



Research Note 2024-03

**Validation of Accessions Screening Tools:
Cycle 24 Evaluation**

Deirdre J. Knapp

Human Resources Research Organization

Cristina D. Kirkendall

U.S. Army Research Institute

December 2023

**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, G-1**

Authorized and approved:

**SCOTT B. SHADRICK, Ph.D.
Acting Director**

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

Human Resources Research Organization

Technical Review by:

Dr. Julia Whitaker, 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 (Month Year) December 2023		2. REPORT TYPE Interim		3. DATES COVERED (Month Year)
				START DATE April 2019
				END DATE October 2022
4. TITLE AND SUBTITLE Validation of Accessions Screening Tools: Cycle 24 Evaluation				
5a. CONTRACT NUMBER W911NF-19-C-0032		5b. GRANT NUMBER		5c. COOPERATIVE AGREEMENT NUMBER
5d. PROGRAM ELEMENT NUMBER 633007	5e. PROJECT NUMBER A792	5f. TASK NUMBER 1031	5g. WORK UNIT NUMBER	
6. AUTHOR(S) Deirdre J. Knapp & Cristina D. Kirkendall				
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 2023 No. 044
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop: 5610) Fort Belvoir, Virginia 22060-5610		10. SPONSOR/MONITOR'S ACRONYM(S) ARI	11. SPONSOR/MONITOR'S REPORT NUMBER(S) Research Note 2024-03	
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 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 for the Behavioral and Social Sciences (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 part of 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</i> (VAST). 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 enlisted personnel selection.				
15. SUBJECT TERMS Behavioral and social science Personnel Manpower Selection and classification				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	Unlimited Unclassified	65
19a. NAME OF RESPONSIBLE PERSON Tonia S. Heffner			19b. PHONE NUMBER (Include area code) 703-545-4408	

PREVIOUS EDITION IS OBSOLETE

STANDARD FORM 298 (REV. 5/2020)

Prescribed by ANSI Std. Z39.18

Research Note 2024-03

**Validation of Accessions Screening Tools:
Cycle 24 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

December 2023

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. 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	CSM Shawn McDonald
MSG Andre Betton	Mr. Clifford McMillan
MSG Jenette Blamer	SFC Samuel Meadows
Mr. Shawn Bova	1SG Luis Miranda
SGM Thomas Bray	SFC Newtina Mitchell
SGM Mark Clark Jr.	SGM Jeffrey Morris
SGM Daniel Colonmendez	CW5 Anthony Moschella
SGM Kimberly Costello	Mr. Wilfredo Norat
SFC Randal Davenport	SFC Anthony Northcutt
MSG Christopher Evans	SGM Sharita Onugha
SFC Jose Flores	MSG Todd Pameticky
SFC Alonna Ford	SGM Terry Pittman
SGM Bryant Fowle	Mr. John Plotts
Mr. Alphonsa Green	Mr. Gerald Purcell
MSG Daniel Griego	Dr. Aryn Pyke
Mr. Todd Grisso	MSG Joel Raglin
SFC Jerome Hanson	MSG Derrian Richardson
COL James Hayes	SGM Jaclyn Rivera
Mr. Lawrence Hayward Jr.	MSG Timothy Robinson
MSG Julio Hernandez	SFC Marcelo Rubalcava
1SG Thomas Hill Jr.	Mr. Mark Russo
MSG Kaulai Hollis	SFC Michael Rust
MSG Adam Jacobs	SFC Fred Scaldaferrri
SFC Saba James	SFC Christopher Shirley
SGM Louis Johnson Jr.	CSM Todd Shirley
Ms. Amanda Kelley	SGM Thomas Skelton
SFC James Kennedy	SFC Marvin Skinner
SGM Gareth Kilpatrick	1SG Christopher Stanifer
SFC Clifford Kurten	CW4 Keisha Towles
SGM Alfredo Lassiseramirez	MSG Wesley Weaver
Mr. Lamar Lauderdale	Dr. Natalya Weber
SFC Ricardo Martinez	Mr. Thomas Wetzel Sr.

Finally, we would like to acknowledge the efforts of HumRRO staff Martin Yu (database construction), Nicholas Howald and Mitchell Pontikes (analysts) as well as Doug Brown (IMT data collection support), Chris Graves (in-unit data collection manager), and their respective teams in conducting the work described in this report.

VALIDATION OF ACCESSIONS SCREENING TOOLS (VAST): CYCLE 24 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. Research has demonstrated that non-cognitive attributes are important to entry-level Soldier performance and retention (e.g., Campbell & Knapp, 2001; Ingerick et al., 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, the U.S. Army Research Institute for the Behavioral and Social Sciences (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 extract 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., attrition). 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 February 2022 and criterion data collected through January 2022. The TAPAS database consists of over 1.4M applicants. The validation sample sizes are considerably smaller, with the IMT Validation Sample comprising 106,797 Soldiers, the In-Unit Validation Sample comprising 9,982 Soldiers,

and the Administrative Validation Sample (which includes Soldiers with criterion data [e.g., attrition] from at least one administrative source) comprising 878,846 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 interrater 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 (ΔR) 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), Adaptation, and Conduct composite scores.

Findings:

Results presented here align with the findings from prior analysis cycles (e.g., Hughes, 2018). The TAPAS composites demonstrated incremental validity over AFQT scores in predicting several important first term Soldier outcomes. In regard to prediction of continuous criterion measures, the Will-Do composite demonstrated the greatest overall incremental validity, with ΔR estimates exceeding .05 for predicting Peer Performance Rating Composite, Peer Leadership, and Army Physical Fitness Test (APFT) scores for both IMT and in-unit Soldiers, and Counterproductive Soldier Behavior and Commitment and Fit for Soldiers in IMT. The Adaptation composite yielded incremental validity estimates of .05 or greater for IMT and in-unit APFT scores, and the Conduct composite yielded these levels of incremental validity for IMT Counterproductive Soldier Behavior and APFT Score. 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), such that higher scores on these composites were associated with a reduction in the likelihood of these outcomes. The Conduct composite provided incremental validity over AFQT in predicting these criteria as well except for IMT Restarts and 6-month and 12-month attrition. 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 scores for predicting the same criteria 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 (VAST):
CYCLE 24 EVALUATION

CONTENTS

	Page
CHAPTER 1: INTRODUCTION	1
Background	1
The Tailored Adaptive Personality Assessment System (TAPAS)	2
The Validation of Accessions Screening Tools (VAST) Program	3
Overview of Report.....	4
CHAPTER 2: RESEARCH METHOD	5
Overview	5
Predictor Measures.....	5
Criterion Measures.....	8
Data Collection	21
Biannual Longitudinal Database Updates.....	21
Analysis Plan Overview.....	22
CHAPTER 3: CYCLE 24 SAMPLE CHARACTERISTICS AND DESCRIPTIVE STATISTICS	23
Description of Analysis Samples	23
Descriptive Statistics for Predictor Measures	26
Descriptive Statistics for Criterion Measures	27
Summary	29
CHAPTER 4: EVIDENCE FOR THE PREDICTIVE VALIDITY OF THE TAPAS.....	30
Bivariate Correlations	30
Incremental Validity Analyses.....	32
Summary	47
REFERENCES	49

APPENDIX

APPENDIX. LIST OF FREQUENTLY USED ACRONYMS.....	A-1
---	-----

LIST OF TABLES

TABLE 2.1	TAPAS VERSIONS BY ADMINISTRATION DATE	6
TABLE 2.2	TAPAS FACET NAMES AND DEFINITIONS.....	7
TABLE 2.3	INITIAL MILITARY TRAINING (IMT) TAPAS VALIDATION CRITERION MEASURES.....	9
TABLE 2.4	IN-UNIT TAPAS VALIDATION CRITERION MEASURES	10
TABLE 2.5	ATTRITION CRITERION MEASURES.....	11
TABLE 2.6	ARMY LIFE QUESTIONNAIRE SCALE AND COMPOSITE SCORE RELIABILITY ESTIMATE.....	12
TABLE 2.7	IMT PEER PERFORMANCE RATING SCALE DIMENSIONS AND ASSOCIATED EXAMPLE BEHAVIORS	13
TABLE 2.8	IN-UNIT PEER PERFORMANCE RATING SCALE DIMENSIONS AND ASSOCIATED EXAMPLE BEHAVIORS	15
TABLE 2.9	PEER PERFORMANCE RATING SCALE INTERRATER RELIABILITY ESTIMATES.....	16
TABLE 2.10	IN-UNIT PRS-S ARMY-WIDE RATING SCALE DIMENSIONS AND ASSOCIATED DEFINITIONS.....	18
TABLE 3.1	BACKGROUND AND DEMOGRAPHIC CHARACTERISTICS OF THE ANALYSIS SAMPLES.....	24
TABLE 3.2	MEANS AND STANDARD DEVIATIONS FOR THE TAPAS COMPOSITES BY TAPAS VERSION	26
TABLE 3.3	DESCRIPTIVE STATISTICS AND INTERCORRELATIONS AMONG AFQT AND TAPAS COMPOSITES IN THE APPLICANT SAMPLE	26
TABLE 3.4	DESCRIPTIVE STATISTICS FOR CRITERION SCORES IN THE IMT VALIDATION SAMPLE.....	27
TABLE 3.5	DESCRIPTIVE STATISTICS FOR CRITERION SCORES IN THE IN- UNIT VALIDATION SAMPLE	28
TABLE 3.6	BASE RATES FOR DICHOTOMOUS CRITERIA IN THE ADMINISTRATIVE VALIDATION SAMPLE.....	28

	Page
TABLE 4.1 UNCORRECTED ZERO-ORDER CORRELATIONS BETWEEN PREDICTORS AND CONTINUOUS CRITERIA	31
TABLE 4.2 UNCORRECTED ZERO-ORDER CORRELATIONS BETWEEN PREDICTORS AND DICHOTOMOUS CRITERIA	32
TABLE 4.3 INCREMENTAL VALIDITY ESTIMATES FOR TAPAS COMPOSITES OVER AFQT FOR PREDICTING IMT CRITERIA.....	34
TABLE 4.4 INCREMENTAL VALIDITY ESTIMATES FOR TAPAS COMPOSITES OVER AFQT FOR PREDICTING IN-UNIT CRITERIA	36
TABLE 4.5 INCREMENTAL VALIDITY ESTIMATES FOR TAPAS COMPOSITES OVER AFQT FOR PREDICTING IMT RESTARTS AND DISCIPLINARY INCIDENTS.....	38
TABLE 4.6 INCREMENTAL VALIDITY ESTIMATES FOR TAPAS COMPOSITES OVER AFQT FOR PREDICTING CUMULATIVE ATTRITION THROUGH 36 MONTHS OF SERVICE (REGULAR ARMY ONLY).....	43

LIST OF FIGURES

FIGURE 2.1 SAMPLE IMT PEER RATER PERFORMANCE RATING SCALES	14
FIGURE 2.2 SAMPLE IN-UNIT SUPERVISOR RATER PERFORMANCE RATING SCALE.....	19
FIGURE 2.3 OVERVIEW OF VAST DATA FILE MERGING AND NESTED SAMPLE GENERATION PROCESS.....	22
FIGURE 4.1 INCREASE IN PREDICTION OF IMT CONTINUOUS CRITERIA USING TAPAS COMPOSITES ($\Delta R > .05$ ONLY)	35
FIGURE 4.2 INCREASE IN PREDICTION OF IN-UNIT CONTINUOUS CRITERIA USING TAPAS COMPOSITES ($\Delta R > .05$ ONLY)	37
FIGURE 4.3 AUC STATISTICS OF AFQT AND TAPAS CAN-DO COMPOSITE IN THE DISCRIMINATION OF BOTH IMT AND IN-UNIT DISCIPLINARY INCIDENTS AND IMT RESTARTS.....	40

	Page
FIGURE 4.4 AUC STATISTICS OF AFQT AND TAPAS WILL-DO COMPOSITE IN THE DISCRIMINATION OF BOTH IMT AND IN-UNIT DISCIPLINARY INCIDENTS AND IMT RESTARTS.....	41
FIGURE 4.5 AUC STATISTICS OF AFQT AND TAPAS ADAPTATION COMPOSITE IN THE DISCRIMINATION OF BOTH IMT AND IN-UNIT DISCIPLINARY INCIDENTS AND IMT RESTARTS	41
FIGURE 4.6 AUC STATISTICS OF AFQT AND TAPAS CONDUCT COMPOSITE IN THE DISCRIMINATION OF BOTH IMT AND IN-UNIT DISCIPLINARY INCIDENTS AND IMT RESTARTS.....	42
FIGURE 4.7 AUC STATISTICS OF AFQT AND TAPAS CAN-DO COMPOSITE IN THE DISCRIMINATION OF CUMULATIVE ATTRITION.....	45
FIGURE 4.8 AUC STATISTICS OF AFQT AND TAPAS WILL-DO COMPOSITE IN THE DISCRIMINATION OF CUMULATIVE ATTRITION.....	46
FIGURE 4.9 AUC STATISTICS OF AFQT AND TAPAS ADAPTATION COMPOSITE IN THE DISCRIMINATION OF CUMULATIVE ATTRITION	46
FIGURE 4.10 AUC STATISTICS OF AFQT AND TAPAS CONDUCT COMPOSITE IN THE DISCRIMINATION OF CUMULATIVE ATTRITION	47

VALIDATION OF ACCESSIONS SCREENING TOOLS (VAST): CYCLE 24 EVALUATION

CHAPTER 1: INTRODUCTION¹

Deirdre J. Knapp (HumRRO) and Cristina D. Kirkendall (ARI)

Background

The U.S. Army's People Strategy is geared toward ensuring a ready, professional, diverse, and integrated force (U.S. Army, 2019). The strategy envisions a variety of ways to address these goals, including "organizational research that employs scientific methods to develop and deploy leading edge approaches to solving Army people challenges" (p. 9). The Selection and Assignment Research Unit (SARU) of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has conducted such research for over 50 years. Over the past decade, ARI has investigated the Tailored Adaptive Personality Assessment System (TAPAS) as a tool to allow the Army to have a more *whole person* approach to selecting enlisted personnel and classifying them into Army occupations. Indeed, the latest People Strategy implementation plan (U.S. Army, 2020) specifically calls for using TAPAS scores in two pilot programs to support recruiting goals by expanding the definition of high-quality, prospective enlistees. Further research on TAPAS is the focus of the present report. We begin, however, with providing further contextual background.

