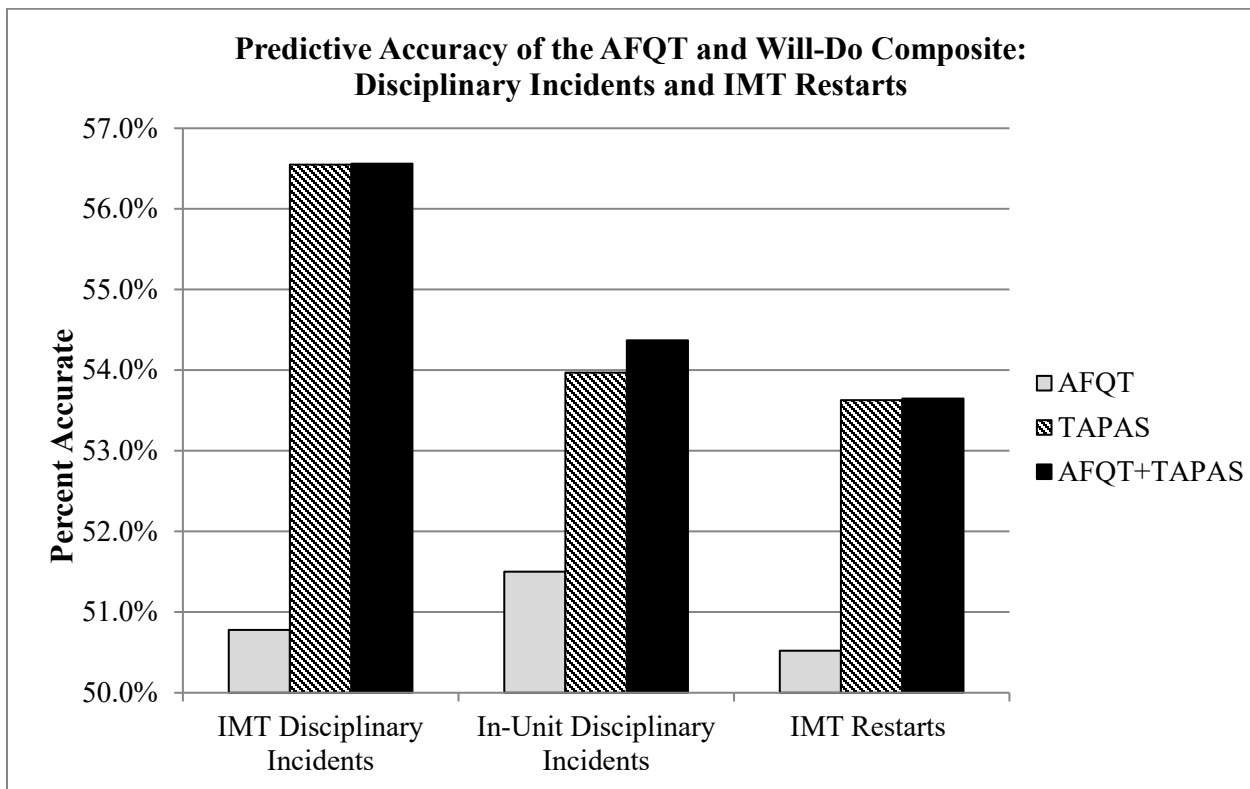
*Note.* OR = odds ratio for each predictor. Confidence interval of the odds ratio shown in parentheses.  $r_{pb}$  = point biserial correlation between the observed outcome and predicted probability.  $\Delta$ -2LL = change in negative two log likelihood (deviance) from adding the selected TAPAS composite scale to the AFQT-only logistic regression model. Odds ratios equal to 1.0 (or confidence intervals of the odds ratio that include 1.0) indicate no relationship between the predictor and criterion. Odds ratios less than 1.0 indicate a negative relationship between the predictor and criterion. Odds ratios greater than 1.0 indicate a positive relationship between the predictor and criterion. For  $\Delta$ -2LL, bolded values indicate significant change in model fit based on a Likelihood Ratio  $\chi^2$  test,  $p < .05$ .

