MILITARY PSYCHOLOGY
2020, VOL. 32, NO. 1, 24–35
https://doi.org/10.1080/08995605.2019.1652484

Routledge
Taylor & Francis Group

Validation of the TAPAS for predicting in-unit soldier outcomes

Cristina Kirkendall[a], Bethany Bynum[b], Colby Nesbitt[b], and Michael Hughes[b]

[a]U. S. Army Research Institute for the Behavioral and Social Sciences, Ft. Belvoir, Virginia; [b]Human Resources Research Organization, Alexandria, Virginia, USA

ABSTRACT

The Tailored Adaptive Personality Assessment System (TAPAS) is a computer-adaptive temperament test that has been used operationally by the Army since 2009 to examine characteristics that are not captured by other testing requirements. Outcomes collected at multiple time points throughout a Soldier's career show the TAPAS to be predictive of performance at the end of initial military training, as well as in-unit. This article focuses on the performance of Soldiers in their first or second units of assignment. Specifically, we will look at the predictive value of TAPAS over and above the Armed Services Vocational Aptitude Battery (ASVAB) for predicting Army outcomes, including Army fit and commitment, as well as job knowledge and attrition. We will also examine the TAPAS composites, which are optimized to predict performance in first-term Soldiers, and the operational history and implementation of the TAPAS.

ARTICLE HISTORY
Received 13 June 2019
Accepted 24 June 2019

KEYWORDS
Personality testing; military selection

What is the public significance of this article?—An organization's effectiveness is based on the people who comprise it. Thus, selection and placement of personnel directly impact that effectiveness. Historically, cognitive testing has been the dominant tool for these purposes. However, cognitive tests have been shown to be limited in predicting elements of success beyond technical proficiency. They do not predict well those aspects of performance which depend on the individual's motivation to perform well over time, or to remain with the organization over time. For these outcomes, noncognitive attributes such as personality and vocational interests provide critical predictive information. This special issue demonstrates the effectiveness of personality and interest measures in a military context, and how these tools are transforming the military selection and classification process. The effort reported in this issue marks major changes in the selection and classification process, changes that can help both military and civilian organizations be more productive and successful.

In the U.S. Army, there are many aspects to take into account when considering what makes a successful Soldier. For example, successful Soldiers must be adept at performing the technical requirements of their specific job, or military occupational specialty (MOS). They must also be physically fit to handle the general requirements of the military, must be adaptive, able to work in teams, and able to follow instructions and safety procedures. Successful Soldiers must also have a commitment to the Army and stay in service long enough to provide a reasonable return on investment for the Army.

The above examples are by no means an exhaustive list of performance criteria but they do cover a broad array of performance indicators for success. Generally, performance in the Army is conceptually broken down into four areas: can-do performance, which is comprised of the technical knowledge and proficiency required for the job; will-do performance, which reflects motivational aspects; fit with the Army or with the job itself; and attrition.

When discussing the capabilities of a Soldier to handle job tasks involving knowledge, there is no better predictor than cognitive ability, which is measured by the Armed Forces Vocational Aptitude Battery (ASVAB). The ASVAB has several mathematical and verbal subtests that are combined to determine an overall Armed Forces Qualifying Test (AFQT), which is used for establishing eligibility for Army service. In addition to the AFQT, ASVAB subtests are combined to determine eligibility requirements for specific MOS.

CONTACT Cristina Kirkendall cristina.kirkendall@gmail.com U.S. Army Research Institute for the Behavioral and Social Sciences, 6000 6th St. Bldg 1464, Ft. Belvoir, VA 22060.

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the Department of the Army, DOD, or the U.S. Government. The research described herein was sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences, Department of the Army Contract No. W911NF-16-C-0056.

Aside from cognitive ability testing, the Army also uses other screening procedures to determine eligibility. These additional procedures are designed to tap into the other aspects of performance listed above. The Army recently implemented the Occupational Physical Assessment Test (OPAT), which assesses applicants' physical aptitude and places them in one of three categories (i.e., Significant, Moderate, or Heavy physical demands) and determines eligibility for MOS. Background, medical, and education information is also collected to make sure the applicant is fit for service. Taken together, this screening process mostly covers what applicants *can do*, or their general ability to complete job tasks. Although the ASVAB has proven to be, and will continue to serve as, a useful metric for selecting new Soldiers, there is a growing recognition of the need to consider whole person assessment that takes other personal attributes, in particular non-cognitive attributes (e.g., temperament, interests, and values) into consideration. Non-cognitive attributes are important to entry-level Soldier performance and retention (e.g., Campbell & Knapp, 2001; Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009, 2010; Knapp & Tremble, 2007). To fill this gap, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) developed the Tailored Adaptive Personality Assessment System (TAPAS), which is a forced-choice measure of personality or temperament designed to tap into the motivational aspects of performance. In this paper, we will examine the development of the TAPAS, its operational use, and its utility in predicting outcomes including Army fit, commitment, attrition, and job knowledge in early career Army Soldiers.

Tailored Adaptive Personality Assessment System (TAPAS)

The TAPAS is different from traditional personality measures in several ways. First, the TAPAS is a multidimensional forced-choice measure (Stark, Chernyshenko, & Drasgow, 2005). This means that the applicant chooses which of two presented statements is more representative of him or her (e.g., "I am not one to volunteer to be group leader but would serve if asked" or "My life has had about an equal share of ups and downs.") Using statements that are from different dimensions but matched on social desirability and intensity ensures that neither of the options is obviously right or wrong and limits faking. Second, the TAPAS uses an item response theory framework which means that instead of including items that are only high or low on the trait continuum, as in Likert-type tests, the items cover all levels of the trait continuum (Stark, Chernyshenko, Drasgow, & Nye, 2017). This allows

Table 1. Personality dimensions of the tailored adaptive personality assessment system (TAPAS).