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.² Although AFQT has proven to be, and will continue to serve as, a useful metric for selecting new Soldiers, other personal characteristics, specifically non-cognitive attributes (e.g., temperament, interests, and values), are important to entry-level Soldier performance and retention (e.g., Knapp & Tremble, 2007). Importantly, broadening the lens through which potential new recruits are considered could result in better decision-making and is critical to ensure a demographically diverse enlisted force. That said, the historical obstacle to using information about traits such as temperament and interests in a large-scale personnel system has been measuring these characteristics in a manner that prevents applicants from distorting their responses to meet enlistment standards.

The ASVAB has undergone improvements over the years, but the basic content has remained largely unchanged since 1976 when it was initially introduced as the selection and classification battery to be used by all the U.S. military services. Potential content to incorporate into or be used in conjunction with ASVAB has included personality measures (e.g., TAPAS, Drasgow et al., 2012; Navy Computer Adaptive Personality Scales [NCAPS], Houston et al., 2006), the Cyber Test (Trippe et al., 2017), interest measures such as the Army Vocational Interest Diagnostic (AVID; Kirkendall et al., 2020) and the Job Opportunities in the Navy (JOIN;

¹ Much of the text in this chapter is drawn from Knapp & Kirkendall (2023). Introduction. In D.J. Knapp & C.D. Kirkendall (Eds.) *Exploring personality and performance criterion relationships using the Tailored Personality Assessment System (TAPAS)*. ARI Technical Report.

² A list of defined acronyms commonly used in this report is provided in Appendix A.

Farmer et al., 2006), and improved physical fitness measures (Sharp et al., 2018). An initiative informally known as ASVAB 2.0 is also underway to take a systematic look at all ASVAB content to reassess what would be of most value to include to support enlisted talent management needs across the U.S. military services.

The Tailored Adaptive Personality Assessment System (TAPAS)

ARI's research on the TAPAS dates to approximately 2008 and was preceded by earlier lines of research seeking ways to improve the Army's selection and classification systems by expanding the psychological assessment information on which these systems are based (i.e., cognitive ability) and broadening consideration of outcomes of interest (e.g., to include motivational aspects of job performance). Related prior ARI research programs include the seminal Army Project A conducted in the 1980s through early 1990s (Campbell & Knapp, 2001) and the Select21 and Army Class programs conducted in the early 2000s (Allen et al., 2016; Knapp et al., 2005; Knapp & Tremble, 2007).

Validating Future Force Performance Measures (Army Class) examined the potential of several experimental non-cognitive measures (e.g., temperament and person-environment fit) developed in the Select21 project for predicting important 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 et al., 2016; Ingerick et al., 2009; Knapp & Heffner, 2009). In a contemporaneous but separate effort, TAPAS was developed by the Drasgow Consulting Group (DCG) under a Small Business Innovation Research program (Drasgow et al., 2012). Given the promise of TAPAS for addressing applicant faking concerns by using the latest advances in psychometric theory, it was integrated into late stages of the Army Class validation effort and surfaced as one of the most promising measures investigated in that research (Knapp & Heffner, 2010).

Based on this research, the Army requested that TAPAS be administered to Army applicants at Military Entrance Processing Stations (MEPS) on the computer-adaptive platform for the ASVAB (CAT-ASVAB). Testing of Education Tier 1 (primarily high school diploma graduates) applicants began in May 2009 and was subsequently expanded to all Army applicants. In an associated research program dubbed the *Tier One Performance Screen (TOPS) initial operational test and evaluation (IOT&E)*, ARI investigated the potential for TAPAS to identify Tier 1 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" applicants with relatively low AFQT scores, allowing more of these applicants to access into the Army (hence the research program title). However, changing economic conditions could spur use of TAPAS as a tool to screen out unmotivated applicants, thus making the selection criteria to enter the Army more stringent. As this change in strategy illustrates, decision rules regarding the derivation and use of TAPAS composite scores could be adjusted in myriad ways to reflect evolving recruiting conditions.

ARI's research findings prompted other U.S. military services to begin thinking about how they might want to use TAPAS, and they have been conducting their own research (e.g., Rose et al., 2013; Trent et al., 2020). Increasing interest also prompted a comprehensive evaluation of TAPAS

sponsored by the Defense Testing & Assessment Center (DTAC), which was completed in 2019 (Roberts et al., 2019). DTAC also sponsored compilation and review of a broad range of evidence related to the validity of TAPAS (Ford et al., 2021).

The Validation of Accessions Screening Tools (VAST) Program

Generation of criterion-related validity evidence for using TAPAS for Army enlisted personnel selection, which initially began as part of the TOPS IOT&E project, is being continued in the current research effort, *Validation of Accessions Screening Tools (VAST)*. While the method of conducting ongoing data collections and biannual criterion-related validation analyses is the same as it was under the TOPS IOT&E, the resulting VAST database is expressly intended to support myriad research efforts pertinent to Army enlisted talent acquisition and management needs, including but not limited to investigations related to TAPAS.

Under VAST, ARI collects training outcome data on Soldiers in 71 military occupational specialties (MOS) as they complete initial military training (IMT). The criterion (outcome) measures include job knowledge tests, an attitudinal assessment, and peer performance ratings. These measures are computer-administered in a process overseen by Army personnel with guidance and support from both ARI and HumRRO.

Criterion data are also collected from Soldiers and their supervisors in units during annual data collection trips to major Army installations. These proctored “in-unit” data collections began in January 2011 and target Soldiers who accessed into the Army since approximately 2010. The in-unit criterion measures parallel those administered in IMT, but also include supervisor performance ratings. Separation status and other administrative data for all Soldiers is updated biannually.

Both IMT and in-unit data collection activity was temporarily suspended in 2020 primarily due to the COVID-19 pandemic. In-unit data collection activity restarted in early 2021 and IMT data collection activity restarted in the second half of 2021.

The VAST (and previously the TOPS IOT&E) schedule specifies biannual database and TAPAS criterion-related analysis updates. Because of the temporary suspension of data collection activities, Cycle 22 included only new administrative data (Nesbitt et al., 2022) and Cycle 23 was skipped altogether. This report documents the Cycle 24 database and analysis updates. Prior ARI TAPAS biannual validation efforts are described in a series of reports (Bynum & Mullins, 2017; Knapp & Heffner, 2011, 2012; Knapp et al., 2011; Knapp & Kirkendall, 2018, 2020, 2021, 2023; Knapp & LaPort, 2013a, 2013b, 2014; Knapp & Wolters, 2017) and internal memoranda.

To date, both the TOPS IOT&E and VAST research has focused on consideration of the incremental validity of TAPAS composite scores over the use of AFQT scores for selection. It is important to recognize that these analyses do not factor in other qualification requirements, which include moral and physical fitness screens.

Overview of Report

Chapter 2 describes the VAST predictor and criterion measures, data collection methods, and database construction strategy. Chapter 3 describes the analysis samples and provides descriptive statistics for the research variables. Chapter 4 describes criterion-related validation analyses, focusing on the incremental validity of four TAPAS composite scores over AFQT for predicting key outcomes.

CHAPTER 2: RESEARCH METHOD³

Deirdre J. Knapp (HumRRO)

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 maintaining 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 24th cycle of biannual database updates.

Predictor Measures

Armed Forces Qualification Test

The ASVAB is a multiple aptitude battery of tests administered by 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 Expression⁴ (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 et al., 2016; DMDC, 2008; Hunter, 1986; Thacker et al., 2020).

AFQT percentile scores are divided into the following categories:⁵

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

³ The author 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), Arielle Rogers (VAST database construction manager), and their respective teams in conducting the work described in this chapter.

⁴ Verbal Expression is a scaled combination of the Word Knowledge (WK) and Paragraph Comprehension (PC) tests.

⁵ For more information on ASVAB scoring, see 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-IIIA accessions.

Tailored Adaptive Personality Assessment System (TAPAS)

TAPAS consists of pools of statements descriptive of various facets of temperament (Drasgow et al., 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.

As summarized in Table 2.1, there have been multiple versions of TAPAS. The versions vary in which of over two dozen facets of personality are measured. The different versions have allowed ARI to explore the value of alternative facets and to retire the original statement pools that were exposed in research settings. All versions starting with v5 use DoD-owned statement pools. In the present report, we use data from all TAPAS versions except for the first one (13D-CAT) in which some of the facets have minor conceptual dissimilarities from the same facets in later versions.

Table 2.1

TAPAS Versions by Administration Date

TAPAS Version	Dates Administered	No. of Facets	Adaptive	No. of Items
13D-CAT	May 4, 2009 to July 10, 2009	13	Yes	104
15D-Static	July 2009 to October 2009	15	No	120
v4	October 2009 to August 2011	15	Yes	120
v5, v7, v8	August 2011 to September 2013	15	Yes	120
v9, v10, v11	September 2013 – October 2019	13	Yes	120
v12, v13, v14	November 2019 – present	13	Yes	120

The three versions of TAPAS administered to Army applicants from 2013 to 2019, and for which we have by far the most data, each assessed 13 facets, 10 of which were common to all three versions. Since 2019, three new versions of the TAPAS have been administered to Army applicants. Each of these versions assess a common set of 13 facets (described in Table 2.2) plus four additional research-only facets that are unique to each version. The facets assessed in each version of the TAPAS were selected with a goal of maximizing the validity of this assessment for predicting several specific Army outcomes. To do so, DCG researchers used data from previous VAST data collections to identify the TAPAS facets that were the best predictors of these criteria. The 13 facets that provided the best prediction are now being assessed in TAPAS v12, v13, and v14. In contrast, the four additional research-only facets that are unique to each version were included to collect data on those facets to help determine their future utility.

Scores are standardized within TAPAS versions to enable cross-version analyses. Versions 12, 13, and 14 are structured such that the core 13 facets are administered separately from the experimental mix of facets, thus making the three versions identical for analysis purposes.

Table 2.2*TAPAS Facet Names and Definitions (TAPAS v12-v14)*

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.
Attention Seeking	High scoring individuals tend to engage in behaviors that attract social attention. They are loud, loquacious, entertaining, and even boastful.
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.
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.
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.

The number and composition of TAPAS composites used by the Army have evolved over the years. In all cases, composites have been constructed to predict specific outcomes of interest. While detailed information about the composites is confidential, the changes can be roughly summarized as follows:

- Unit-weighted Can-Do and Will-Do composites were created in 2009 (Knapp & Heffner, 2010) for the early original TAPAS versions.
- Three regression-weighted composites (Can-Do, Will-Do, and Adaptation) were introduced in 2011 (Nye et al., 2012) with the second generation of TAPAS versions (v4-v8). Adaptation was designed to predict early-term attrition.

- Composition (regression weights and/or constituent facets) of the three composites was updated in 2013 (Nye et al. 2014) with the third generation of TAPAS versions (v9-v11).
- Composition of the three composites was updated in 2019 with the latest generation of TAPAS versions (v12-v14), and an additional composite (Conduct) was added to predict attrition due to misbehavior.

Up until the latest set of revisions, it was generally possible to apply new composite formulae to preceding TAPAS versions for most of the composites. However, none of the TAPAS versions prior to the latest generation measure all the facets needed to compute the three reformulated composite scores. Accordingly, until there are sufficient cases for these composite scores that can be computed using the new formulations, our TAPAS-related analyses will need to treat the old and new composite formulations as comparable. Note also that the Conduct composite score can only be calculated for those who have taken TAPAS v8 and v12-v14, so sample sizes for analyses using this score are relatively small.

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 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. 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, Wolters et al. 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, 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 estimated 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, VAST 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 2.3 – 2.5), which are based on tests of job knowledge, self-reported attitudes, performance ratings provided by peers and supervisors, and relevant administrative indicators.

Table 2.3*Initial Military Training (IMT) TAPAS Validation Criterion Measures*

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. Items are averaged to form the composite.
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.

Note. ALQ = Army Life Questionnaire; PRS-P = Performance Rating Scale – Peer; WTBD = Warrior Tasks and Battle Drills.

Table 2.4*In-Unit TAPAS Validation Criterion Measures*

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.
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. Items are averaged to form the composite.
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.

Note. ALQ = Army Life Questionnaire; PRS-P = Performance Rating Scale – Peer; WTBD = Warrior Tasks and Battle Drills.

Table 2.5*Attrition Criterion Measures*

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

Technical Job Knowledge

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. 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 et al., 2009).

On average, the JKTs include roughly 30 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 for the MOS-specific JKTs and across MOS for the WTBD JKT. Scores are standardized within each biannual reporting cycle.

As estimated using the Cycle 21 evaluation database, the MOS-specific JKT reliability estimates are reasonable (mean $\alpha = .81$; Marshall et al., 2023). The WTBD reliability estimates were similar (IMT $\alpha = .81$, in-unit $\alpha = .79$).

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 here. Self-reported administrative information is discussed in a later section.

The ALQ was originally developed under the Select21 project (Knapp et al., 2005), though it had its origins in Project A (Campbell & Knapp, 2001). Earlier forms of the ALQ (Van Iddekinge et al., 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 et al., 2020). There are separate IMT and in-unit versions of the ALQ, though the two versions are highly similar.

Table 2.6 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. Validation analyses reported in Chapter 4 use the three composite scores shown at the top of Table 2.6, as well as the Counterproductive Soldier Behavior score. See Nesbitt et al. (2020) for more detailed information about ALQ content and derivation of the three composite scores.⁶

Table 2.6

Army Life Questionnaire Scale and Composite Score Reliability Estimate

Variable	IMT			In-Unit		
	Alpha	Number Scales/Items	<i>n</i>	Alpha	Number Scales/Items	<i>n</i>
<i>Composites</i>						
Commitment and Fit ^a	0.81	3	121,893	0.83	3	17,558
Peer Leadership ^a	0.81	4	53,019	0.81	4	4,980
Retention Cognitions	0.86	2	121,448	0.87	2	4,978
<i>Likert Scales</i>						
Affective Commitment	0.79	3	121,893	0.80	3	17,558
Army Fit	0.77	4	121,893	0.80	4	17,558
Army Life Adjustment	0.78	4	121,893	-	-	-
Counterproductive Soldier Behavior	0.83	9	53,020	0.77	9	4,980
Deployment Satisfaction	-	-	-	0.80	4	1,765
MOS Fit	0.86	5	121,893	0.90	5	17,558
MOS Satisfaction	-	-	-	0.84	5	17,558
MTL - Affective	0.83	3	53,020	0.83	3	4,980
MTL - Noncalculative	0.77	3	53,020	0.77	3	4,980
MTL – Social-Normative	0.77	3	53,019	0.76	3	4,980
Organizational Citizenship Behavior/Leadership	0.86	8	53,020	0.86	8	4,980
Resilience	0.80	7	53,020	0.78	7	4,980
Work/Life Balance	0.87	5	52,971	0.86	5	4,980

Note. MTL = Motivation to Lead. Reliability estimates were computed using the Cycle 20 database. All other estimates computed using the Cycle 21 database.