**Table 4.4. Incremental Validity Estimates for TAPAS over AFQT for Predicting Cumulative Attrition through 36 Months of Service**

Attrition Measure / Model	OR <sub>AFQT</sub>	OR <sub>TAPAS</sub>	$r_{pb}$	$\Delta$ -2LL
<b>6 Month</b>				
<i>n</i> = 323,366 - 389,549				
Can-Do				
AFQT	.842 (.833-.851)		0.05	
AFQT + TAPAS	.834 (.825-.844)	1.001 (1.001-1.002)	0.05	<b>18.48</b>
Will-Do				
AFQT	.845 (.837-.854)		0.05	
AFQT + TAPAS	.855 (.847-.864)	.989 (.988-.989)	0.09	<b>2077.74</b>
Adaptation				
AFQT	.842 (.833-.851)		0.05	
AFQT + TAPAS	.871 (.861-.880)	.990 (.989-.990)	0.08	<b>1362.04</b>
<b>12 Month</b>				
<i>n</i> = 311,204 - 377,192				
Can-Do				
AFQT	.871 (.862-.880)		0.05	
AFQT + TAPAS	.864 (.855-.873)	1.001 (1.001-1.002)	0.05	<b>15.84</b>
Will-Do				
AFQT	.874 (.866-.882)		0.05	
AFQT + TAPAS	.884 (.876-.892)	.989 (.988-.989)	0.09	<b>2392.52</b>
Adaptation				
AFQT	.871 (.862-.880)		0.05	
AFQT + TAPAS	.901 (.892-.910)	.989 (.989-.990)	0.09	<b>1699.37</b>
<b>24 Month</b>				
<i>n</i> = 268,389 - 333,986				
Can-Do				
AFQT	.852 (.844-.860)		0.07	
AFQT + TAPAS	.844 (.836-.853)	1.001 (1.001-1.002)	0.07	<b>22.76</b>
Will-Do				
AFQT	.856 (.849-.863)		0.06	
AFQT + TAPAS	.864 (.857-.871)	.991 (.990-.991)	0.10	<b>1993.15</b>
Adaptation				
AFQT	.852 (.844-.860)		0.07	
AFQT + TAPAS	.878 (.870-.886)	.991 (.990-.991)	0.10	<b>1630.51</b>
<b>36 Month</b>				
<i>n</i> = 224,476 - 289,366				
Can-Do				
AFQT	.827 (.820-.835)		0.09	
AFQT + TAPAS	.819 (.811-.827)	1.001 (1.001-1.002)	0.09	<b>32.81</b>
Will-Do				
AFQT	.833 (.827-.840)		0.08	
AFQT + TAPAS	.840 (.833-.846)	.991 (.991-.992)	0.12	<b>1825.71</b>
Adaptation				
AFQT	.827 (.820-.835)		0.09	
AFQT + TAPAS	.850 (.843-.858)	.991 (.991-.992)	0.12	<b>1464.21</b>

*Note.* OR = odds ratio for each predictor. Confidence interval of the odds ratio shown in parentheses.  $r_{pb}$  = point biserial correlation between the observed outcome and predicted probability.  $\Delta$ -2LL = change in negative two log likelihood (deviance) from adding the selected TAPAS composite scale to the AFQT-only logistic regression model. Odds ratios equal to 1.0 (or confidence intervals of the odds ratio that include 1.0) indicate no relationship between the predictor and criterion. Odds ratios less than 1.0 indicate a negative relationship between the predictor and criterion. Odds ratios greater than 1.0 indicate a positive relationship between the predictor and criterion. For  $\Delta$ -2LL, bolded values indicate significant change in model fit based on a Likelihood Ratio  $\chi^2$  test,  $p < .05$ .

Figures 4.3 and 4.4 display the accuracy of the logistic regression models in predicting the dichotomous outcomes (Disciplinary Incidents, IMT Restarts, and attrition). For each outcome, results are presented for three models which include the following predictors: (a) AFQT, (b) TAPAS composite, and (c) AFQT + TAPAS composite. Specifically, the *percent accurate* values represent the *c* statistic, or area under the receiver operating characteristic (ROC) curve, which reflects the ability of the given model to correctly discriminate between a case and a noncase (e.g., attriter vs. stayer).<sup>13</sup> The area under the curve (AUC) can range from .50 to 1.0, corresponding to 50% (or chance) and 100% accuracy, respectively. For example, the closer AUC is to 1.0, the more likely it is that Soldiers with lower TAPAS scores will have at least one IMT Disciplinary Incident.