Facet Name	Brief Description
Achievement	High scoring individuals are seen as hard working, ambitious, confident, and resourceful.
Adjustment	High scoring individuals are well adjusted, worry free, and handle stress well.
Adventure Seeking	High scoring individuals enjoy participating in extreme sports and outdoor activities.
esthetics	High scoring individuals appreciate various forms of art and music and participate in art-related activities more than most people.
Attention Seeking	High scoring individuals tend to engage in behaviors that attract social attention. They are loud, loquacious, entertaining, and even boastful.
Commitment to Serve	High scoring individuals identify with the military and have a strong desire to serve their country.
Consideration	High scoring individuals are affectionate, compassionate, sensitive, and caring.
Cooperation	High scoring individuals are pleasant, trusting, cordial, non-critical, and easy to get along with.
Courage	High scoring individuals stand up to challenges and are not afraid to face dangerous situations.
Curiosity	High scoring individuals are inquisitive and perceptive; they are interested in learning new information and attend courses and workshops whenever they can.
Dominance	High scoring individuals are domineering, "take charge" and are often referred to by their peers as "natural leaders."
Even Tempered	High scoring individuals tend to be calm and stable. They don't often exhibit anger, hostility, or aggression.
Ingenuity	High scoring individuals are inventive and can think "outside of the box."
Intellectual Efficiency	High scoring individuals believe they process information and make decisions quickly; they see themselves (and they may be perceived by others) as knowledgeable, astute, or intellectual.
Non-Delinquency	High scoring individuals tend to comply with rules, customs, norms, and expectations, and they tend not to challenge authority.
Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.
Order	High scoring individuals tend to organize tasks and activities and desire to maintain neat and clean surroundings.
Physical Conditioning	High scoring individuals tend to engage in activities to maintain their physical fitness and are more likely participate in vigorous sports or exercise.
Responsibility	High scoring individuals are dependable, reliable, and make every effort to keep their promises.
Self Control	High scoring individuals tend to be cautious, levelheaded, able to delay gratification, and patient.
Selflessness	High scoring individuals are generous with their time and resources.
Situational Awareness	High scoring individuals pay attention to their surroundings and rarely get lost or surprised.
Sociability	High scoring individuals tend to seek out and initiate social interactions.
Team Orientation	High scoring individuals prefer working in teams and make people work together better.
Tolerance	High scoring individuals are interested in other cultures and opinions that may differ from their own. They are willing to adapt to novel environments and situations.
Virtue	High scoring individuals strive to adhere to standards of honesty, morality, and "good Samaritan" behavior.

the TAPAS to better assign values to individuals who are in the middle range on a trait. Third, the TAPAS is an adaptive test, meaning that the item pairs presented are based on previous responses, making each test unique to an individual and reducing the potential for test compromise.

The TAPAS contains 26 personality dimensions (see Table 1), the combination of which varies across test versions with a core set of dimensions being administered in each version. This core set of dimensions is used to compute three composites: "Can-Do", "Will-Do", and "Adaptation". The "Can-Do" composite is designed to predict technical performance, "Will-Do" is designed to predict motivational aspects of performance, and "Adaptation" is designed to predict continuance behaviors (Knapp & Wolters, 2017). For test security, the scales comprising each of the composites are not presented.

The TAPAS instrument was originally designed to be a screen-in measure for Army applicants, meaning that applicants who scored low on AFQT could potentially still be accepted into the Army if they had passing TAPAS scores. However, changes in economic conditions and Army manning requirements led to the TAPAS being re-conceptualized as a measure to screen out applicants who scored low on AFQT and whose TAPAS scores indicated they were low in motivation.

TAPAS administration in the Army began in May 2009. At that time, only applicants in Education Tier 1, typically those with a high school diploma, took the TAPAS. Education Tier 2 applicants, typically those with alternative education credentials, such as GEDs, were screened using an alternative temperament test, Assessment of Individual Motivation (AIM; White & Young, 1998), and therefore did not take TAPAS. In early 2011, the Army replaced AIM with TAPAS, which they began administering regardless of the applicants' Education Tier. There have been several different operational screens of the TAPAS used since that time. Most recently, TAPAS was used to screen Army applicants in two categories: 1) all Education Tier 2 applicants, who were required to score in the top 70% percentile on the Adaptation composite and 2) Education Tier 1 applicants who scored below the mean on AFQT, who were required to score in the top 90% percentile on the Will Do and Adaptation composites.

Types of criteria

The present study seeks to investigate the contribution of the TAPAS for predicting a broad domain of relevant Soldier in-unit performance outcomes beyond what is predicted by the AFQT. Like the TAPAS composites, the criteria presented in this study are also organized around the conceptual structure of Can Do, Will Do, and Adaption. Table 2 (adapted from Allen, Knapp, & Owens, 2013) provides an overview of the criteria taxonomy in this paper and we will discuss each criterion type in turn below.

Can do criteria

As mentioned above, Can Do criteria are those that are generally related to the completion of job tasks and this is where the Army's traditional testing methods are focused. The AFQT has been consistently shown to predict task performance and it will likely remain the best predictor; however, there is potential for non-cognitive assessments to predict some Can Do criteria. Previous researchers found that non-cognitive predictors are related to technical performance and MOS-specific job knowledge tests at the end of initial military training (IMT) and in-unit (Allen et al., 2013; Stark et al.,

Table 2. Criteria taxonomy.