^a Alphas reflect the average of the composite scale alphas.

⁶ Two prior biannual evaluations (Knapp & Kirkendall, 2020, 2021) failed to reverse code the Motivation to Lead – Noncalculative score, which is a component of the Peer Leadership composite. These variables are correctly scored in the present report.

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.

Soldiers do not rate peers for whom they indicate insufficient familiarity (rating = 1).

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 2.7. The rating scale anchors are all on a 1-5 scale but vary by dimension. Figure 2.1 presents two examples.

Table 2.7

IMT Peer Performance Rating Scale Dimensions and Associated Example Behaviors

Scale	Example Behaviors
Effort and Discipline	• Puts forth individual effort in study, practice, and other training activities • Shows respect in word and action towards superiors • Obeys instructions, rules, and regulations • Maintains self, uniforms, living areas, and barracks to Army standards
Working Effectively with other Soldiers	• Respects, assists, and cooperates with fellow Soldiers • Treats all Soldiers with courtesy and respect, regardless of gender, race, ethnicity, ability, or background differences • Communicates ideas and information clearly and directly
Physical Fitness	• Maintains and enhances physical readiness • Performs required physical tasks

(Continued)

Table 2.7. (continued)

Scale	Example Behaviors
MOS Qualification Knowledge and Skill	• Learns and demonstrates AIT/OSUT knowledge and skills required for MOS qualification • Effectively applies training, rules, and strategies to solve problems
Resilience and Adjustment	• Persists in carrying out difficult tasks, even in stressful circumstances • Remains calm during high-pressure situations • Bounces back after facing a difficult situation or failure • Adjusts well to life in the Army
Counterproductive Soldier Behaviors	• Takes shortcuts that may be harmful • Takes unauthorized breaks • Leaves a mess for someone else to clean up
Going Above and Beyond	• Notices when a fellow Soldier is falling behind and offers assistance • Offers to help fellow Soldiers struggling to learn training material and skills • Goes out of the way to give another Soldier encouragement or express appreciation • Spends free time learning about procedures, equipment, etc.
Overall Performance	Instructions: Considering your evaluation of your peers on the dimensions important to being a successful Soldier, please rate the overall performance of each Soldier.

Figure 2.1

Sample IMT Peer Rater Performance Rating Scales

A. Effort and Discipline				
Example behaviors: • Puts forth individual effort in study, practice, and other training activities • Shows respect in word and action towards superiors • Obeys instructions, rules, and regulations • Maintains self, uniforms, living areas, and barracks to Army standards				
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: • Takes shortcuts that may be harmful • Takes unauthorized breaks • Leaves a mess for someone else to clean up				
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

In-Unit PRS-P

The in-unit PRS-P rating dimensions and example behaviors are presented in Table 2.8. 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 2.8

In-Unit Peer Performance Rating Scale Dimensions and Associated Example Behaviors

Scale	Example Behaviors
Effort and Discipline	• Puts forth individual effort in completing tasks and handling responsibilities • Shows respect in word and action towards superiors • Obeys instructions, rules, and regulations • Maintains self, uniforms, and work space to Army standards
Working Effectively with other Soldiers	• Respects, assists, and cooperates with fellow Soldiers • Treats all Soldiers with courtesy and respect, regardless of gender, race, ethnicity, ability, or background differences • Communicates ideas and information clearly and directly
Physical Fitness	• Maintains and enhances physical readiness • Performs required physical tasks
MOS Qualification Knowledge and Skill	• Demonstrates knowledge and skills required for MOS • Handles most MOS task-related problems effectively • Effectively applies training, previous experience, rules, and strategies to complete job tasks
Processing Information and Solving Problems	• Efficiently handles day-to-day information load • Develops original and creative approaches to dealing with problems • Can integrate information from multiple sources • Considers relevant costs and benefits of alternative solutions
Resilience and Adjustment	• Persists in carrying out difficult tasks, even in stressful circumstances • Remains calm during high-pressure situations • Bounces back after facing a difficult situation or failure • Adjusts well to life in the Army
Counterproductive Soldier Behaviors	• Takes shortcuts that may be harmful • Takes unauthorized breaks • Leaves a mess for someone else to clean up • Speaks poorly about the Army to others
Going Above and Beyond	• Volunteers for additional duty • Goes out of the way to give another Soldier encouragement or express appreciation • Demonstrates concern about the image of the unit
Emergent Leadership	• Notices when a fellow Soldier is falling behind • Keeps the unit focused on goals and objectives even when supervisors are not present • Gains the cooperation and trust of other Soldiers in the unit
<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.

PRS-P Interrater Reliability Estimates

Table 2.9 presents interrater reliability (IRR) estimates for both the IMT and in-unit PRS-P based on Cycle 21 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 2.9

Peer Performance Rating Scale Interrater Reliability Estimates

Measure	$G(q,k)$	k	n_{Ratees}	n_{Raters}
<i>IMT</i>				
Effort and Discipline	0.41	2.24	46,291	41,834
Working Effectively with Other Soldiers	0.37	2.25	46,303	41,835
Physical Fitness	0.51	2.24	46,281	41,802
MOS Qualification Knowledge & Skill	0.38	2.24	46,303	41,789
Resilience and Adjustment	0.37	2.24	46,264	41,784
Counterproductive Soldier Behaviors	0.17	2.24	46,224	41,670
Going Above and Beyond	0.32	2.24	46,269	41,803
Overall Performance Rating	0.46	2.25	46,242	41,680
<i>In-Unit</i>				
Effort and Discipline	0.29	1.74	3,113	2,832
Working Effectively with Other Soldiers	0.19	1.74	3,115	2,840
Physical Fitness	0.39	1.74	3,121	2,838
MOS Qualification Knowledge & Skill	0.28	1.73	3,112	2,829
Processing Information & Solving Problems	0.24	1.74	3,113	2,834
Resilience and Adjustment	0.22	1.74	3,113	2,832
Counterproductive Soldier Behaviors	0.10	1.73	3,113	2,819
Going Above and Beyond	0.19	1.74	3,117	2,837
Emergent Leadership	0.28	1.73	3,111	2,834
Overall Performance Rating	0.31	1.74	3,110	2,834

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 estimates were computed using the Cycle 21 database.

PRS-P Scoring

Validation analyses use a single Peer Performance Rating composite score. Specifically, we average scores across the scales (including Overall Performance but 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 2.10 lists the dimensions rated on the PRS-S.

Table 2.10

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.

Figure 2.2

Sample In-unit Supervisor Rater Performance Rating Scale

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
- Is not able to perform most Core Warrior Tasks. - Requires constant supervision.		- Performs most Core Warrior Tasks competently. - Requires some supervision under difficult conditions.			- Performs almost all Core Warrior Tasks extremely effectively. - Requires little or no supervision, even under difficult conditions.	

The vast majority of Soldiers in units are rated by only one supervisor, so we cannot compute interrater reliability estimates.

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 reenlistment or completion of service are not considered as having separated from the Army for the purposes of our analyses⁷. Separation data are reported for Regular Army Soldiers only because reliable data for the Reserve and National Guard components are not available.

⁷ In previous cycles, Soldiers with SPD codes indicating completion of service were excluded from our analyses. Beginning with Cycle 24, our construction of attrition variables has been revised such that Soldiers with SPD codes indicating completion of service are now included in our analyses as not having separated from the Army.

We create separate monthly attrition status variables through 36 months within a Soldier's time in service. A Soldier's time in service is calculated as the difference between the date from which their TAPDB record was extracted, they completed service, or they separated (whichever comes first), and their "ship" date as indicated in TAPDB. For each monthly attrition status variable, a Soldier will be indicated as:

- 0 (in service) if they had that many months in service without separating;
- 1 (separated) if they had separated by that many months in service; or
- Missing/No data (cannot be determined) if attrition data was missing for the Soldier.

Missingness on the attrition status variables may occur for the following reasons:

- Missing data on ship date.
- Insufficient time in service to indicate a given monthly attrition status variable. For example, if a Soldier has a total of 12 months in service and has not separated, they will be indicated as 0 (in service) for each month from first through 12th and indicated as missing data (cannot be determined) for each month from 13th through 36th.
- Censored data due to completion of active service. For example, if a Soldier has a 2-year term of service and has not separated for other reasons after completing their term of service, they will be indicated as 0 (in service) for each month from first through 24th month and indicated as missing data (cannot be determined) for each month from 25th through 36th.

Once the monthly attrition status variables are constructed, a Soldier's 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 fitness test scores. 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.

Soldiers retake fitness tests at approximate 6-month intervals. The Army Physical Fitness Test (APFT) has recently been replaced by the Army Combat Fitness Test (ACFT). Both result in a single continuous score which reflects performance on multiple fitness exercises. Scores are adjusted for gender and age. Given the recent introduction of ACFT, only APFT scores are included in the present report.

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.⁸ 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 to manage the data collection, while also providing Soldiers with the correct spelling of the rated peers' names.

In-Unit Data Collections

Most Army installations host an annual “umbrella week” which is a time period set aside to support research data collection activities. We schedule data collection visits during these umbrella weeks, resulting in rolling data collections throughout the year at various locations. Soldiers and their supervisors are instructed to report to a computer classroom at which time data collection staff (ARI and/or HumRRO 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 whom 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 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.

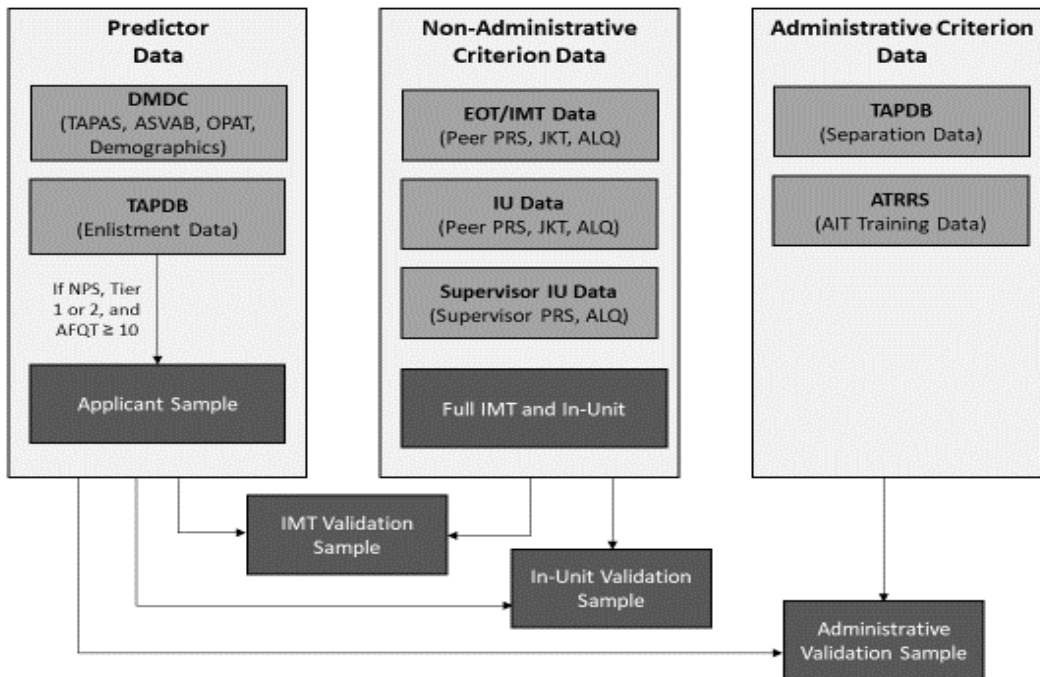
An illustrative view of our data sources is provided in Figure 2.3 (see Appendix A for acronym definitions). The analysis data files are built on a foundational database drawn from the TAPDB onto which predictor and criterion data from other sources are added. The lighter boxes within the figure represent 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

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

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.

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 24 SAMPLE CHARACTERISTICS AND DESCRIPTIVE STATISTICS

Timothy Burgoyne and Tetsuhiro Yamada (HumRRO)

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 we filtered Soldier data for analysis, then describe multiple subsamples created to support various types of analyses. Finally, we 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 such as always picking the first statement, large amounts of missing data), and
- who were administered an Army TAPAS version (i.e., excludes Air Force, Marine Corps, and Navy versions) other than the 13D-CAT⁹.

From the sample of TAPAS data, these filters excluded 17,274 Soldiers in Education Tier 3, 42,253 Soldiers in AFQT Category V, 1,654 Soldiers who completed TAPAS version 13D, 234,606 Soldiers who completed non-Army TAPAS versions, and 1,205,073 Soldiers with invalid TAPAS data. There were 1,436,067 Soldiers who met the screening criteria and were included in the Applicant Sample.

The Administrative Validation Sample included 878,846 Soldiers who met all the inclusion criteria for the Applicant Sample and also had a signed contract and at least one record from an administrative criterion data source (e.g., ATRRS). There were 106,797 Soldiers with IMT criterion data linked to an administrative TAPAS record and included in the IMT Validation Sample. There were 9,982 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

⁹ 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).

date of birth to match TAPAS records to the IMT and in-unit criterion data, which results in a nontrivial percentage of unsuccessful matches.