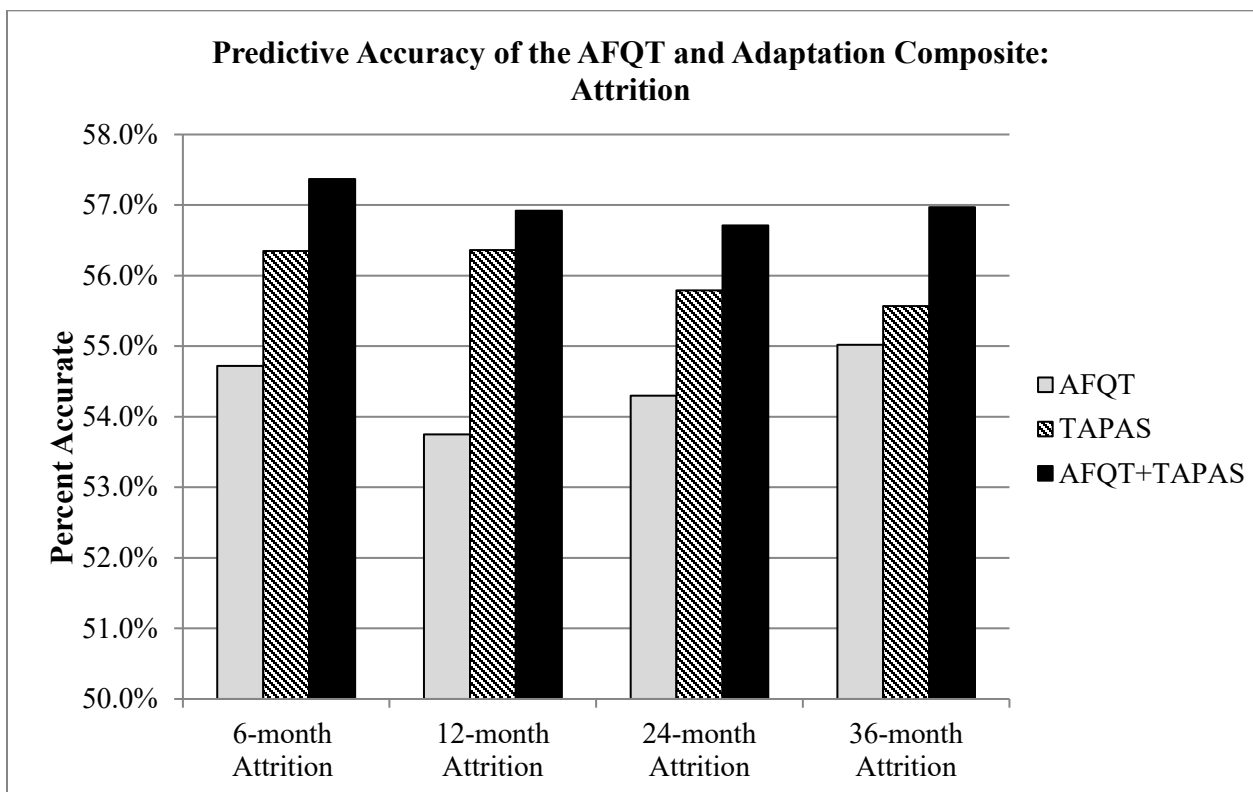
**Figure 4.3. Predictive accuracy of AFQT and TAPAS Will-Do composite in the discrimination of both IMT and in-unit Disciplinary Incidents and IMT Restarts.**

In particular, Figure 4.3 displays the results of AFQT and TAPAS Will-Do composite models in predicting IMT Restarts and Disciplinary Incidents (IMT and in-unit). For all three criteria, the probability of discriminating between Soldiers with and without incidence of the outcome of interest is lowest (less than 52.1%) when using AFQT alone. However, the predictive accuracy increases for the Will-Do composite and AFQT + Will-Do composite models for all outcomes. Note the combined model demonstrated the most accuracy, ranging from 53.7% for IMT

<sup>13</sup> The ROC curve is a plot of sensitivity (i.e., probability of detecting a true positive) versus specificity (i.e., probability of detecting a true negative) across a range of potential cut scores for continuous predictors in a logistic regression model (Cook, 2007). For the purposes of evaluating the logistic regression models discussed in this report, only the area under the ROC curve (i.e., AUC) is presented and discussed.