Criteria	Definition
Performance	Overall indicator of performance included aspects of both Can Do and Will Do performance
Can-Do Performance:	
Technical Task Proficiency	The degree to which the Solider performs the core substantive or technical tasks and duties, including those that are central to his or her MOS as well as those that are Army-wide.
Will-Do Performance:	
Maintaining Physical Fitness, Strength, and Weight	This dimension relates to behaviors and actions that involve physical exertion with the intention of promoting physical well-being and ability. Activities fostering fitness goals may include exercise routines and engagement in team or individual sporting activities.
Initiative, Persistence, and Effort	This dimension is defined by observable behaviors such as working extra hours, voluntarily taking on additional tasks, and working under extreme or adverse conditions.
Peer/Team Member Leadership Performance	This dimension refers to behaviors aimed at influencing peer/team members through interpersonal interaction and influence. It includes behaviors such as consideration and support, guiding, directing, goal emphasis, empowerment, facilitation, training, coaching, and serving as a model.
Fit	This component captures cognitions, attitudes, and views toward perceived fit with the Army as an organization, with the norms of Army life and subsequent appraisals of experiences with the Army. This includes satisfaction with all aspects of the Army experience, and being committed to orders, directives, and the Army in general.
Attrition	The final dimension captures separation from the Army. This includes cognitions toward both separating from and remaining with the Army, as well as actual separation from the Army and Army retention.

2014); however, neither of these studies looked at the incremental validity of non-cognitive predictors over and above AFQT scores for in-unit Can Do outcomes. The current study expands on this research by investigating the relationship of personality for in-unit Can Do criteria, taking AFQT scores into account as well.

Will do criteria

One of the areas where TAPAS was designed to have the most impact is Will Do criteria, or those that are motivation-based and are less likely to be related to cognitive ability scores. As mentioned earlier, there are several things that make a successful Soldier and many, such as physical fitness and adjustment to Army life, have more to do with determination than knowledge. Past research on personality measures in the military has shown a relationship between these measures and subsequent performance. For example, Hough, Eaton, Dunnette, Kamp, and McCloy (1990) and Motowidlo and Van Scotter (1994) both showed that the Assessment of Background and Life Experiences (ABLE; White, Nord, Mael, & Young, 1993) has predictive validity for Will Do performance criteria, even over and above cognitive ability.

TAPAS was designed to build upon the non-cognitive research using the ABLE and other early personality assessments, and researchers have shown that TAPAS is also an effective predictor of motivational aspects of performance. Stark et al. (2014) found that TAPAS Will-Do scores were positively related to Army Physical Fitness Test (APFT) scores, and negatively related to disciplinary incidents in IMT. Stark et al. (2014) also found that the TAPAS Will-Do composite was predictive of Army Physical Fitness Test (APFT) scales at the end of IMT.

Fit and attrition

Because of the costs associated with recruiting and training each Soldier, attrition is one of the most important criteria. While the education tier system described earlier was intended to be a predictor of attrition, revisions made over time have reduced its ability to serve this function (Laurence, Ramsberger, & Arabian, 1997). To augment the system, ARI has looked to non-cognitive measures and has found that non-cognitive measures, including TAPAS, are predictive of attrition (e.g., White et al., 1993; White, Rumsey, Mullins, Nye, & LaPort, 2014).

The reasons for attrition vary over time, and the causes differ for those that leave during Army training and those that leave after they are in their units of assignment (Strickland, 2005). Strickland (2005) also found attitudinal variables to be indicators of attrition in-unit. Commitment, fit, satisfaction, and continuance intentions were all related to attrition, demonstrating the role of psychological processes in the decision to leave the Army and highlighting the importance of including non-cognitive predictors in the accessioning process.

In-unit validation of TAPAS

While past research has demonstrated the value of non-cognitive predictors for selection into the Army, much of the work was done with criteria collected at the end of training. The current investigation expands on this research by looking at Soldiers in their first units of assignment. The current research also investigates the incremental validity that TAPAS provides over and above the AFQT for in-unit outcomes, which was not included in previous work. The expansion of the literature shows the practical, on-the-job implications of including non-cognitive measures in the Army selection process.

Method

Sample

Since January 2010, researchers and contractors of ARI have collected data from in-unit Soldiers during proctored online data collections at all active duty Army installations. Participation in the data collection was voluntary and the information collected has been used for research purposes only. Most Soldiers take the Tailored Adaptive Personality Assessment System (TAPAS) as part of their accession into the Army. Criterion data were collected from 7,228 Soldiers who completed IMT between 1 January 2010 and 1 December 2016. Because the criterion data were collected in the Soldiers' units of assignment, time between TAPAS administration and criterion data collection ranged from 1 to 6.73 years ($M = 1.88$, $SD = 1.01$). Four hundred and seven Soldiers' data were removed from the analysis sample due to insufficient responding or lack of valid predictor data, resulting in an analysis sample of 6,821 Soldiers. Table 3 describes the demographic characteristics of the sample. For most analyses, Soldiers from all components are included. However, for analyses involving separation data, results are only presented for the Regular Army Soldiers.

Measures

Predictors

Tailored adaptive personality assessment system (TAPAS). The TAPAS measures 26 personality dimensions (as shown in Table 1). Nine versions of the TAPAS

Table 3. Sample background and demographic characteristics.

Characteristic	*n*	%
Component		
Regular	6,787	99.6
Army National Guard	19	0.3
United States Army Reserve	5	0.1
Education Tier		
Tier 1	6,624	97.1
Tier 2	197	2.9
Military Occupational Specialty		
11B/C/X + 18X: Infantry	1,710	25.1
12B: Combat Engineer	266	3,9
13B: Cannon Crewman	174	2.6
13D: Field Artillery Tactical Data	80	1.2
13F: Fire Support Specialist	84	1.2
19D: Cavalry Scout	337	4.9
19K: Armor Crew Member	158	2.3
31B: Military Police	142	2.1
42A: Human Resource Specialist	171	2.5
68W: Health Care Specialist	296	4.3
88M: Motor Transport Operator	370	5.4
91B: Wheeled Vehicle Mechanic	342	5.0
Other	2,833	41.5
AFQT Category		
I	335	4.6
II	1,909	26.4
IIIA	1,569	21.7
IIIB	2,982	41.3
IV	433	6
Gender		
Female	823	12.1
Male	5,721	83.9
Missing	277	4.1
Race		
African American	1,564	22.9
American Indian	46	0.7
Asian	370	5.4
Hawaiian/Pacific Islander	23	0.3
Caucasian	4,686	68.7
Multiple	30	0.4
Declined to Answer/Missing	102	1.5
Ethnicity		
Hispanic/Latino	1,137	16.9
Not Hispanic	5,594	82.0
Declined to Answer/Missing	90	1.3