Table 3.1

Background and Demographic Characteristics of the Analysis Samples

Characteristic	Applicant ^a		Administrative Validation ^b		IMT Validation ^c		In-Unit Validation ^d	
	<i>n</i> = 1,436,067		<i>n</i> = 878,846		<i>n</i> = 106,797		<i>n</i> = 9,982	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>Component</i>								
Regular	819,896	57.1	516,699	58.8	63,811	59.7	9,693	97.1
ARNG	427,149	29.7	247,541	28.2	30,844	28.9	91	0.9
USAR	189,022	13.2	114,606	13.0	12,142	11.4	198	2
Tier 1		95.1	841,263	95.7	102,619	96.1	9,660	96.8
Tier 2	70,754	4.9	37,583	4.3	4,178	3.9	322	3.2
Infantryman (11B/C/X + 18X)	148,637	10.4	137,773	15.7	19,704	18.4	2,290	22.9
Combat Engineer (12B)	28,060	2.0	25,122	2.9	6,226	5.8	373	3.7
Cannon Crewmember (13B)	18,528	1.3	16,940	1.9	410	0.4	204	2.0
Field Artillery Tactical Data Systems Specialist (13D)	4,119	0.3	4,008	0.5	974	0.9	66	0.7
Fire Support Specialist (13F)	10,510	0.7	9,829	1.1	1,158	1.1	83	0.8
Fire Control Specialist (13J)	4,348	0.3	3,554	0.4	204	0.2	31	0.3
Cavalry Scout (19D)	24,807	1.7	22,907	2.6	2,404	2.3	488	4.9
M1 Armor Crewmember (19K)	10,851	0.8	10,358	1.2	2,499	2.3	259	2.6
Military Police (31B)	39,155	2.7	35,697	4.1	12,424	11.6	166	1.7
Human Resources Specialist (42A)	24,451	1.7	22,369	2.5	4,727	4.4	245	2.5
Combat Medic Specialist (68W)	42,301	2.9	38,527	4.4	8,903	8.3	372	3.7
Motor Transport Operator Soldiers (88M)	44,998	3.1	40,258	4.6	11,964	11.2	692	6.9
Wheeled Vehicle Mechanic (91B)	47,816	3.3	43,051	4.9	3,061	2.9	549	5.5
Other	502,133	35.0	457,870	52.1	21,757	20.4	4,110	41.2
Unknown	485,353	33.8	10,583	1.2	10,382	9.7	54	0.5
<i>AFQT Category^f</i>								
I	83,876	5.8	53,124	6.1	6,545	6.1	432	4.3
II	414,888	28.9	277,945	31.6	36,441	34.1	2,609	26.1
IIIA	292,481	20.4	195,454	22.2	24,584	23	2,253	22.6
IIIB	440,326	30.7	297,094	33.8	32,814	30.7	4,000	40.1
IV	204,496	14.2	58,967	6.3	6,413	6	688	6.9

(continued)

Table 3.1 (continued)

Characteristic	Applicant ^a		Administrative Validation ^b		IMT Validation ^c		In-Unit Validation ^d	
	<i>n</i> = 1,436,067		<i>n</i> = 878,846		<i>n</i> = 106,797		<i>n</i> = 9,982	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>Contract Status</i>								
Signed	984,943	68.6						
Not signed	451,124	31.4	--	--	--	--	--	--
<i>Gender</i>								
Female	337,645	23.5	177,410	20.2	18,183	17	1,244	12.5
Male	1,098,422	76.5	701,436	79.8	88,614	83	8,738	87.5
<i>Race</i>								
African American	338,382	23.6	190,887	21.7	19,445	18.2	2,380	23.8
American Indian	11,665	0.8	7,001	0.8	834	0.8	76	0.8
Asian	69,440	4.8	42,027	4.8	4,693	4.4	573	5.7
Hawaiian/Pacific Islander	2,956	0.2	1,414	0.2	152	0.1	25	0.3
Caucasian	995,643	69.3	630,181	71.7	80,694	75.6	6,839	68.5
Multiple	4,271	0.3	2,420	0.3	247	0.2	32	0.3
Declined to Answer/Missing	13,710	1	4,916	0.6	732	0.7	57	0.6
<i>Ethnicity</i>								
Hispanic/Latino	248,377	17.3	148,504	16.9	18,793	17.6	1,896	19
Not Hispanic	1,170,905	81.5	723,178	82.3	87,356	81.8	8,020	80.3
Declined to Answer/Missing	16,785	1.2	7,164	0.8	648	0.6	66	0.7

^aNon-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

^bSoldiers in Applicant Sample with a signed contract and at least one criterion from an administrative record.

^cSoldiers in Applicant Sample with a signed contract and criterion data collected at schoolhouses.

^dSoldiers in Applicant Sample with a signed contract and criterion data collected in units.

^eThe 13D MOS was combined with 13P to create the 13J MOS in early 2018.

^fAFQT 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 four TAPAS composite scores by TAPAS version.

Table 3.2

Means and Standard Deviations for the TAPAS Composites by TAPAS Version

TAPAS Composite	15D-Static/CAT v4/5/7/8		13D-CAT v9		13D-CAT v10		13D-CAT v11		13D-CAT v12-v14	
	(n = 100,976 - 422,160) ^a		(n = 268,899)		(n = 268,596)		(n = 271,026)		(n = 205,377 - 205,386)	
	M	SD	M	SD	M	SD	M	SD	M	SD
Can-Do	99.62	21.05	99.88	19.63	99.25	19.63	100.01	19.58	98.51	19.65
Will-Do	99.81	19.97	98.83	20.18	98.55	19.57	99.13	20.78	98.92	19.64
Adaptation	100.22	19.84	99.57	19.89	99.48	19.79	100.02	21.07	99.57	19.61
Conduct	100.77	20.97	-	-	-	-	-	-	98.19	19.87

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).

^aCan-Do and Adaptation composites were not calculated for Soldiers taking TAPAS v7 and v8, resulting in reduced sample sizes.

Table 3.3 presents descriptive statistics and correlations among the predictor variables in the Applicant Sample. Consistent with its characterization as a predictor of technical performance, the Can-Do composite showed a moderate correlation with AFQT. Can-Do correlated moderately with Adaptation as well, likely due to the shared TAPAS facets between the two. The Will-Do and Adaptation composites correlated at $r = .56$, which is remarkable considering there is only one facet shared between the two composites. The Adaptation and Conduct composites also correlated moderately, which may be a reflection of the shared TAPAS component facets between the two.

Table 3.3

Descriptive Statistics and Intercorrelations Among AFQT and TAPAS Composites in the Applicant Sample

Variable	n	M	SD	1	2	3	4
1. AFQT	1,436,067	54.89	22.58	-			
2. Can-Do	1,234,101	99.50	19.89	.38	-		
3. Will-Do	1,436,067	99.14	20.05	.10	.13	-	
4. Adaptation	1,335,077	99.80	20.06	.15	.35	.56	-
5. Conduct	306,353	99.04	20.27	.07	.26	-.09	.38

Note. All correlations are statistically significant at $p < 0.05$.

Descriptive Statistics for Criterion Measures

IMT Criteria

Descriptive statistics for IMT criteria are reported in Table 3.4. 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.¹⁰ 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. For all IMT criteria, all means leaned toward the desirable direction (e.g., greater than 3.0 for 5-point rating scales, less than 3.0 for Counterproductive Soldier Behavior, and less than 1.0 for negative-valenced dichotomous criteria).

Table 3.4

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	47,484	0.04	0.85	-4.00	2.40
WTBD JKT	80,642	-0.02	1.00	-4.49	3.17
Peer Performance Rating (PRS-P) Composite	43,653	3.70	0.77	1.00	5.00
IMT Restarts (Y/N)	707,101	0.06	0.23	0.00	1.00
<i>Army Life Questionnaire (ALQ)</i>					
Commitment & Fit	96,551	3.76	0.65	1.00	5.00
Retention Cognitions	96,143	3.58	1.08	1.00	5.00
Peer Leadership	54,600	3.66	0.55	1.00	5.00
Counterproductive Soldier Behavior	54,602	1.94	0.63	1.00	5.00
Army Physical Fitness Test (APFT)	89,284	250.79	28.54	180.00	300.00
Disciplinary Incidents (Y/N)	103,093	0.23	0.42	0.00	1.00

Note. Ratings on PRS-P range from 1 to 5.

In-Unit Criteria

The in-unit criteria presented in Table 3.5 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 individual ratings concerning a Soldier's performance. Like the IMT criteria, all variables had favorable means except for Retention Cognition, which had a mean slightly below the midpoint of the scale, and WTBD JKT, which had a mean slightly below zero. Note that Counterproductive Work Behavior is negatively valenced; as such, its mean of 2.01 is favorable.

¹⁰ For IMT Soldiers who do not have an MOS JKT score, AIT grade is substituted when available.

Table 3.5*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 (PRS-S) Composite	6,367	4.85	1.14	1.00	7.00
Peer Performance Rating (PRS-P) Composite	2,743	3.76	0.76	1.00	5.00
WTBD JKT	6,699	-0.06	1.00	-4.45	2.96
<i>Army Life Questionnaire (ALQ)</i>					
Commitment & Fit Composite	9,518	3.31	0.78	1.00	5.00
Retention Cognitions Composite	4,382	2.93	1.32	1.00	5.00
Peer Leadership Composite	4,384	3.60	0.60	1.00	5.00
Counterproductive Work Behavior	4,384	2.01	0.60	1.00	5.00
Army Physical Fitness Test (APFT) Score	8,974	251.43	28.82	180.00	300.00
Disciplinary Incidents (Y/N)	9,964	0.26	0.44	0.00	1.00

Note. Ratings on PRS-P range from 1 to 5. Ratings on PRS-S range from 1 to 7.

Attrition

Table 3.6 summarizes base rates for the dichotomous criteria. The 6- and 12-month attrition rates are comparable to those of previous evaluation cycles. The rates for 24- and 36-month attrition, however, are lower. In particular, the 36-month attrition rate is approximately 5% lower compared to what we observed in Cycle 22. This is due in part to the change in the construction of these variables, as Soldiers who completed their term of service are now included in the sample as not having separated from the Army (they had been excluded from attrition analyses in prior cycles). The IMT Restart rate and Disciplinary Incident rates are comparable to the most recently reported values in Cycles 22 and 21, respectively.

Table 3.6*Base Rates for Dichotomous Criteria in the Administrative Validation Sample*

<i>Criterion</i>	<i>n</i>	<i>n_{Event}</i>	<i>%_{Event}</i>
<i>Cumulative Attrition</i>			
6-Month	488,737	57,130	11.7
12-Month	483,071	69,770	14.4
24-Month	434,347	92,063	21.2
36-Month	376,108	106,337	28.3
IMT Restarts	707,101	38,981	5.51
Disciplinary Incidents (IMT)	103,093	23,530	22.8
Disciplinary Incidents (In-unit)	9,964	2,546	25.6

Note. *n* = number of Soldiers with criterion data at the time data were extracted. For Cumulative Attrition, only Regular Army Soldiers are included. *n_{Event}* = number of Soldiers who experienced the focal event (i.e., attrition, training restart, disciplinary incident) at the specified months of service. *%_{Event}* = percentage of Soldiers who experienced the focal event through the specified months of service [$(n_{Event}/n) \times 100$].

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 1,436,067 applicants who took TAPAS and the Validation Sample including 879,855 incumbents with at least one type of criterion data. Analyses on the predictor variables revealed consistent means and standard deviations of TAPAS composite scores across forms and composites, as well as expected degrees of intercorrelation among the predictor variables. Descriptive statistics of criterion variables were also consistent with those of past cycles.

CHAPTER 4: EVIDENCE FOR THE PREDICTIVE VALIDITY OF THE TAPAS¹¹

Tetsuhiro Yamada and Nicholas Howald (HumRRO)

This chapter describes criterion-related validation of TAPAS scores to predict Soldiers' performance and retention through their first enlistment term. For these analyses, we focus on the four TAPAS composites described in Chapter 2 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 et al., 2001; Campbell et al., 1990; Knapp & Tremble, 2007; Strickland, 2005). We used regression analysis to examine the incremental validity of TAPAS to predict outcomes beyond AFQT. We conducted these analyses on the Administrative Validation Sample described in Chapter 3 (i.e., subset of the Applicant Sample that also have a signed contract and at least one record from an administrative criterion data source).

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. Refer to Tables 2.2 through 2.4 in Chapter 2 which summarize the IMT and in-unit criteria, including a brief description of each.

While the four TAPAS composite scores were each designed to predict different outcomes (e.g., Adaptation is intended to predict attrition), we examine results for the full set of criterion measures for all three TAPAS composite scores.

Bivariate Correlations

Table 4.1 presents uncorrected zero-order correlations between the predictor variables (i.e., AFQT and TAPAS composite scores) and continuous criteria (both IMT and in-unit). As expected, AFQT correlated moderately with cognitively loaded criteria (Knowledge & Skill, WTBD JKT for both IMT and in-unit). The Can-Do composite also yielded correlations close to or above .20 with cognitively loaded criteria. The Will-Do composite yielded correlations larger than .20 with APFT (IMT and in-unit) and Peer Leadership (IMT and in-unit). The Adaptation and Conduct composites generally showed weaker correlations with continuous criteria compared to the other predictors, although Conduct had a correlation of $r = .23$ with PRS-P: Overall Performance Composite (in-unit).

¹¹ 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).

Table 4.1*Uncorrected Zero-order Correlations Between Predictors and Continuous Criteria*

Criterion Measure	Predictor				
	AFQT	Can-Do	Will-Do	Adaptation	Conduct
<i>IMT</i>					
<i>n</i>	43,653 – 707,101	41,202 – 578,302	43,653 – 707,101	43,533 – 642,763	1,050 – 66,602
Knowledge & Skill	.43	.21	.03	.09	.07
WTBD JKT (Standardized)	.40	.19	.03	.08	.06
APFT Score	.08	.01	.26	.17	-.03
Counterproductive Soldier Behavior	.01	-.01	-.06	.00	-.13
Commitment and Fit	-.04	.02	.12	.03	.02
Retention Cognitions	-.11	-.02	.04	-.03	.01
Peer Leadership	.10	.10	.26	.02	-.04
PRS-P: Overall Performance Composite	.08	.03	.13	.11	.07
IMT Restarts	.00	.00	-.03	-.03	.00
Disciplinary Incidents	.00	-.01	-.09	-.06	-.05
<i>In-Unit</i>					
<i>n</i>	2,774 – 9,964	2,626 – 7,699	2,774 – 9,964	2,703 – 8,884	119 – 1,289
WTBD JKT (Standardized)	.43	.22	.02	.11	-.03
APFT Score	-.01	-.01	.22	.14	-.09
Counterproductive Soldier Behavior	.08	-.02	-.03	.02	-.07
Commitment and Fit	-.03	.00	.04	.01	.07
Retention Cognitions	-.09	-.02	.00	-.02	.01
Peer Leadership	.08	.11	.22	.03	.01
PRS-P: Overall Performance Composite	.04	.00	.10	.06	.23
PRS-S: Overall Performance Composite	.09	.05	.09	.05	.03
Disciplinary Incidents	-.02	-.01	-.05	-.05	-.07

Note. Correlation coefficients in bold are statistically significant at $p < .05$.