Restarts to 56.6% for IMT Disciplinary Incidents. In addition, the increase in discrimination accuracy for the combined model compared to the Will-Do composite-only model was generally small, suggesting that AFQT adds little value to the discrimination of these dichotomous outcomes.

Figure 4.4 displays the results of the AFQT and TAPAS Adaptation composite models in predicting attrition for Regular Army Soldiers. For attrition, the combined AFQT + Adaptation composite model resulted in the highest discrimination accuracy (approximately 57% for each time point). When comparing the AFQT-only and Adaptation composite-only models, the Adaptation composite-only model evidenced at least somewhat higher accuracy than the AFQT-only at all four time points, ranging from 1.4% for 36-month attrition to 0.5% for 12-month attrition (increases were 1.0% for 6-month attrition and 0.9% for 24-month attrition).



**Figure 4.4. Predictive accuracy of AFQT and TAPAS Adaptation composite in the discrimination of attrition outcomes for Regular Army Soldiers.**

## Summary

This chapter summarizes results from the 20<sup>th</sup> cycle of the evaluation of criterion-related validity of TAPAS. Results presented here align with the findings from prior analysis cycles showing the TAPAS composites demonstrated incremental validity over AFQT in predicting important first-term Soldier outcomes. For both IMT and in-unit outcomes, the Will-Do TAPAS composite demonstrated the strongest relationships with APFT score, Peer Leadership, and Resilience. Results of the incremental validity analyses in particular showed that the Will-Do composite contributed beyond AFQT in predicting these outcomes. Regarding dichotomous outcomes, the Will-Do composite helped predict Disciplinary Incidents and IMT Restarts better than AFQT alone. The Adaptation composite also contributed to the prediction of attrition beyond AFQT alone. Given the well-established strength of AFQT to predict cognitive outcomes, it is unsurprising that TAPAS does not help in the prediction of the knowledge-based outcomes.

## 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 29-51). 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.
- Bynum, B. H., & Mullins, H. M. (Eds.) (2017). *Tier One Performance Screen Initial operational test and evaluation: 2013 annual report* (Technical Report 1359). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- 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., 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. 307-334). Hillsdale, NJ: Erlbaum.
- Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology*, 43, 313-333.
- Carretta, T.R., Ree, M.J., & Teachout, M.S. (2016). Training affects variability in training performance both within and across jobs. *International Journal of Selection and Assessment*, 24, 72-76.
- Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd edition)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115, 928-935.
- DMDC (2008). *ASVAB Technical Bulletin No. 3 CAT-ASVAB Forms 5-9*. Seaside, CA: Defense Manpower Data Center.
- 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.

- Graves, C., & Taylor, W. (2020). Description and psychometric properties of objective criterion measures. In D.J. Knapp & C.D. Kirkendall (Eds.) *Tier One Performance Screen initial operational test and evaluation: Capstone report* (Technical Report 1380). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Hughes, M. G. (2018). Evidence for the predictive validity of the TAPAS. In D.J. Knapp & C.D. Kirkendall (Eds.) *Tier One Performance Screen initial operational test and evaluation: 2015-2016 biennial report* (Technical Report 1367). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Intelligence*, 29, 340-362.
- 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., & 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., Salmon, E., Kirkendall, C.D., & Burgoyne, T. (2020). Description and psychometric properties of performance rating criterion measures. In D.J. Knapp & C.D. Kirkendall (Eds.) *Tier One Performance Screen initial operational test and evaluation: Capstone report* (Technical Report 1380). 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., & Kirkendall, C.D. (Eds.) (2018). *Tier One Performance Screen initial operational test and evaluation: 2015-2016 biennial report* (Technical Report 1367). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

- Knapp, D.J., & Kirkendall, C.D. (Eds.) (2020). *Tier One Performance Screen initial operational test and evaluation: Capstone report* (Technical Report 1380). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & LaPort, K. (Eds.) (2013a). *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., & LaPort, K. (Eds.) (2013b). *Tier One Performance Screen initial operational test and evaluation: 2012 interim report* (Technical Report 1332). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & LaPort, K. (Eds.) (2014). *Tier One Performance Screen initial operational test and evaluation: 2012 annual report* (Technical Report 1342). 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.
- Knapp, D. J., Sager, C. E., & Tremble, T. R. (Eds.). (2005) *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & Tremble, T. R. (Eds.) (2007). *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & Wolters, H. M. K. (Eds.) (2017). *Tier One Performance Screen initial operational test and evaluation: 2014 annual report* (Technical Report 1358). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Maier, M.H. (1993). *Military aptitude testing: The past fifty years* (DMDC Technical Report 93-007). Monterey, CA: Defense Manpower Data Center.
- 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.
- Nesbitt, C.K., Salmon, E., & Kirkendall, C.D. (2020). Army Life Questionnaire. In D.J. Knapp & C.D. Kirkendall (Eds.) *Tier One Performance Screen initial operational test and evaluation: Capstone report* (Technical Report 1380). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