have been administered since May 2009. Each version of the TAPAS measured 13 or 15 dimensions, with a core set of dimensions being administered on each version. The TAPAS scales are scored using an item response theory (IRT) model for multidimensional pairwise preference (MDPP) items (Stark et al., 2005). Three regression-weighted composites were constructed using a subset of the TAPAS personality dimensions. The composites were developed to predict: (a) technical training performance and completion (Can Do) and (b) motivation-based performance (Will Do) and (c) Attrition (Adaptation). This research focused on the predictive utility of the three composite scores. Table 4 provides the descriptive statistics for the three TAPAS predictor composites.

Table 4. Descriptive Statistics of AFQT and TAPAS Composite scores.

Domain/Measure	n	M	SD	Min	Max
Predictors					
AFQT	6,821	55.22	19.75	10.00	99.00
TAPAS: Can-Do	4,607	100.72	20.24	30.87	166.42
TAPAS: Will-Do	6,821	100.59	19.23	30.65	158.38
TAPAS: Adaptation	6,821	102.40	19.39	23.10	167.75

Armed forces qualifying test (AFQT). The Armed Services Vocational Aptitude Battery (ASVAB) is a multiple aptitude battery of tests administered at the Military Entrance Processing Command (MEPCOM). Most military applicants take the computer adaptive version of ASVAB (i.e., the CAT-ASVAB). Scores on the ASVAB tests are combined to create composite scores for use in selecting applicants into the Army and qualifying them for an MOS. The AFQT, the composite used for selecting applicants into the Army, comprises the Verbal Expression[1] (VE), Arithmetic Reasoning (AR), and Math Knowledge (MK) tests. Table 4 provides the descriptive statistics for the AFQT scores.

Criteria

Army life questionnaire (ALQ). Each Soldier took a battery of assessments including the Army Life Questionnaire (ALQ) and a series of self-report performance items. The ALQ was designed to measure Soldiers' self-reported attitudes and experiences in the Army (Bynum & Beatty, 2014). The ALQ consists of Likert-type response scales that measure: (a) Army Fit, (b) MOS Fit, (c) Career intentions, (d) Reenlistment intentions, and (e) Attrition Cognitions. Table 5 provides a description of each of these scales. In addition, Soldiers were asked to write-in their Army Physical Fitness Test (APFT) score. Because career intentions, reenlistment intentions, and attrition cognition measure a similar underlying construct related to a Soldier's general intentions of continuance in the Army, we combined these scales into an overall Retention Cognitions composite score. The average correlation among these scales was .57, suggesting the scales measure a similar construct.

ALQ data were flagged as unusable if the Soldier (a) omitted more than 10% of the assessment items, (b) took fewer than five minutes to complete the entire assessment, or (c) chose an implausible response to the careless responding item. APFT scores were dropped if less than 180, which suggested the score was being reported in error.

Warrior tasks and battle drills (WTBD) job knowledge test. All Soldiers were administered a generic job knowledge test (JKT) called the Warrior Tasks and Battle Drills (WTBD). The WTBD JKT is intended to measure Army-Wide job performance (Bynum & Beatty, 2014). Most of the JKT items are in a multiple-choice format with two to four response options. However, other formats, such as multiple-response (i.e., check all that apply), rank ordering, and matching are also used. Many items use

Table 5. Army life questionnaire (ALQ) criterion composites.

Composite/Scale Name		Description	Number of Items	Example Item
MOS Fit		Measures Soldiers' perceived fit with their MOS.	9	My MOS provides the right amount of challenge for me.
Army Fit		Measures Soldiers' perceived fit with the Army.	8	The Army is a good match for me.
Retention Cognition	Career Intentions	Measures Soldiers' intentions to reenlist and to make the Army a career.	3	How likely is it that you will make the Army a career?
	Reenlistment Intentions	Measures Soldiers' intention to reenlist in the Army.	4	How likely is it that you will leave the Army after completing your current term of service?
	Attrition Cognitions	Measures the degree to which Soldiers think about attriting before the end of their first term.	4	How likely is it that you will complete your current term of service?

Table 6. In-unit army-wide performance rating scale dimensions and composite scores.

Rating Composites	*α*
Can Do	.90
Performing Core Warrior Tasks	
Performing MOS-Specific Tasks	
Processing Information	
Solving Problems	
Effort & Discipline	.79
Exhibiting Effort	
Exhibiting Personal Discipline	
Working with Others	.68
Communicating with Others	
Contributing to the Team	
Self-Management	.79
Following Safety Procedures	
Developing Own Skills	
Managing Personal Matters	
Single Item Dimensions	
Adjustment to Army Life	
Physical Fitness and Bearing	
Overall Leadership Potential Rating	

Of the seven performance ratings used in analyses, four are composites of multiple dimensions and three are single dimension ratings.

Table 7. Descriptive statistics of in-unit outcomes.