Table 4.2 presents comparable results for dichotomous outcomes. Despite its design, the Adaptation composite did not relate much more strongly with attrition criteria than AFQT and the Will-Do composite. The Conduct composite showed non-zero but weak correlations with Disciplinary Incidents, reflecting its intended characterization as a predictor for attrition due to misbehavior.

Table 4.2

Uncorrected Zero-order Correlations Between Predictors and Dichotomous Criteria

Criterion Measure	Predictor				
	AFQT	Can-Do	Will-Do	Adaptation	Conduct
<i>n</i>	9,964 – 707,101	7,699 – 578,302	9,964 – 707,101	8,884 – 642,763	1,289 – 77,387
6-Month Attrition	-.05	-.01	-.08	-.08	.00
12-Month Attrition	-.05	-.01	-.08	-.08	.00
24-Month Attrition	-.07	-.01	-.08	-.09	-.02
36-Month Attrition	-.09	-.02	-.07	-.10	-.03
IMT Restarts	.00	.00	-.03	-.03	.00
Disciplinary Incidents (IMT)	.00	-.01	-.09	-.06	-.05
Disciplinary Incidents (In-Unit)	-.02	-.01	-.05	-.05	-.07

Note. Correlation coefficients in bold are statistically significant at $p < .05$.

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 scores. This approach is generally consistent with previous evaluations of TAPAS and similar experimental non-cognitive predictors (e.g., Hughes, 2020; Ingerick et al., 2009; Knapp & Heffner, 2009; 2010; Trippe et al., 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. We present the results of these primarily in tabular form. For models that yielded $\Delta R > .05$, we also present results in graphic form. Although sample sizes for all IMT analyses were quite large, those for models involving the Can-Do and Adaptation composites were smaller relative to models involving the Will-Do composite because TAPAS v7 and v8 did not include all the facets required to compute these scores using the same formula. Sample sizes are even smaller for analyses involving the Conduct composite, since this composite was introduced in the latest generation of TAPAS versions (v12-v14). Furthermore, sample sizes for the in-unit analyses were much smaller than for the IMT analyses, especially if they involved the Conduct composite. The smallest sample size was for the Conduct – PRS-P: Overall Performance Composite model ($n = 119$). Caution must be exercised when interpreting the results of such models.

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. We provide additional details concerning the specific indices presented for the logistic regression analyses with the results of the dichotomous outcomes later in the chapter. We examined attrition criteria for Regular Army Soldiers only.

IMT Incremental Validity Analyses

Table 4.3 and Figure 4.1 summarize incremental validity results of the TAPAS composites for predicting IMT performance criteria over and above AFQT. Overall, the results suggest that the Will-Do and Adaptation TAPAS composites enhance prediction of various IMT criteria compared to the Can-Do and Conduct composites.

The TAPAS composites evidenced no notable increments over AFQT in predicting cognitive criteria (i.e., Knowledge and Skill and WTBD JKT; $\Delta R_s = .00$). These results are not surprising given that AFQT is an established strong predictor of Can-Do outcomes (Thacker et al., 2020).

For motivation-oriented performance criteria, the Will-Do composite enhanced prediction of the APFT score ($\Delta R = .19$) and Peer Leadership composite ($\Delta R = .17$). To a lesser extent, it also provided substantial incremental validity for Counterproductive Soldier Behavior ($\Delta R = .06$), Commitment and Fit ($\Delta R = .09$), and the Peer Performance Rating composite ($\Delta R = .07$). The relatively small incremental gain in predicting the performance ratings score may be due to the low interrater reliability of the measure.

The Adaptation composite demonstrated a substantial degree of incremental validity in the prediction of APFT score ($\Delta R = .11$) and to a lesser extent, the Peer Performance Rating composite ($\Delta R = .05$). The Can-Do composite is not designed to predict Will-Do criteria, and as such it did not demonstrate any meaningful incremental validity for the motivation-based outcomes.

Incremental validity of the Conduct composite over AFQT for IMT continuous criteria was generally trivial. However, it did provide substantial incremental validity in predicting Counterproductive Soldier Behavior ($\Delta R = .12$). This is consistent with its conceptualization as a correlate of misbehavior.

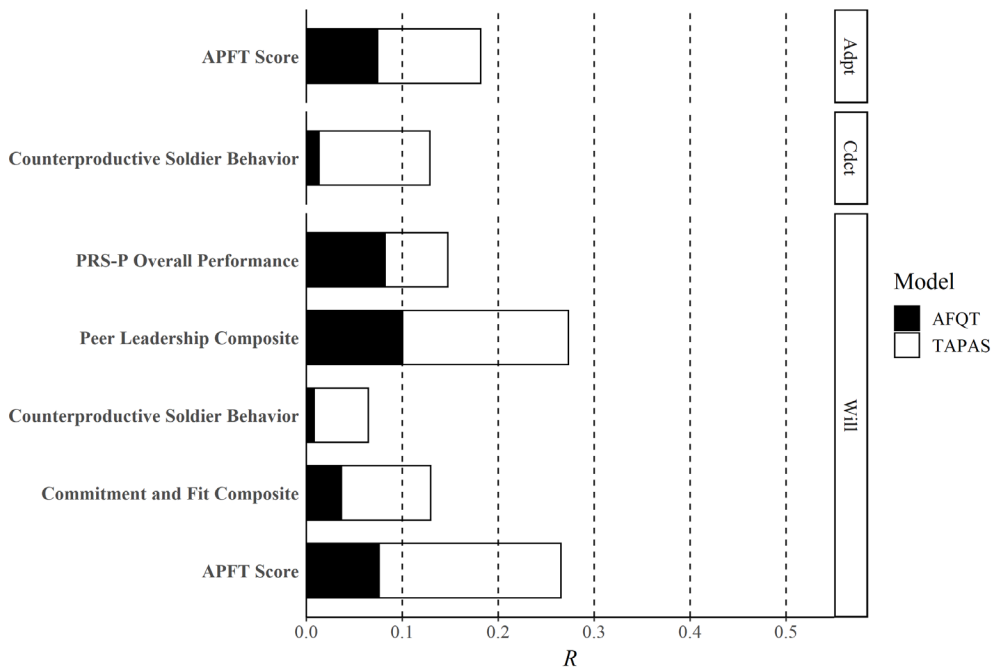
Table 4.3*Incremental Validity Estimates for TAPAS Composites over AFQT for Predicting IMT Criteria*

Criterion Measure/Model	<i>n</i>	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR
<i>Knowledge & Skill</i>				
Can-Do	41,202	.43	.43	.00
Will-Do	47,484	.43	.43	.00
Adaptation	44,307	.43	.43	.00
Conduct	3,829	.46	.47	.00
<i>WTBD JKT (Standardized)</i>				
Can-Do	72,193	.40	.40	.00
Will-Do	80,642	.40	.40	.00
Adaptation	76,381	.40	.40	.00
Conduct	6,090	.41	.41	.00
<i>APFT Score</i>				
Can-Do	80,454	.07	.08	.00
Will-Do	89,284	.08	.27	.19
Adaptation	84,881	.07	.18	.11
Conduct	5,002	.07	.08	.01
<i>Counterproductive Soldier Behavior</i>				
Can-Do	54,241	.01	.02	.01
Will-Do	54,602	.01	.06	.06
Adaptation	54,433	.01	.01	.00
Conduct	2,923	.01	.13	.12
<i>Commitment and Fit</i>				
Can-Do	87,400	.03	.04	.01
Will-Do	96,551	.04	.13	.09
Adaptation	91,974	.03	.05	.01
Conduct	7,415	.05	.06	.01
<i>Retention Cognitions</i>				
Can-Do	87,030	.11	.11	.00
Will-Do	96,143	.11	.12	.01
Adaptation	91,589	.11	.11	.00
Conduct	7,383	.11	.11	.00
<i>Peer Leadership</i>				
Can-Do	54,239	.10	.12	.02
Will-Do	54,600	.10	.27	.17
Adaptation	54,431	.10	.10	.00
Conduct	2,922	.08	.09	.01
<i>PRS-P Overall Performance Composite</i>				
Can-Do	43,390	.08	.08	.00
Will-Do	43,653	.08	.15	.07
Adaptation	43,533	.08	.13	.05
Conduct	1,050	.11	.13	.02

Note. PRS = Performance Rating Scales, where -P denotes peer ratings. *R* = multiple correlations between the AFQT and selected TAPAS composites with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS composite to the regression model [(AFQT + TAPAS) – AFQT-Only]. Bolded values indicate $p < .05$.

Figure 4.1

Increase in Prediction of IMT Continuous Criteria using TAPAS Composites ($\Delta R > .05$ only)



Note. Adpt = Adaptation, Cdct = Conduct, Will = Will-do. None of the models with the Can-do composite yielded $\Delta R > .05$.

Table 4.4 and Figure 4.2 present incremental validity results for predicting in-unit performance. Similar to the results for the IMT performance criteria, these results also suggest that the Will-Do and Adaptation composites are particularly useful predictors of in-unit performance outcomes.

As was the case with IMT, none of the TAPAS composites provided a meaningful degree of incremental validity over AFQT in predicting WTBD JKT scores.

For motivation-based criteria, the Will-Do composite provided incremental validity beyond AFQT in predicting APFT score ($\Delta R = .21$), Peer Leadership ($\Delta R = .15$), and to a lesser extent, the Peer Performance Rating composite ($\Delta R = .07$) scores.

The Adaptation composite demonstrated meaningful levels of incremental validity for predicting APFT score ($\Delta R = .14$).

The Conduct composite showed modest incremental validity over AFQT for APFT score ($\Delta R = .08$), Commitment and Fit ($\Delta R = .05$), and Peer Performance Rating composite ($\Delta R = .19$), although the result for the Peer Performance Rating composite should be interpreted with caution due to the small sample size.

Table 4.4

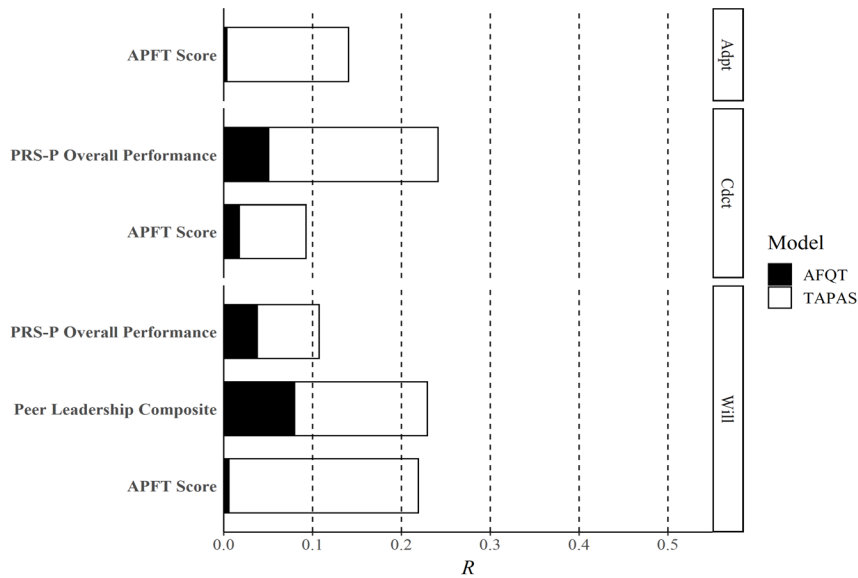
Incremental Validity Estimates for TAPAS Composites over AFQT for Predicting In-Unit Criteria

Criterion Measure/Model	<i>n</i>	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR
<i>WTBD JKT (Standardized)</i>				
Can-Do	4,891	.42	.43	.01
Will-Do	6,699	.43	.43	.00
Adaptation	5,831	.43	.43	.00
Conduct	1,012	.43	.43	.00
<i>APFT Score</i>				
Can-Do	6,876	.01	.01	.00
Will-Do	8,974	.01	.22	.21
Adaptation	7,971	.00	.14	.14
Conduct	1,154	.02	.09	.08
<i>Counterproductive Soldier Behavior</i>				
Can-Do	3,992	.07	.09	.01
Will-Do	4,384	.08	.08	.01
Adaptation	4,198	.08	.08	.00
Conduct	278	.18	.19	.01
<i>Commitment and Fit</i>				
Can-Do	7,338	.04	.04	.00
Will-Do	9,518	.03	.06	.02
Adaptation	8,477	.04	.04	.00
Conduct	1,222	.03	.08	.05
<i>Retention Cognitions</i>				
Can-Do	3,990	.09	.09	.00
Will-Do	4,382	.09	.09	.00
Adaptation	4,196	.09	.09	.00
Conduct	278	.08	.08	.00
<i>Peer Leadership</i>				
Can-Do	3,992	.09	.12	.03
Will-Do	4,384	.08	.23	.15
Adaptation	4,198	.08	.08	.00
Conduct	278	.00	.01	.01
<i>PRS-P: Overall Performance Composite</i>				
Can-Do	2,597	.03	.04	.00
Will-Do	2,743	.04	.11	.07
Adaptation	2,673	.04	.07	.03
Conduct	119	.05	.24	.19
<i>PRS-S: Overall Performance Composite</i>				
Can-Do	4,837	.08	.08	.00
Will-Do	6,367	.09	.13	.04
Adaptation	5,644	.09	.10	.01
Conduct	845	.13	.13	.00

Note. PRS = Performance Rating Scales, where -S denotes supervisor ratings and -P denotes peer ratings. *R* = multiple correlations between the AFQT and selected TAPAS composites with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS composite to the regression model [(AFQT + TAPAS) – AFQT-Only]. Bolded values indicate $p < .05$.