- 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. 135-155). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- 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.
- Stark, S.E., Chernyshenko, O.S., Nye, C., & Drasgow, F. (2018). The Tailored Adaptive Personality Assessment System (TAPAS). In D.J. Knapp & C.D. Kirkendall (Eds.) *Tier One Performance Screen initial operational test and evaluation: 2015-2016 biennial report* (Technical Report 1367). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Stark, S. E., Chernyshenko, O. S., & Drasgow, 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 29-51). 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). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6<sup>th</sup> ed.). Boston, MA: Pearson.
- Thacker, A.A., Russell, T.L., Knapp, D.J., & Gribben, M.A. (2020). Validity argument 1.0 for the Armed Forces Qualification Test (AFQT) (2020 No. 07). Alexandria, VA: Human Resources Research Organization.
- 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). Arlington, VA: 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.

Wolters, H., White, L., Drasgow, F., Stark, S., Nye, C., & Chernyshenko, O.S. (July, 2017). *TAPAS research update: Ongoing and future research*. Presentation given to the Defense Advisory Committee on Military Personnel Testing.

Zhang, B., Sun, T., Drasgow, F., Chernyshenko, O.S., Nye, C.D., Stark, S., & White, L.A. (2020). Though forced, still valid: Psychometric equivalence of forced-choice and single-statement measures. *Organizational Research Methods*, 23(3), 569-590.

## APPENDIX A

### LIST OF FREQUENTLY USED ACRONYMS

AFQT	Armed Forces Qualification Test
AHRC	Army Human Resources Command
AIT	Advanced Individual Training
ALQ	Army Life Questionnaire
APFT	Army Physical Fitness Test
ASVAB	Armed Services Vocational Aptitude Battery
ATRRS	Army Training Requirements and Resources System
ATSC	Army Training Support Center
CSB	Counterproductive Soldier Behavior
DMDC	Defense Manpower Data Center
EEEM	Expanded Enlistment Eligibility Metrics
ICTL	Information, Communication, and Technical Literacy
IMT	Initial Military Training
IRR	Inter-rater reliability
JKT	Job knowledge test
MEPCOM	Military Entrance Processing Command
MOS	Military Occupational Specialty
MTL	Motivation to Lead
NCO	Noncommissioned officer
NPS	Non-prior service
OCB	Organizational citizenship behavior
PRS-P	Performance rating scales completed by peers
PRS-S	Performance rating scales completed by supervisors or cadre
RITMS	Resident Individual Training Management System
SME	Subject matter expert
TAPAS	Tailored Adaptive Personality Assessment System
TAPDB	Total Army Personnel Database
TOPS IOT&E	Tier One Performance Screen initial operational test and evaluation
WTBD JKT	Warrior tasks and battle drills job knowledge test

APPENDIX B

BIVARIATE CORRELATIONS AMONG STUDY VARIABLES

*Table B.1. Correlations among AFQT, TAPAS Composites, and Continuous IMT and In-Unit Criteria*