Domain/Measure	*n*	*M*	*SD*	*Min*	*Max*
Performance					
PRS: Overall Performance	5263	5.17	1.15	1.00	7.00
Technical Task Proficiency					
WTBD JKT	6452	62.26	10.65	15.38	96.15
PRS: Can Do	5273	4.88	1.26	1.00	7.00
Maintaining Physical Fitness, Strength, and Weight					
APFT scores	6296	253.75	27.85	180.00	300.00
PRS: Physical Fitness and Bearing	5293	5.25	1.55	1.00	7.00
Initiative, Persistence, and Effort					
PRS: Effort and Discipline	5270	5.15	1.37	1.00	7.00
PRS: Self-Management	5250	5.26	1.15	1.00	7.00
Peer/Team Member Leadership Performance					
PRS: Leadership Potential	5193	4.70	1.68	1.00	7.00
PRS: Working with Others	5272	5.26	1.22	1.00	7.00
Fit					
Army Fit	6574	3.84	0.72	1.00	5.00
MOS Fit	6574	3.21	0.93	1.00	5.00
PRS: Adjustment to Army Life	5001	5.19	1.55	1.00	7.00
Attrition					
Retention Cognitions	6574	2.42	0.66	1.00	4.14

All self-report outcomes were collected on a 5-point Likert-type scale of agreement. PRS outcomes are on a 7-point behaviorally anchored rating scale. Valid APFT scores range between 180–300.

visual images to make them more realistic and reduce reading requirements for the test.

We computed a single, overall raw score for each JKT by summing the total number of points Soldiers earned across the JKT items and computing a percent correct score based on the maximum number of points that could be obtained on each test. For the correlational analyses among criterion variables and criterion-related validity analyses, we converted the total raw scores to standardized scores (or *z*-scores) by standardizing the scores *within* each MOS. A JKT score was flagged and not included in analysis if the Soldier (a) omitted more than 10% of the assessment items, (b) took fewer than five minutes to complete the entire assessment, or (c) selected an implausible response to one of the careless responding items. The Cronbach's reliability of WTBD JKT was .57. Given the heterogeneous nature of the content of the test, the lower reliability was expected.

Performance rating scales. Supervisors were asked to complete a performance ratings scale (PRS) that includes 13 performance dimensions as well as a leadership potential item (see Table 6) for Soldiers within the supervisor's unit. The PRS was originally developed for the Army Class project and was designed to cover performance elements that would be common to all MOS. Details regarding the development of the PRS is documented in Moriarty, Campbell, Heffner, and Knapp (2009). The PRS uses a 7-point Behaviorally Anchored Rating Scale format, giving the supervisors examples of a low, average, and high performing Soldier on each dimension. Ratings on several of the individual scales were combined to form four PRS composites and three scales were left as single-item dimensions. These scores were also combined into an Overall Performance composite. Confirmatory factor analyses (CFA) results supported an overall performance dimension (χ^2 (*df*) = 12.17 (5); RMSEA = .032; CFI = .999, TLI = .997, SRMR = .007)[2]. Cronbach's alpha coefficients for the PRS composite scales are reported in Table 6. The PRS also includes a four-point familiarity item to ensure that the supervisor was familiar enough with each Soldier to rate their performance. Supervisors who indicated that they

Table 8. Base rates for attrition criteria for regular army soldiers.

Cumulative Attrition	n	n_{Attrit}	$\%_{Attrit}$
6-Month	227,990	24,168	10.6
12-Month	202,838	26,948	13.3
24-Month	161,493	31,255	19.4

n = number of Soldiers with attrition data at the time data were extracted. n_{Attrit} = number of Soldiers who attrited at the specified months of service. $\%_{Attrit}$ = percentage of Soldiers who attrited through the specified months of service [(n_{Attrit}/*n*) x 100].

were not familiar with a Soldier's performance were removed from the analyses. Most Soldiers were rated by only one supervisor, so interrater reliability estimates were not calculated.

Descriptive statistics for the criteria described above are summarized in Table 7.

Administrative criteria. In addition to information gathered from Soldiers and supervisors directly, ARI also obtained administrative records for each Soldier. The most valuable of these administrative records is attrition. Attrition is a broad category that encompasses involuntary and voluntary separations for a variety of reasons (e.g., underage enlistment, conduct, family concerns, drugs or alcohol, performance, physical standards or weight, mental disorder, or violations of the Uniform Code of Military Justice [UCMJ]). The reason for separation was determined by the Soldiers' Separation Program Designator (SPD) code. Soldiers who 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) were excluded from our analyses. Separation data are reported for Regular Army Soldiers only. The current analyses cover attrition through 36 months of service. Table 8 presents the basic descriptive statistics for attrition.

Results

This section is composed of two parts. The first examines the relationship between TAPAS composites and in-unit criteria via simple bivariate correlational analyses; the second. the incremental validity that TAPAS could provide over and above AFQT scores alone.

Correlations

Table 9 presents the correlations among AFQT, TAPAS composite scores, and in-unit outcome measures. The three TAPAS composites show low to moderate correlations with AFQT. The relationship with AFQT and the Can-Do composite is the highest (r = .39), which is expected because they both measure aspects achievement and are intended to measure technical competencies. The Will-Do composite showed no relationship with AFQT, and Adaptation showed a small but significant relationship (r = .14). These results suggest that that the TAPAS composites measured different constructs than AFQT. The Can-Do composite showed a moderate correlation with the WTBD JKT (r = .24); however, the composite did not predict job knowledge as well as AFQT (r = .44). We observed low correlations (r = − .06 – .06) between the Can-Do composite and all other aspects of performance, fit, and attrition. The small relationship between the Can-Do composite and supervisor ratings of Can Do performance suggests that the supervisor ratings may have been influenced by other aspects of performance, rather than technical competencies. The Can-Do composite showed a small but significant correlation with 24-month Attrition (r = − .04). The negative correlation indicates that people with higher Can-Do composite scores attrited less. The Will-Do composite showed the highest relationships with APFT scores and supervisor ratings of Physical Fitness (r = .22, .11, respectively). The Will-Do composite showed higher correlations with Army Fit (r = .07), and supervisor ratings of Leadership Potential (r = .09) and Adjustment to Army Life (r = .08) compared to AFQT.