Figure 4.2

Increase in Prediction of In-unit Continuous Criteria using TAPAS Composites ($\Delta R > .05$ only)



Note. Adpt = Adaptation, Cdct = Conduct, Will = Will-do. None of the models with the Can-do composite yielded $\Delta R > .05$.

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 IMT and in-unit 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 χ^2 tests of the change in model deviance (i.e., negative two log likelihood; -2LL) between the AFQT-only and the AFQT + TAPAS composite models.

Odds ratios can be used to assess the likelihood (or odds) of the occurrence of a given outcome depending on change in a predictor. Specifically, for a given logistic regression model, a unique OR is estimated for each predictor and represents the amount of change in the odds of the occurrence 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 (i.e., lower likelihood of occurrence). Conversely, ORs above 1.0 indicate positive relationships (i.e., higher likelihood of occurrence). Although values of ORs cannot fall below 0.0, there is no upper limit for ORs (Cohen et al., 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 statistically 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 χ^2 tests to determine the statistical significance of change in model fit between models. In the present application, statistically significant likelihood ratio χ^2 tests of the change in deviances suggest that the inclusion of a given TAPAS composite to a regression model provides a 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 4.5. 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 IMT Disciplinary Incidents ($OR_{Will-Do} = .989$; $OR_{Adaptation} = .993$), in-unit Disciplinary Incidents ($OR_{Will-Do} = .994$; $OR_{Adaptation} = .995$), 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. The Can-Do composite enhanced the prediction of IMT Disciplinary Incidents ($OR_{Can-Do} = .999$) but not in-unit. Consistent with its characterization, the Conduct composite added incremental validity in predicting both IMT ($OR_{Conduct} = .994$) and in-unit ($OR_{Conduct} = .992$) Disciplinary Incidents.

Table 4.5

Incremental Validity Estimates for TAPAS Composites over AFQT for Predicting IMT Restarts and Disciplinary Incidents

Criterion Measure/Model	<i>n</i>	$OR_{AFQT}(CI)$	$OR_{TAPAS}(CI)$	r_{pb}	$\Delta-2LL$
<i>IMT Restarts</i>					
Can-Do	578,302				
AFQT		.975 (.964-.986)		.01	
AFQT+TAPAS		.978 (.966-.990)	1.000 (.999-1.000)	.01	2.20
Will-Do	707,101				
AFQT		.979 (.969-.989)		.00	
AFQT+TAPAS		.988 (.978-.998)	.994 (.993-.994)	.03	600.67
Adaptation	642,763				
AFQT		.978 (.967-.988)		.01	
AFQT+TAPAS		.994 (.983-1.005)	.994 (.994-.995)	.03	466.76
Conduct	66,602				
AFQT		1.005 (.971-1.039)		.00	
AFQT+TAPAS		1.005 (.972-1.040)	.999 (.998-1.001)	.00	0.96

(continued)

Table 4.5. (continued)

Criterion Measure/Model	<i>n</i>	OR _{AFQT} (CI)	OR _{TAPAS} (CI)	<i>r_{pb}</i>	Δ-2LL
<i>IMT Disciplinary Incidents</i>					
Can-Do	93,726				
AFQT		.992 (.977-1.008)		.00	
AFQT+TAPAS		.999 (.982-1.015)	.999 (.998-1.000)	.01	4.48
Will-Do	103,093				
AFQT		.990 (.975-1.005)		.00	
AFQT+TAPAS		1.004 (.990-1.019)	.989 (.989-.990)	.09	848.28
Adaptation	98,414				
AFQT		.991 (.976-1.006)		.00	
AFQT+TAPAS		1.011 (.996-1.027)	.993 (.992-.994)	.06	341.95
Conduct	7,749				
AFQT		.952 (.901-1.006)		.02	
AFQT+TAPAS		.956 (.904-1.010)	.994 (.992-.997)	.05	17.60
<i>In-Unit Disciplinary Incidents</i>					
Can-Do	7,699				
AFQT		.959 (.908-1.013)		.02	
AFQT+TAPAS		.960 (.906-1.018)	1.000 (.997-1.003)	.02	0.02
Will-Do	9,964				
AFQT		.945 (.902-.991)		.02	
AFQT+TAPAS		.947 (.903-.993)	.994 (.992-.996)	.06	28.83
Adaptation	8,884				
AFQT		.947 (.900-.996)		.02	
AFQT+TAPAS		.958 (.911-1.008)	.995 (.993-.997)	.05	16.45
Conduct	1,289				
AFQT		.893 (.784-1.016)		.05	
AFQT+TAPAS		.899 (.788-1.024)	.992 (.986-.998)	.09	6.84

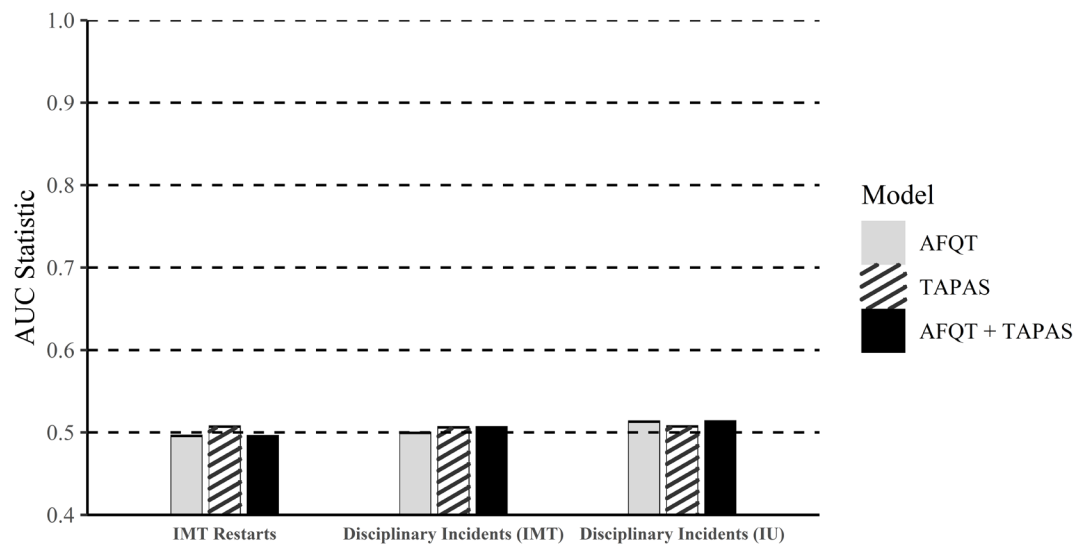
Note. OR = odds ratio for each predictor. CI = 95% confidence interval of the odds ratio. *r_{pb}* = point biserial correlation between the observed outcome and predicted probability. Δ-2LL = change in negative two log likelihood (deviance) from adding the selected TAPAS composite score to the AFQT-only 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 Δ-2LL, bolded values indicate significant change in model fit based on a Likelihood Ratio χ^2 test, $p < .05$.

Figures 4.3 to 4.6 display for each TAPAS composite logistic regression models in predicting Disciplinary Incidents and IMT Restarts. AUC statistics are commonly used to help evaluate the quality of binary prediction models and can be thought of as follows: If a pair of Soldiers, one who experienced the event of interest (e.g., Disciplinary Incident, IMT Restart) and another who did not experience the event of interest, were repeatedly sampled from the population of Soldiers, the AUC statistic reflects the proportion of times the Soldiers who experienced the event will score higher on the prediction model. As such, AUC values can range from 0 (i.e., Soldiers who experienced the event never receiving a higher score than those who did not experience the event) to 1.0 (i.e., Soldiers who experienced the event always scoring higher than those who did not experience the event). A value of .50 indicates that the prediction model is no better than a coin flip.¹² For each outcome, results are presented for three models which include the following predictors: (a) AFQT, (b) TAPAS composite, and (c) AFQT + TAPAS composite.

In interpreting the magnitude of AUC, Rice and Harris (2005) provided guidance on what AUC values could be considered the lower bound for small, medium, and large effects, similar to how Cohen's *d* values are interpreted. More specifically, in Table 1 of their manuscript, they present different values of Cohen's *d*, as well as AUC and point biserial correlation values equivalent to this Cohen's *d* value. For the common *d* values typically considered small (*d* = .20), medium (*d* = .50), and large (*d* = .80), the table presents AUC values of .556, .639, and .714, respectively.

Figure 4.3

AUC Statistics of AFQT and TAPAS Can-Do Composite in the Discrimination of both IMT and In-unit Disciplinary Incidents and IMT Restarts



¹² The ROC curve is a plot of the true positive rate against the false positive rate 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.

Figure 4.4

AUC Statistics of AFQT and TAPAS Will-Do Composite in the Discrimination of both IMT and In-unit Disciplinary Incidents and IMT Restarts

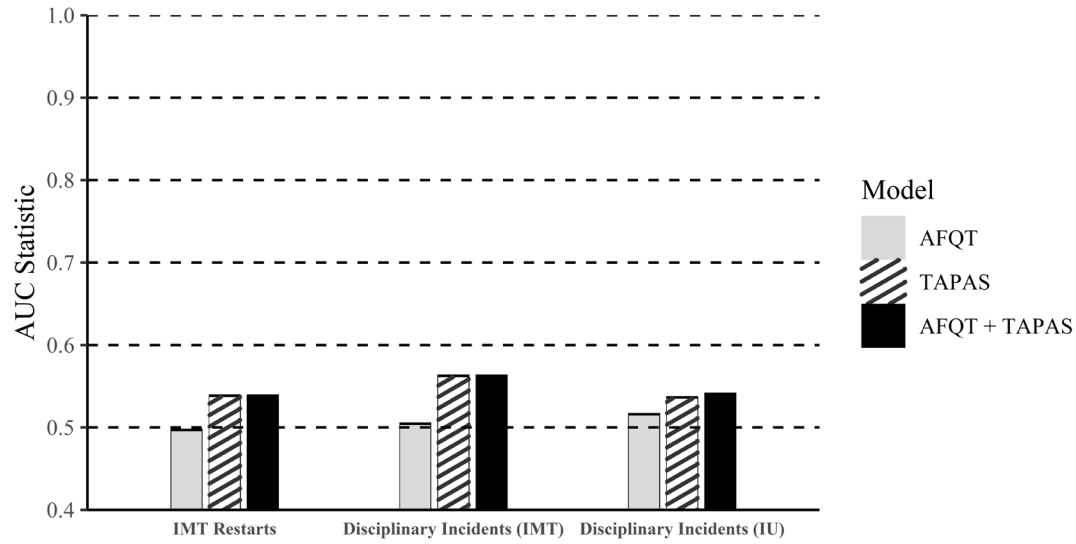


Figure 4.5

AUC Statistics of AFQT and TAPAS Adaptation Composite in the Discrimination of both IMT and In-unit Disciplinary Incidents and IMT Restarts

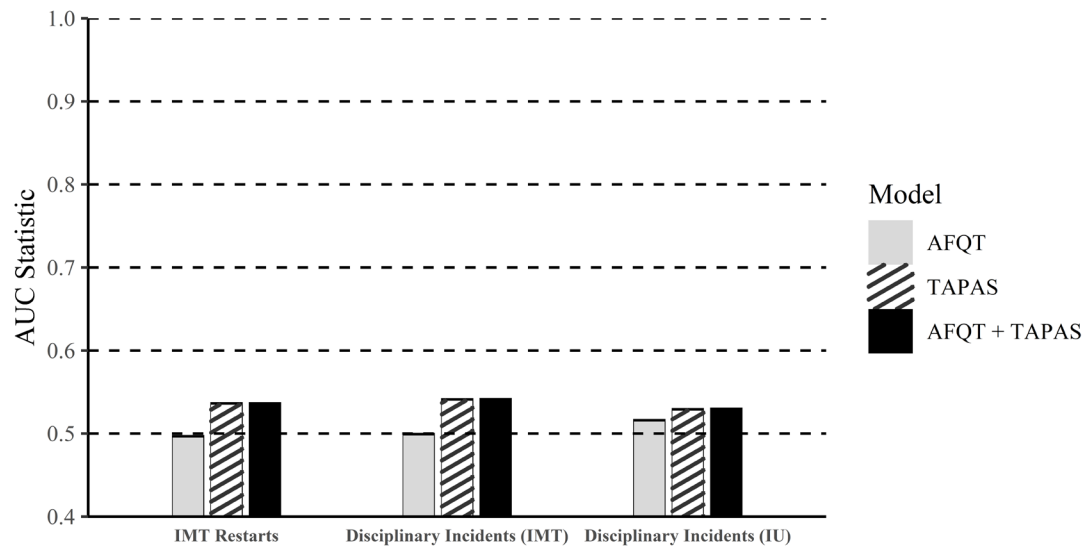
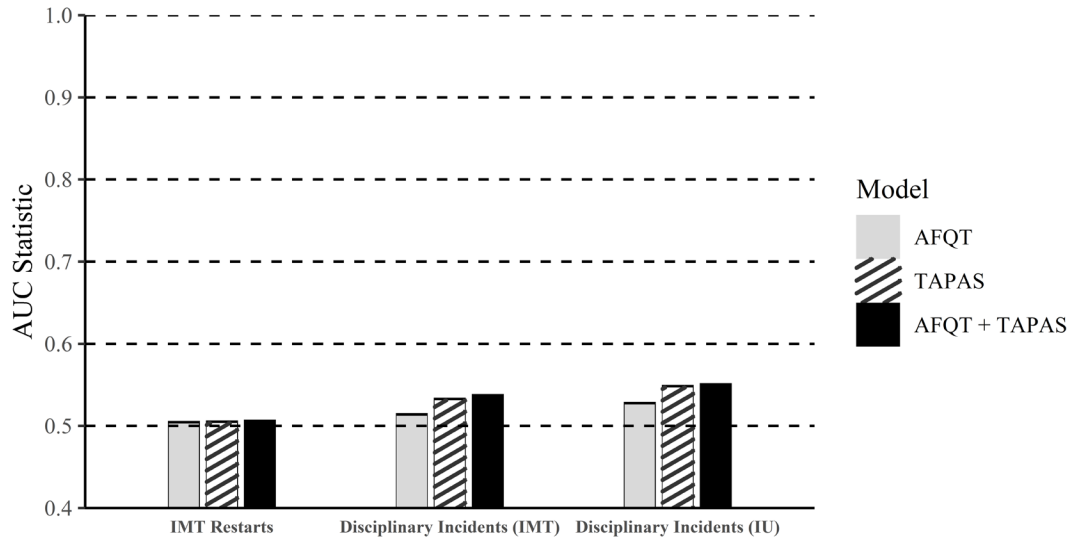


Figure 4.6

AUC Statistics of AFQT and TAPAS Conduct Composite in the Discrimination of both IMT and In-unit Disciplinary Incidents and IMT Restarts



For all three criteria, the quality of the logistic regression models in predicting the incidence of the outcome of interest increased when the Will-Do, Adaptation, or Conduct composite was added to the model. That said, these AUC values represent low prediction quality according to Rice and Harris's (2005) benchmarks. The highest prediction accuracy was achieved by the Will-Do composite in addition to AFQT for IMT Disciplinary Incidents at 56.1%. This result indicates that if we were to randomly sample pairs of Soldiers who experienced IMT Disciplinary Incidents and those who did not, the ones who did have an incident would score higher on the model-based composite only 56.1% of the time. On the other hand, results were different regarding the Can-Do composite. For in-unit Disciplinary Incidents and IMT Restarts, use of the Can-Do composite in addition to AFQT resulted in negligible change in predictive accuracy.