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. AFQT		<b>0.36</b>	<b>0.08</b>	<b>0.17</b>	<b>0.43</b>	<b>0.40</b>	<b>-0.03</b>	<b>-0.09</b>	<b>-0.05</b>	<b>0.06</b>	-0.01	-0.01	<b>0.04</b>	<b>0.10</b>
<i>TAPAS Composites</i>														
2. Can-Do	<b>0.36</b>		<b>0.19</b>	<b>0.40</b>	<b>0.22</b>	<b>0.21</b>	-0.01	-0.03	-0.01	-0.02	0.01	-0.01	0.00	<b>0.04</b>
3. Will-Do	<b>0.08</b>	<b>0.19</b>		<b>0.50</b>	0.02	<b>0.05</b>	<b>0.06</b>	0.00	<b>0.19</b>	-0.02	<b>0.18</b>	<b>0.21</b>	<b>0.10</b>	<b>0.09</b>
4. Adaptation	<b>0.17</b>	<b>0.40</b>	<b>0.50</b>		<b>0.14</b>	<b>0.13</b>	0.01	-0.01	-0.02	0.03	<b>0.05</b>	<b>0.12</b>	0.02	<b>0.05</b>
<i>Criteria</i>														
5. Knowledge & Skill <sup>a</sup>	<b>0.43</b>	<b>0.22</b>	<b>0.03</b>	<b>0.10</b>		<b>0.82</b>	<b>0.08</b>	0.03	<b>0.08</b>	-0.02	<b>0.06</b>	0.02	<b>0.07</b>	<b>0.11</b>
6. WTBD JKT	<b>0.37</b>	<b>0.18</b>	<b>0.05</b>	<b>0.10</b>	<b>0.82</b>		<b>0.13</b>	0.03	<b>0.08</b>	-0.03	<b>0.06</b>	<b>0.07</b>	<b>0.09</b>	<b>0.17</b>
7. Commitment & Fit <sup>a</sup>	<b>-0.04</b>	<b>0.02</b>	<b>0.13</b>	<b>0.02</b>	<b>0.12</b>	<b>0.15</b>		<b>0.60</b>	<b>0.40</b>	<b>-0.29</b>	<b>0.45</b>	<b>0.08</b>	<b>0.11</b>	<b>0.15</b>
8. Retention Cognitions <sup>a</sup>	<b>-0.11</b>	<b>-0.01</b>	<b>0.05</b>	<b>-0.03</b>	<b>0.01</b>	<b>0.01</b>	<b>0.52</b>		<b>.27</b>	<b>-0.25</b>	<b>0.29</b>	<b>0.07</b>	<b>0.11</b>	<b>0.06</b>
9. Peer Leadership <sup>a</sup>	<b>-0.04</b>	<b>0.03</b>	<b>0.22</b>	<b>-0.02</b>	<b>0.04</b>	0.01	<b>0.40</b>	<b>0.23</b>		<b>-0.10</b>	<b>0.47</b>	<b>0.15</b>	<b>0.14</b>	<b>0.13</b>
10. CSB	0.00	-0.01	<b>-0.07</b>	0.01	<b>-0.16</b>	<b>-0.13</b>	<b>-0.33</b>	<b>-0.17</b>	<b>-0.04</b>		<b>-0.26</b>	<b>-0.05</b>	<b>-0.10</b>	<b>-0.06</b>
11. Resilience	<b>-0.03</b>	<b>0.01</b>	<b>0.21</b>	<b>0.07</b>	<b>0.08</b>	<b>0.06</b>	<b>0.52</b>	<b>0.26</b>	<b>0.47</b>	<b>-0.29</b>		<b>0.13</b>	<b>0.15</b>	<b>0.16</b>
12. APFT	<b>0.09</b>	<b>0.01</b>	<b>0.25</b>	<b>0.17</b>	<b>0.04</b>	<b>0.06</b>	<b>0.07</b>	<b>0.03</b>	<b>0.12</b>	<b>-0.04</b>	<b>0.13</b>		<b>0.18</b>	<b>0.18</b>
13. Peer Performance Rating <sup>a</sup>	<b>0.07</b>	<b>0.03</b>	<b>0.14</b>	<b>0.10</b>	<b>0.08</b>	<b>0.07</b>	<b>0.08</b>	<b>0.04</b>	<b>0.13</b>	<b>-0.08</b>	<b>0.14</b>	<b>0.28</b>		<b>0.35</b>
14. Supervisor Performance Rating (in-unit only) <sup>a</sup>														

*Note.* Correlations below the diagonal reflect the IMT criteria,  $n = 9,304-697,975$ . Correlations above the diagonal reflect the in-unit criteria,  $n = 347-697,527$ . Lower sample sizes are associated with new ALQ scales and higher sample sizes are associated with administrative data. Missing values reflect the scales that were not administered either in IMT or in units. Correlations in bold are statistically significant,  $p < .05$  (two-tailed).

<sup>a</sup>Indicates a criterion composite score.