The Adaptation composite showed the highest relationships with the APFT scores (r = .15) and WTBD JKT (r = .10). The overlap between the Adaptation composite and the Can-do composite (r = .41) and the Will-do composite (r = .51) is reflected in the correlations among the outcome variables. The Adaptation composite showed a small but significant correlation with retention cognitions (r = − .04).

Incremental validity

Here the results are organized by criteria, with the relationship between TAPAS composites and performance areas provided first, then fit and attrition. Within each of these sub-sections we highlight the TAPAS composite scores that provided the best prediction utility over and beyond AFQT alone. The analytic approach involved testing a series of Ordinary Least Squares (OLS) regression models, where scores on each criterion measure were first regressed onto Soldiers' AFQT scores, followed by scores on the TAPAS composites. For each criterion, predictors in the second step were evaluated with respect to their incremental prediction beyond the AFQT. These results provide insight into the increased prediction quality that is achieved when the TAPAS is added to AFQT to identify Soldiers that will perform well, fit, and continue as an Army Soldier once in a unit.

Table 9. Correlation estimates between AFQT, TAPAS composite scores, and in-unit outcomes.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	AFQT																			
2	TAPAS: Can-Do	**0.39**																		
3	TAPAS: Will-Do	0.01	**0.19**																	
4	TAPAS: Adaptation	**0.14**	**0.41**	**0.51**																
5	PRS: Overall Performance	**0.09**	**0.04**	**0.09**	**0.06**															
6	WTBD JKT	**0.44**	**0.24**	0.02	**0.10**	**0.12**														
7	PRS: Can Do	**0.10**	**0.06**	**0.08**	**0.05**	**0.88**	**0.12**													
8	APFT scores	−0.02	−0.01	**0.22**	**0.15**	**0.22**	0.01	**0.15**												
9	PRS: Physical Fitness and Bearing	0.01	0.02	**0.11**	**0.08**	**0.77**	**0.07**	**0.54**	**0.37**											
10	PRS: Effort and Discipline	**0.09**	0.03	**0.06**	**0.04**	**0.90**	**0.13**	**0.78**	**0.13**	**0.58**										
11	PRS: Self-Management	**0.08**	0.02	**0.05**	**0.03**	**0.87**	**0.10**	**0.74**	**0.14**	**0.57**	**0.76**									
12	PRS: Leadership Potential	**0.07**	0.03	**0.09**	**0.05**	**0.77**	**0.11**	**0.67**	**0.22**	**0.56**	**0.68**	**0.66**								
13	PRS: Working with Others	**0.12**	**0.05**	**0.07**	**0.05**	**0.86**	**0.12**	**0.78**	**0.11**	**0.52**	**0.76**	**0.72**	**0.62**							
14	Army Fit	**−0.03**	0.00	**0.07**	0.02	**0.17**	**0.11**	**0.09**	**0.10**	**0.15**	**0.15**	**0.13**	**0.18**	**0.09**						
15	MOS Fit	**0.03**	**0.03**	**0.05**	**0.05**	**0.09**	**0.10**	**0.07**	**0.04**	**0.07**	**0.09**	**0.06**	**0.09**	**0.05**	**0.45**					
16	PRS: Adjustment to Army Life	**0.05**	0.01	**0.08**	**0.05**	**0.88**	**0.11**	**0.70**	**0.21**	**0.65**	**0.75**	**0.72**	**0.76**	**0.66**	**0.23**	**0.10**				
17	6 Month Attrition	−0.01	−0.01	−0.02	−0.02	0.00	0.01	0.00	0.01	0.01	−0.01	−0.02	0.01	−0.01	0.00	0.01	0.01			
18	12 Month Attrition	−0.01	0.00	0.01	0.00	−0.03	0.02	**−0.03**	0.01	−0.01	**−0.03**	**−0.03**	−0.02	−0.02	**−0.03**	−0.02	−0.02	**0.43**		
19	24 Month Attrition	**−0.03**	**−0.04**	0.01	−0.02	**−0.15**	**−0.05**	**−0.14**	**−0.03**	**−0.12**	**−0.13**	**−0.13**	**−0.15**	**−0.11**	**−0.07**	−0.03	**−0.14**	**0.08**	**0.25**	
20	Retention Cognition	**−0.13**	**−0.06**	0.00	**−0.04**	**0.05**	**−0.04**	0.02	**0.03**	**0.04**	**0.05**	**0.04**	**0.08**	0.01	**0.43**	**0.19**	**0.09**	0.00	0.01	0.00

n = 3,178 – 6,821. WTBD JKT = Warrior Tasks and Battle Drills Job Knowledge Test. APFT = Army Physical Fitness Test. Bolded values indicate $p < .05$.

Table 10. Incremental validity estimates for the can-do TAPAS composite over AFQT for predicting performance criteria.

In-Unit Criterion Measure	*n*	AFQT *R*	AFQT + TAPAS *R*	Δ*R*
PRS: Overall Performance	3,502	**0.08**	**0.08**	0.00
Technical Task Proficiency				
WTBD JKT	4,332	**0.44**	**0.44**	0.01
PRS: Can Do	3,508	**0.09**	**0.09**	0.00
Maintaining Physical Fitness, Strength, and Weight				
APFT Score	4,093	0.02	0.02	0.00
PRS: Physical Fitness and Bearing	3,526	0.01	**0.02**	0.01
Initiative, Persistence, and Effort				
PRS: Effort and Discipline	3,505	**0.09**	**0.09**	0.00
PRS: Self-Management	3,497	**0.08**	**0.08**	0.00
Peer/Team Member Leadership Performance				
PRS: Leadership Potential	3,453	**0.07**	**0.07**	0.00
PRS: Working with Others	3,507	**0.11**	**0.11**	0.00

WTBD JKT = Warrior Tasks and Battle Drills Job Knowledge Test. *R* = multiple correlations between the AFQT and selected TAPAS composite scales with the targeted criterion measure. Δ*R* = Increment in *R* from adding the selected TAPAS composite scale to the regression model [(AFQT + TAPAS) – AFQT Only]. Bolded values indicate $p < .05$.