Table 4.6 presents the results of the logistic regression analyses examining attrition through 6, 12, 24, and 36 months of service for Regular Army Soldiers. Will-Do ($OR_{Will-Do} = .988, .989, .991, \text{ and } .992$, respectively) and Adaptation ($OR_{Adaptation} = .989, .989, .990, \text{ and } .991$, 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. The Conduct composite was negatively related to attrition at the 24-month and 36-month versions of this variable ($OR_{Conduct} = .998 \text{ and } .997$, respectively). Conversely, the Can-Do composite was positively related to attrition at each of the time points for Regular Army Soldiers ($OR_{Can-Do} = 1.002$ at all four time points). That being said, the strength of the positive relation was weak, with point estimates being 1.002 and the upper bounds of the estimates being as high as 1.003.

Table 4.6

Incremental Validity Estimates for TAPAS Composites over AFQT for Predicting Cumulative Attrition through 36 Months of Service (Regular Army Only)

Criterion Measure/Model	<i>n</i>	OR _{AFQT} (CI)	OR _{TAPAS} (CI)	<i>r_{pb}</i>	Δ-2LL
<i>6-month Attrition</i>					
Can-Do	422,950				
AFQT		.844 (.836-.853)		.05	
AFQT+TAPAS		.834 (.826-.843)	1.002 (1.001-1.002)	.05	36.08
Will-Do	488,737				
AFQT		.845 (.837-.852)		.05	
AFQT+TAPAS		.856 (.849-.864)	.988 (.988-.989)	.09	2797.02
Adaptation	455,826				
AFQT		.843 (.836-.851)		.05	
AFQT+TAPAS		.869 (.861-.877)	.989 (.988-.989)	.09	2245.43
Conduct	77,387				
AFQT		.868 (.848-.889)		.04	
AFQT+TAPAS		.868 (.847-.889)	1.000 (.999-1.002)	.04	0.57
<i>12-month Attrition</i>					
Can-Do	417,322				
AFQT		.864 (.856-.872)		.05	
AFQT+TAPAS		.854 (.846-.862)	1.002 (1.001-1.002)	.05	41.08
Will-Do	483,071				
AFQT		.866 (.859-.873)		.05	
AFQT+TAPAS		.878 (.870-.885)	.989 (.988-.989)	.09	3178.68
Adaptation	450,175				
AFQT		.864 (.857-.872)		.05	
AFQT+TAPAS		.891 (.883-.899)	.989 (.988-.989)	.09	2782.69
Conduct	72,179				
AFQT		.876 (.857-.896)		.04	
AFQT+TAPAS		.876 (.857-.896)	1.000 (.999-1.001)	.04	0.02
<i>(continued)</i>					

Table 4.6 (continued)

Criterion Measure/Model	<i>n</i>	OR _{AFQT} (CI)	OR _{TAPAS} (CI)	<i>r_{pb}</i>	Δ-2LL
<i>24-month Attrition</i>					
Can-Do	368,859				
AFQT		.829 (.823-.836)		.07	
AFQT+TAPAS		.818 (.810-.825)	1.002 (1.001-1.002)	.07	74.60
Will-Do	434,347				
AFQT		.833 (.827-.839)		.07	
AFQT+TAPAS		.842 (.835-.848)	.991 (.991-.992)	.10	2297.28
Adaptation	401,568				
AFQT		.831 (.824-.837)		.07	
AFQT+TAPAS		.854 (.847-.861)	.990 (.990-.990)	.11	2640.33
Conduct	34,161				
AFQT		.842 (.819-.865)		.06	
AFQT+TAPAS		.844 (.821-.868)	.998 (.996-.999)	.07	14.23
<i>36-month Attrition</i>					
Can-Do	310,973				
AFQT		.799 (.793-.806)		.09	
AFQT+TAPAS		.786 (.779-.793)	1.002 (1.002-1.003)	.10	101.76
Will-Do	376,108				
AFQT		.804 (.798-.810)		.09	
AFQT+TAPAS		.811 (.805-.817)	.992 (.992-.993)	.11	1771.25
Adaptation	343,489				
AFQT		.801 (.795-.808)		.09	
AFQT+TAPAS		.822 (.816-.828)	.991 (.990-.991)	.12	2366.00
Conduct	33,579				
AFQT		.815 (.795-.836)		.09	
AFQT+TAPAS		.818 (.797-.839)	.997 (.996-.998)	.09	21.86

Note. OR = odds ratio for each predictor. CI = 95% confidence interval of the odds ratio. rpb = point biserial correlation between the observed outcome and predicted probability. Δ-2LL = change in negative two log likelihood (deviance) from adding the selected TAPAS composite score 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 Δ-2LL, bolded indicate significant change in model fit based on a Likelihood Ratio χ^2 test, $p < .05$.

Figures 4.7 to 4.10 present the accuracy of logistic regression models in predicting attrition criteria. The results were similar for the Will-Do and Adaptation composites in that these composites alone have higher prediction accuracy of attrition criteria compared to AFQT alone, and the highest prediction accuracy was achieved when AFQT and the composites were used concurrently as predictors. The highest prediction accuracy was achieved by the Will-Do composite for 6-month attrition at 58.0%. However, like the results for IMT Restarts and Disciplinary Incidents, this is still a small effect size according to Rice and Harris's (2005) benchmarks. Meanwhile, the Can-Do and Conduct composites had similar results in that these composites performed poorly as standalone predictors compared to AFQT alone. However, using these composites in addition to the AFQT slightly enhanced prediction accuracy.

Figure 4.7

AUC Statistics of AFQT and TAPAS Can-Do Composite in the Discrimination of Cumulative Attrition

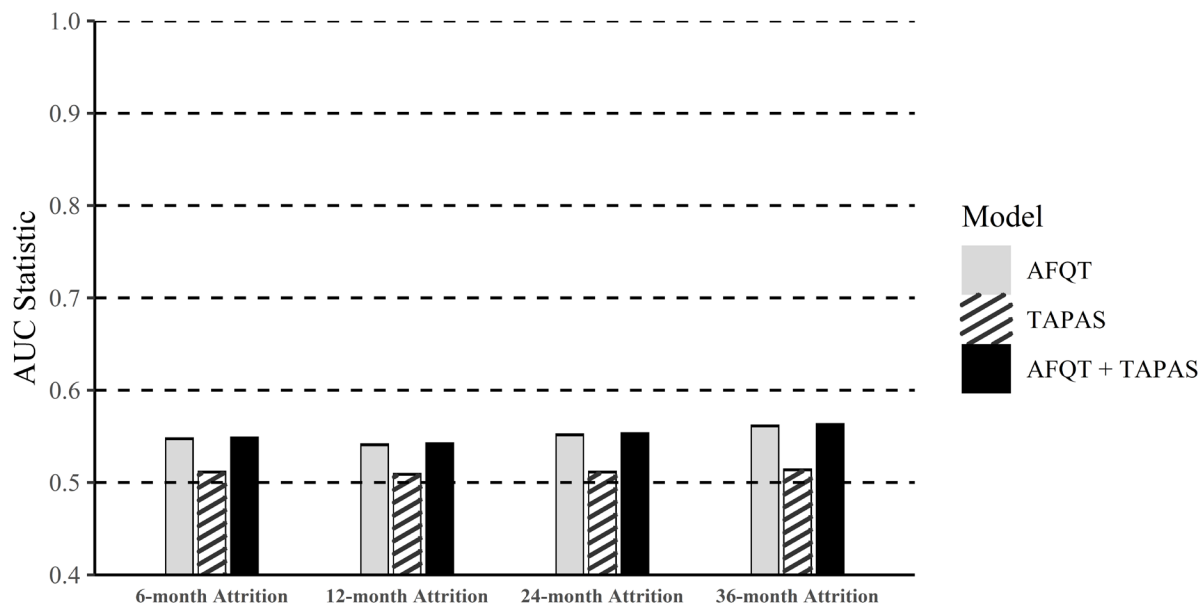


Figure 4.8

AUC Statistics of AFQT and TAPAS Will-Do Composite in the Discrimination of Cumulative Attrition

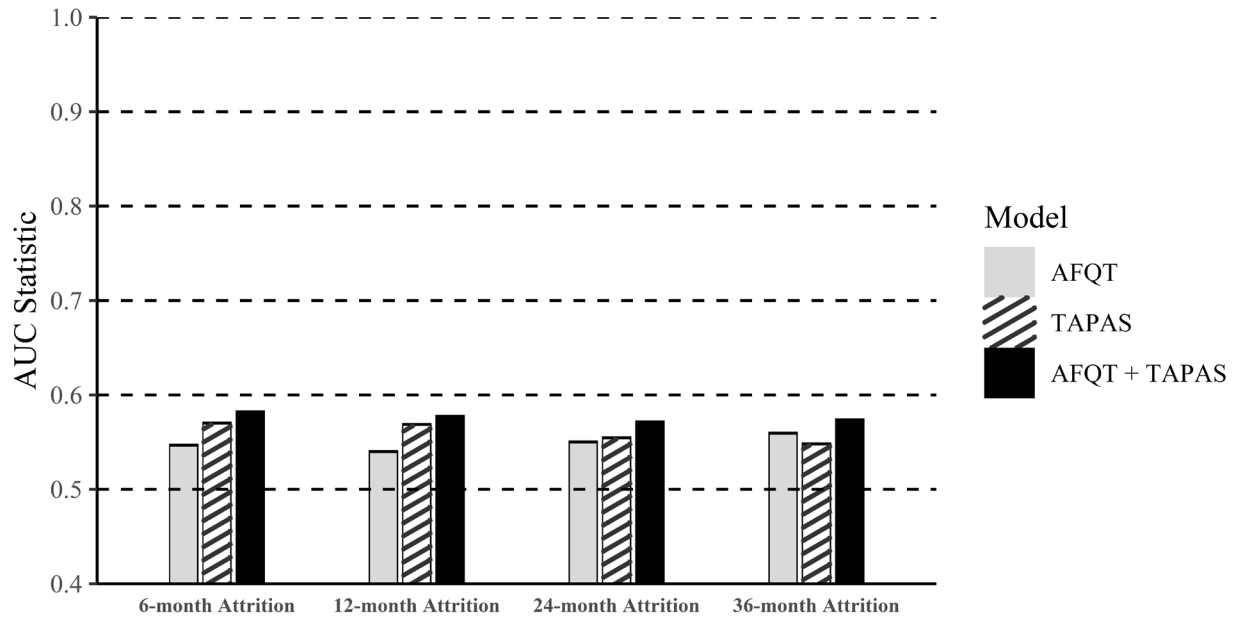


Figure 4.9

AUC Statistics of AFQT and TAPAS Adaptation Composite in the Discrimination of Cumulative Attrition