**Table B.2. Correlations among AFQT, TAPAS Composites, and Continuous Criteria with Dichotomous IMT and In-Unit Criteria**

Scale	Disciplinary Incidents	IMT Restarts	Attrition							
			IMT				In-Unit			
			6-month	12-month	24-month	36-month	6-month	12-month	24-month	36-month
1. AFQT	<b>-0.01</b>	0.00	<b>-0.06</b>	<b>-0.06</b>	<b>-0.08</b>	<b>-0.10</b>	<b>-0.06</b>	<b>-0.06</b>	<b>-0.08</b>	<b>-0.10</b>
<i>TAPAS Composites</i>										
2. Can-Do	<b>-0.01</b>	<b>-0.01</b>	<b>-0.01</b>	<b>-0.02</b>	<b>-0.02</b>	<b>-0.02</b>	<b>-0.01</b>	<b>-0.02</b>	<b>-0.02</b>	<b>-0.02</b>
3. Will-Do	<b>-0.10</b>	<b>-0.03</b>	<b>-0.07</b>	<b>-0.09</b>	<b>-0.09</b>	<b>-0.09</b>	<b>-0.07</b>	<b>-0.09</b>	<b>-0.09</b>	<b>-0.09</b>
4. Adaptation	<b>-0.05</b>	<b>-0.03</b>	<b>-0.07</b>	<b>-0.08</b>	<b>-0.09</b>	<b>-0.10</b>	<b>-0.07</b>	<b>-0.08</b>	<b>-0.09</b>	<b>-0.10</b>
<i>Criteria</i>										
5. Knowledge & Skill <sup>a</sup>	<b>-0.04</b>	<b>-0.01</b>	<b>-0.01</b>	<b>-0.03</b>	<b>-0.08</b>	<b>-0.09</b>	0.00	-0.01	<b>-0.07</b>	<b>-0.11</b>
6. WTBD JKT	<b>-0.03</b>	<b>-0.01</b>	<b>-0.01</b>	<b>-0.03</b>	<b>-0.09</b>	<b>-0.12</b>	0.00	0.01	<b>-0.03</b>	<b>-0.06</b>
7. Commitment & Fit <sup>a</sup>	<b>-0.07</b>	<b>-0.03</b>	<b>-0.05</b>	<b>-0.07</b>	<b>-0.08</b>	<b>-0.10</b>	0.02	-0.01	<b>-0.05</b>	<b>-0.09</b>
8. Retention Cognitions <sup>a</sup>	<b>-0.04</b>	<b>0.01</b>	<b>-0.03</b>	<b>-0.06</b>	<b>-0.06</b>	<b>-0.05</b>	0.00	0.00	<b>-0.06</b>	<b>-0.12</b>
9. Peer Leadership <sup>a</sup>	<b>-0.07</b>	<b>-0.02</b>	<b>-0.02</b>	<b>-0.04</b>	<b>-0.04</b>	<b>-0.08</b>	0.00	0.00	<b>-0.04</b>	<b>-0.09</b>
10. CSB	<b>0.07</b>	<b>0.02</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.08</b>	0.00	0.00	0.00	0.03
11. Resilience	<b>-0.07</b>	<b>-0.03</b>	<b>-0.04</b>	<b>-0.05</b>	<b>-0.07</b>	<b>-0.07</b>	0.00	0.00	-0.03	<b>-0.08</b>
12. APFT	<b>-0.15</b>	<b>-0.04</b>	<b>-0.05</b>	<b>-0.06</b>	<b>-0.11</b>	<b>-0.15</b>	0.00	-0.02	<b>-0.05</b>	<b>-0.10</b>
13. Peer Performance Rating <sup>a</sup>	<b>-0.17</b>	<b>-0.05</b>	<b>-0.05</b>	<b>-0.10</b>	<b>-0.17</b>	<b>-0.28</b>	0.00	0.00	<b>-0.10</b>	<b>-0.16</b>
14. Supervisor Performance Rating <sup>a</sup>	<b>-0.18</b>	<b>-0.06</b>	--	--	--	--	0.00	<b>-0.05</b>	<b>-0.13</b>	<b>-0.18</b>

Note.  $n = 2,156-1,056,293$  for IMT criteria and  $n = 1,583-953,885$  for in-unit criteria. Lower sample sizes are associated with new ALQ scales and higher sample sizes are associated with administrative data. Correlations in bold are statistically significant,  $p < .05$  (two-tailed).

<sup>a</sup>Indicates a composite score.