Table 11. Incremental validity estimates for the Will-Do TAPAS composite over AFQT for predicting performance criteria.

In-Unit Criterion Measure	*n*	AFQT *R*	AFQT + TAPAS *R*	Δ*R*
PRS: Overall Performance	5,263	**0.09**	**0.12**	**0.03**
Technical Task Proficiency				
WTBD JKT	6,268	**0.44**	**0.44**	0.00
PRS: Can Do	5,273	**0.10**	**0.12**	**0.02**
Maintaining Physical Fitness, Strength, and Weight				
APFT Score	6,114	0.02	**0.23**	**0.21**
PRS: Physical Fitness and Bearing	5,293	0.01	**0.11**	**0.10**
Initiative, Persistence, and Effort				
PRS: Effort and Discipline	5,270	**0.09**	**0.11**	**0.02**
PRS: Self-Management	5,250	**0.08**	**0.09**	**0.01**
Peer/Team Member Leadership Performance				
PRS: Leadership Potential	5,193	**0.07**	**0.11**	**0.04**
PRS: Working with Others	5,272	**0.12**	**0.14**	**0.02**

WTBD JKT = Warrior Tasks and Battle Drills Job Knowledge Test. *R* = multiple correlations between the AFQT and selected TAPAS composite scales with the targeted criterion measure. Δ*R* = Increment in *R* from adding the selected TAPAS composite scale to the regression model [(AFQT + TAPAS) – AFQT Only]. Bolded values indicate $p < .05$.

Performance

Tables 10–12 provide the incremental validity of the Can-Do, Will-Do, and Adaptation TAPAS composites, respectively, for predicting performance criteria. The Can-Do TAPAS composite did not add significant incremental validity over and beyond AFQT for any of the performance criteria. The Will-Do TAPAS composite added significant incremental validity over and beyond the AFQT for all performance criteria except for the WTBD JKT. This lack of incremental validity over AFQT with respect to the WTBD JKT is not surprising given that the AFQT was designed to predict job knowledge, whereas the TAPAS was not designed for this purpose. The Will-Do TAPAS

Table 12. Incremental validity estimates for the adaptation TAPAS composite over AFQT for predicting performance criteria.

In-Unit Criterion Measure	*n*	AFQT *R*	AFQT + TAPAS *R*	Δ*R*
PRS: Overall Performance	5,263	**0.09**	**0.10**	**0.01**
Technical Task Proficiency				
WTBD JKT	6,268	**0.44**	**0.44**	0.00
PRS: Can Do	*5,273*	**0.10**	**0.10**	**0.01**
Maintaining Physical Fitness, Strength, and Weight				
APFT Score	6,114	0.02	**0.16**	**0.14**
PRS: Physical Fitness and Bearing	*5,293*	0.01	**0.08**	**0.06**
Initiative, Persistence, and Effort				
PRS: Effort and Discipline	5,270	**0.09**	**0.10**	**0.00**
PRS: Self-Management	5,250	**0.08**	**0.08**	0.00
Peer/Team Member Leadership Performance				
PRS: Leadership Potential	5,193	**0.07**	**0.08**	**0.01**
PRS: Working with Others	5,272	**0.12**	**0.12**	**0.00**

WTBD JKT = Warrior Tasks and Battle Drills Job Knowledge Test. PRS = Performance Rating Scales. *R* = multiple correlations between the AFQT and selected TAPAS composite scales with the targeted criterion measure. Δ*R* = Increment in *R* from adding the selected TAPAS composite scale to the regression model [(AFQT + TAPAS) – AFQT Only]. Bolded values indicate $p < .05$.

Table 13. Incremental validity estimates for the TAPAS over AFQT for predicting fit criteria.

In-Unit Criterion Measure/ Model	*n*	AFQT *R*	AFQT + TAPAS *R*	Δ*R*
Can-Do				
Army Fit	4,424	**0.04**	**0.05**	0.01
MOS Fit	4,424	0.02	0.03	0.01
PRS: Adjustment to Army Life	3,248	**0.04**	**0.04**	0.00
Will-Do				
Army Fit	6,574	**0.03**	**0.08**	**0.05**
MOS Fit	6,574	**0.03**	**0.05**	**0.03**
PRS: Adjustment to Army Life	5,001	**0.05**	**0.09**	**0.04**
Adaptation				
Army Fit	6,574	**0.03**	**0.04**	**0.01**
MOS Fit	6,574	**0.03**	**0.06**	**0.03**
PRS: Adjustment to Army Life	5,001	**0.05**	**0.06**	**0.01**

PRS = Performance Rating Scales. *R* = multiple correlations between the AFQT and selected TAPAS composite scales with the targeted criterion measure. Δ*R* = Increment in *R* from adding the selected TAPAS composite scale to the regression model [(AFQT + TAPAS) – AFQT Only]. Bolded values indicate $p < .05$.

composite, which was designed to predict more motivational aspects of performance, showed the highest incremental validity with Maintaining Physical Fitness (Δ*R* = .21 and .10, for APFT and PRS: Physical Fitness, respectively). The Will-Do composite showed small gains for the other measures of performance (Δ*R* = .01 – .04). The Adaptation TAPAS composite showed small, but significant incremental validity over and beyond AFQT for all performance criteria except for the WTBD JKT and the PRS of Self-Management. The Adaptation TAPAS composites showed the highest incremental validity with Maintaining Physical Fitness (Δ*R* = .10 and .06, for APFT and PRS: Physical Fitness, respectively) and small gains for the other measures of performance (Δ*R* = .01).