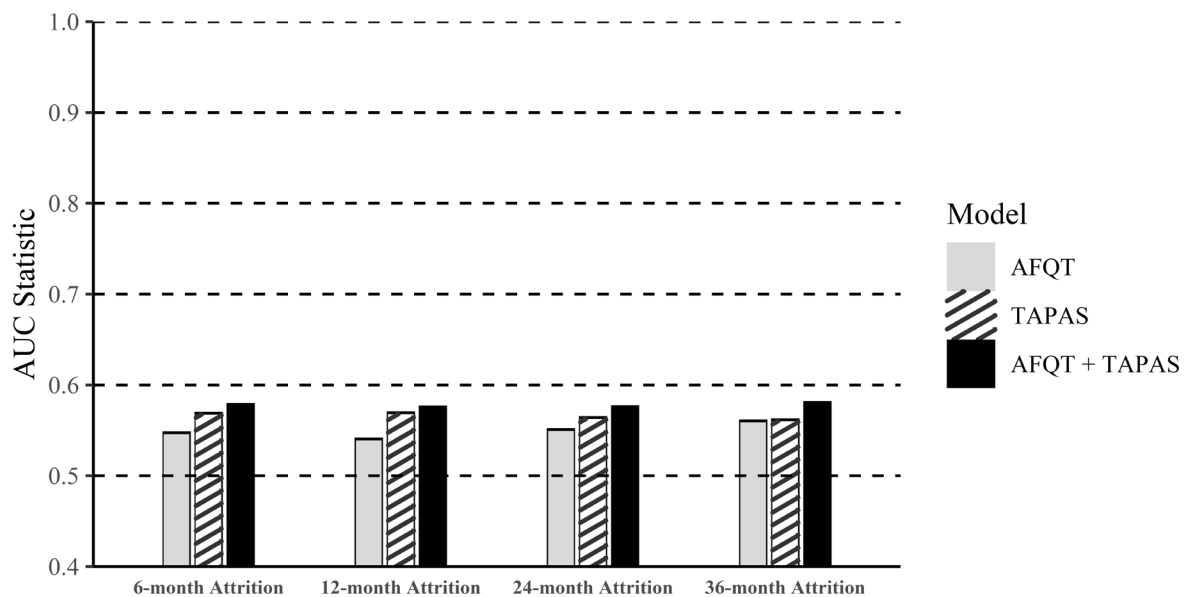
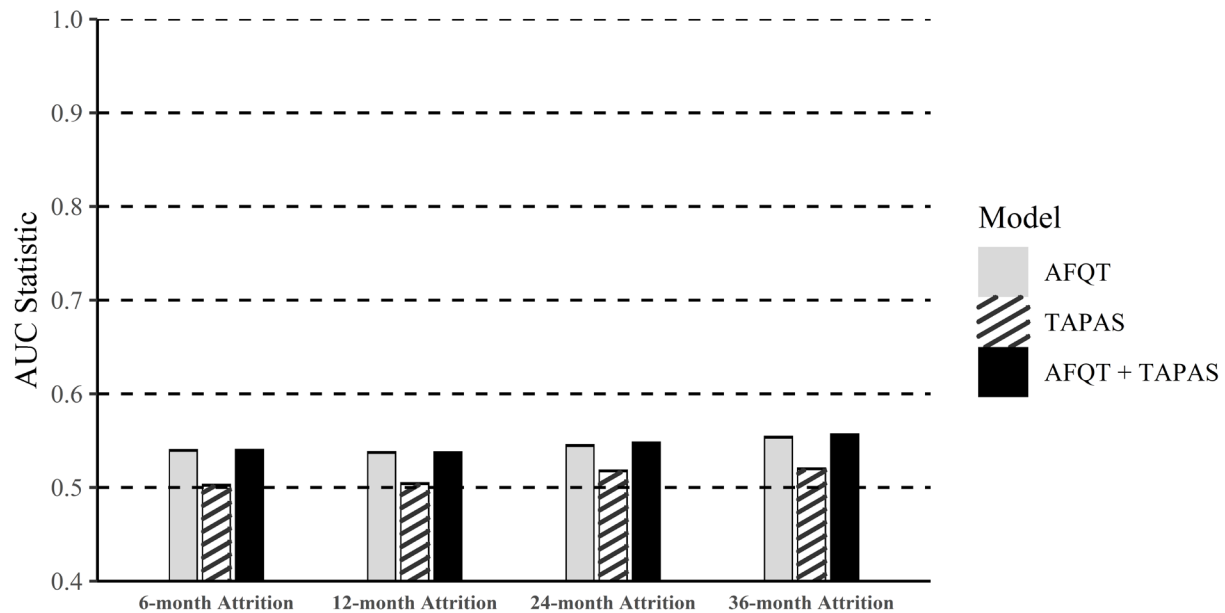


Figure 4.10

AUC Statistics of AFQT and TAPAS Conduct Composite in the Discrimination of Cumulative Attrition



Summary

In this chapter, we examined the relationships between predictor and criterion variables as well as the incremental validity of TAPAS composites over and above AFQT in predicting criteria. As expected, AFQT exhibited positive moderate correlations with cognitive criteria. AFQT also negatively correlated with retention cognitions.

The Can-Do composite also correlated positively with cognitive criteria, although to a lesser extent compared to AFQT. The Can-Do composite did not provide incremental validity in predicting criteria over and above AFQT. Interestingly, when used to predict attrition criteria concurrently with AFQT, the Can-Do composite was positively associated with attrition.

The Will-Do composite correlated moderately with motivational criteria (APFT score, Commitment and Fit, Peer Leadership, and Peer Performance Rating composite). Moreover, this TAPAS composite provided incremental validity over and above AFQT in predicting these criteria. The Will-Do composite also increased predictive accuracy of dichotomous criteria when used as a predictor in addition to AFQT.

The Adaptation composite showed similar relations to criteria as did the Will-Do composite, although to a lesser extent. This was somewhat expected due to the moderate correlation between the two ($r = .56$). Although the Adaptation composite is designed to predict attrition, it did not correlate with attrition criteria much stronger than Will-Do did. That said, the

Adaptation composite did provide incremental validity for predicting APFT and the IMT Peer Performance Rating composite as well as improve prediction accuracy of all dichotomous criteria.

Consistent with its design, the Conduct composite correlated negatively with Counterproductive Soldier Behavior (IMT) and Disciplinary Incidents (IMT and in-unit). The Conduct composite also provided incremental validity above and beyond AFQT for Counterproductive Soldier Behavior (IMT) and APFT (in-unit) and increased predictive accuracy of IMT and in-unit Disciplinary Incidents. The Conduct composite also increased predictive accuracy for 24- and 36-month attrition but not for the 6- and 12-month attrition criteria.

REFERENCES

- Allen, M. T., Knapp, D. J., & Owens, K. S. (2016). *Validating future force measures (Army Class): Concluding analyses* (Technical Report 1355). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1016104)
- Bynum, B. H., & Mullins, H. M. (Eds.) (2017). *Tier One Performance Screen Initial operational test and evaluation: 2013 annual report* (Technical Report 1359). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1091944)
- Campbell, J. P., & Knapp, D. J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. 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). Lawrence Erlbaum Associates, Inc.
- Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology*, 43, 313-333. <https://doi.org/10.1111/j.1744-6570.1990.tb01561.x>
- 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. <https://doi.org/10.1111/ijsa.12131>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd edition)*. Lawrence Erlbaum Associates, Inc.
- Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115, 928-935. <http://doi.org/10.1161/CIRCULATIONAHA.106.672402>
- Defense Manpower Data Center (DMDC) (2008). *ASVAB Technical Bulletin No. 3 CAT-ASVAB Forms 5-9*. Defense Manpower Data Center. https://www.officialasvab.com/wp-content/uploads/2019/08/asvab_techbulletin_3.pdf
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA564422)
- Farmer, W. L., Watson, S. E., Alderton, D. L., Michael, P. G., & Hindelang, R. (2006). *Improving person-job congruence during the classification process: Item development and initial testing of a pictorial interest instrument* (NPRST-TN-06-8). Navy Personnel Research, Studies, and Technology Division, Bureau of Naval Personnel. <https://discover.dtic.mil> (ADA456987)

- Ford, L. A., Knapp, D. J., Russell, T. L., & Thacker, A. A. (2021). *Validity argument 1.0 for the Tailored Adaptive Personality Assessment System (TAPAS)* (2021 No. 095). Human Resources Research Organization.
- Houston, J. S., Borman, W. C., Farmer, W. F., & Bearden, R. M. (2006). *Development of the Navy Computer Adaptive Personality Scales (NCAPS)* (NPRST-TR-06-2). Navy Personnel Research, Studies, and Technology Division Bureau of Naval Personnel. <https://discover.dtic.mil> (ADA456977)
- Hughes, M. G. (2020). 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: Capstone report* (Technical Report 1380). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1091944)
- Hughes, M. G., O'Brien, E. L., Reeder, M. C., & Purl, J. (2020). Attrition and reenlistment in the Army: Using the Tailored Adaptive Personality Assessment System (TAPAS) to improve retention. *Military Psychology*, 32, 36-50. <https://doi.org/10.1080/08995605.2019.1652487>
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Intelligence*, 29, 340-362. [https://doi.org/10.1016/0001-8791\(86\)90013-8](https://doi.org/10.1016/0001-8791(86)90013-8)
- Ingerick, M., Diaz, T., & Putka, D. (2009). *Investigations into Army enlisted classification systems: Concurrent validation report* (Technical Report 1244). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA500985)
- Kirkendall, C. D., Nye, C. D., Rounds, J., Drasgow, F., Chernyshenko, O. S., & Stark, S. (2020). Adaptive Vocational Interest Diagnostic: Informing and improving the job assignment process. *Military Psychology*, 32(1), 91-100. <https://doi.org/10.1080/08995605.2019.1652480>
- Knapp, D. J., & Heffner, T. S. (Eds.) (2009). *Predicting Future Force Performance (Army Class): End of Training Longitudinal Validation* (Technical Report 1257). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA507669)
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA523962)
- Knapp, D. J., & Heffner, T. S. (Eds.) (2011). *Tier One Performance Screen initial operational test and evaluation: 2010 annual report* (Technical Report 1296). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA552618)
- Knapp, D. J., & Heffner, T. S. (Eds.) (2012). *Tier One Performance Screen initial operational test and evaluation: 2011 interim report* (Technical Report 1306). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA563937)

- Knapp, D. J., Heffner, T. S., & White, L. (Eds.) (2011). *Tier One Performance Screen initial operational test and evaluation: Early results* (Technical Report 1283). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA544437)
- Knapp, D. J., & Kirkendall, C. D. (Eds.) (2023). *Exploring personality and performance criterion relationships using the Tailored Adaptive Personality Assessment System (TAPAS)* (Technical Report 1437). U.S. Army Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1211161)
- Knapp, D. J., & Kirkendall, C. D. (Eds.) (2018). *Tier One Performance Screen initial operational test and evaluation: 2015-2016 biennial report* (Technical Report 1367). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1060737)
- Knapp, D. J., & Kirkendall, C. D. (Eds.) (2020). *Tier One Performance Screen initial operational test and evaluation: Capstone report* (Technical Report 1380). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1091944)
- Knapp, D. J., & Kirkendall, C. D. (2021). *Validation of Accessions Screening Tools: Cycle 20 evaluation* (Research Note 2022-01). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1152016)
- Knapp, D. J., & LaPort, K. (Eds.) (2013a). *Tier One Performance Screen initial operational test and evaluation: 2011 annual report* (Technical Report 1325). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA569024)
- Knapp, D. J., & LaPort, K. (Eds.) (2013b). *Tier One Performance Screen initial operational test and evaluation: 2012 interim report* (Technical Report 1332). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA591889)
- Knapp, D. J., & LaPort, K. (Eds.) (2014). *Tier One Performance Screen initial operational test and evaluation: 2012 annual report* (Technical Report 1342). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA602475)
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA602475)
- Knapp, D. J., & Tremble, T. R. (Eds.) (2007). *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA471963)

- Knapp, D. J., & Wolters, H. M. K. (Eds.) (2017). *Tier One Performance Screen initial operational test and evaluation: 2014 annual report* (Technical Report 1358). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1036955)
- Maier, M. H. (1993). *Military aptitude testing: The past fifty years* (DMDC Technical Report 93-007). Defense Manpower Data Center. <https://discover.dtic.mil> (ADA594291)
- Marshall, A. D., Nesbitt, C. K., Burgoyne, T., & Knapp, D. J. (2023). VAST program criterion measures. In D. J. Knapp & C. D. Kirkendall, C. D. (Eds.) *Exploring personality and performance criterion relationships using the Tailored Adaptive Personality Assessment System (TAPAS)* (Technical Report 1437). U.S. Army Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1211161)
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA507668)
- Nesbitt, C. K., Knapp, D. J., & Kirkendall, C. D. (2022). *Validation of Accessions Screen Tools (VAST): Cycle 22 evaluation* (Research Note 2023-02). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1189406)
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (AD1091944)
- Nye, C. D., Drasgow, F., Chernyshenko, O. S., Stark, S., Kubisiak, U. C., White, L. A., & Jose, I. (2012). *Assessing the Tailored Adaptive Personality Assessment System (TAPAS) as an MOS qualification instrument* (Technical Report No. 1312). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/pdfs/ADA566090.pdf>
- 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. <https://doi.org/10.1037/0021-9010.93.5.959>
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29, 615-620. <https://psycnet.apa.org/doi/10.1007/s10979-005-6832-7>
- Roberts, J. S., Winfred, A., Reckase, M., Sackett, P., & Zenisky, A. (2019). *TAPAS Evaluation project report on the readiness of TAPAS for use in selection and classification of military applicants* (2019 No. 076). Human Resources Research Organization.

- Rose, M. R., Manley, G. G., & Weissmuller, J. J. (2013). *Development of two-and three-factor classification models for Air Force Battlefield Airmen (BA) and related AFSs*. Air Force Personnel Center Strategic Research and Assessment Branch.
<https://dtic.mil/sti/pdfs/AD1007586.pdf>
- Sharp, M. A., Foulis, S. A., Redmond, J. E., Canino, M. C., Cohen, B. S., Hauret, K., Frykman, P. N., Pierce, J. R., Westrick, R. B., Pacheco, B. M., Gebhardt, D. L., Zambraski, E. J. (2018). *Longitudinal validation of the Occupational Physical Assessment Test (OPAT)*. (Report #18-05). U.S. Army Research Institute of Environmental Medicine.
<https://discover.dtic.mil> (AD1057800)
- Strickland, W. J. (Ed.) (2005). *A longitudinal examination of first term attrition and reenlistment among FY1999 enlisted accessions* (Technical Report 1172). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA448564)
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). 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). Human Resources Research Organization.
- Trent, J. D., Barron, L. G., Rose, M. R., & Carretta, T. R. (2020). Tailored Adaptive Personality Assessment System (TAPAS) as an indicator for counterproductive work behavior: Comparing validity in applicant, honest, and directed faking conditions. *Military Psychology*, 32(1), 51-59. <https://doi.org/10.1080/08995605.2019.1652481>
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://discover.dtic.mil> (ADA544437)
- Trippe, D. M., Moriarty, K. O., Russell, T. L., Carretta, T. R., & Beatty, A. S. (2017). Development of a cyber/information technology knowledge test for military enlisted technical training qualification. *Military Psychology*, 26(3), 182-198.
<http://dx.doi.org/10.1037/mil0000042>
- U.S. Army (2019). *Army People Strategy*. <https://people.army.mil/wordpress/wp-content/uploads/2019/10/The-2020-Army-People-Strategy-Final.pdf>
- U.S. Army (2020). *2020 Army People Strategy. Military implementation plan*.
<https://people.army.mil/overview-2/implementation/>
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences.
<https://discover.dtic.mil> (ADA507669)

- Wolters, H., White, L., Drasgow, F., Stark, S., Nye, C., & Chernyshenko, O. S. (2017, July). *TAPAS research update: Ongoing and future research*. [Presentation]. Defense Advisory Committee on Military Personnel Testing meeting.
- 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.
<https://doi.org/10.1177/1094428119836486>

APPENDIX

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
EEEM	Expanded Enlistment Eligibility Metrics
ICTL	Information, Communication, and Technical Literacy
IMT	Initial Military Training
IRR	Interrater 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