Table 14. Incremental validity estimates for the TAPAS over AFQT for predicting retention cognition.

In-Unit Criterion Measure/ Model	*n*	AFQT *R*	AFQT + TAPAS *R*	Δ*R*
Can-Do	4,424	**0.13**	**0.13**	0.00
Will-Do	6,574	**0.13**	**0.13**	0.00
Adaptation	6,574	**0.13**	**0.13**	**0.00**

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

Fit

Table 13 provides the incremental validity estimates for the Can-Do, Will-Do, and Adaptation TAPAS composites for predicting fit. The Will-Do and Adaptation TAPAS composites added significant variance to the prediction of Army Fit, MOS Fit, and the PRS of Adjustment to Army Life. The Will-Do composite showed the highest incremental validity to the prediction of Army Fit (Δ*R* = .05) and PRS of Adjustment to Army Life (Δ*R* = .04). The Adaptation TAPAS composite showed marginal gains for predicting MOS Fit (Δ*R* = .03).

Attrition

Table 14 provides the incremental validity estimates for the TAPAS composites for predicting retention cognition. The Adaptation TAPAS composites added significant variance to the prediction of Retention Cognition (Δ*R* = .002). The Can-Do and Will-Do TAPAS composites did not add to the prediction of Retention Cognition.

In addition to the OLS regression analyses of ALQ and PRS criteria, we conducted logistic regression analyses attrition at 6, 12, and 24 months. 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) from the AFQT-only to the AFQT + TAPAS composite models.

Odds ratios can be used to assess the likelihood (or odds) of a given outcome depending on change in a predictor. Specifically, for a given logistic regression model, a unique odds ratio is estimated for each predictor, and represents the amount of change in the odds of the outcome that is associated with change in the given predictor. For the present analyses, the ORs represent the amount of change in the likelihood of each outcome that can be attributed to every 1.0 change in the predictor score. Note that ORs equal to 1.0 reflect no relationship between a given predictor and outcome, ORs greater than 1.0 reflect positive relationships, and ORs between 0.0 and less than 1.0 reflect negative relationships (i.e., decreasing odds of the outcome with increasing values of the predictor). For ORs

Table 15. Incremental validity estimates for the TAPAS composite scores over AFQT for predicting cumulative attrition through 24 months of service.

Attrition Measure/ Model	OR_{AFQT} (CI)	OR_{TAPAS} (CI)	r_{pb}	Δ-2LL
6 Month	*n = 148,319–216,372*			
Can-Do				
AFQT	.814 (.800-.828)		.06	
AFQT + TAPAS	.803 (.788-.819)	1.002 (1.001–1.003)	.06	**12.53**
Will-Do				
AFQT	.818 (.806-.830)		.06	
AFQT + TAPAS	.826 (.815-.839)	.990 (.989-.991)	.08	**750.04**
Adaptation				
AFQT	.818 (.806-.830)		.06	
AFQT + TAPAS	.839 (.827-.852)	.991 (.990-.992)	.08	**606.08**
12 Month	*n = 124,391–191,856*			
Can-Do				
AFQT	.842 (.828-.856)		.05	
AFQT + TAPAS	.833 (.818-.849)	1.001 (1.000–1.002)	.05	**8.30**
Will-Do				
AFQT	.850 (.838-.861)		.05	
AFQT + TAPAS	.858 (.846-.870)	.990 (.989-.990)	.09	**887.69**
Adaptation				
AFQT	.850 (.838-.861)		.05	
AFQT + TAPAS	.873 (.861-.885)	.990 (.990-.991)	.08	**743.64**
24 Month	*n = 88,111–151,449*			
Can-Do				
AFQT	.822 (.808-.836)		.07	
AFQT + TAPAS	.816 (.800-.831)	1.001 (1.000–1.002)	.07	**4.05**
Will-Do				
AFQT	.830 (.820-.841)		.07	
AFQT + TAPAS	.838 (.827-.849)	.992 (.991-.992)	.09	**614.39**
Adaptation				
AFQT	.830 (.820-.841)		.07	
AFQT + TAPAS	.851 (.840-.862)	.991 (.991-.992)	.10	**670.08**

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, $p < .05$.

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

Table 15 presents the results of the logistic regression analyses examining attrition for 6, 12, and 24 months of service for Regular Army Soldiers. The Will-Do (.990 ≤ $OR_{Will\text{-}Do}$ ≤ .992) and Adaptation (.990 ≤ $OR_{Adaptation}$ ≤ .991) composites were negatively related to attrition at all three time points, and their respective inclusion in the models resulted in significantly better fit over the AFQT alone. Conversely, the TAPAS Can-Do composite was positively related to attrition at each of the time points for Regular Army Soldiers ($OR_{Can\text{-}Do}$ > 1.00). However, the OR lower bounds of the CIs for the Can-Do composite were very near 1.0 in each of the attrition models, and these results may not represent a true effect.

Summary

The Will-Do TAPAS composite showed the highest utility for predicting in-unit performance, fit, and continuance over and above AFQT. The Can-Do TAPAS composite did not add to the prediction of performance or fit and only added small incremental validity to the prediction of attrition. The Adaptation TAPAS composite showed some increased prediction beyond AFQT, namely for attrition, maintaining physical fitness, and MOS fit.

Discussion

Inclusion of non-cognitive measures in Army initial entry screening allows for the prediction of a broader range of valued Army outcomes than traditional cognitive ability measures and educational credential screening. The TAPAS, specifically, demonstrates the ability to predict an expanded concept of Soldier performance to include motivation, adaptability, and fit with military life. Indeed, the results of the present investigation indicate that the TAPAS composites exhibited gains in predictive validity over the AFQT for several key performance outcomes. This was true particularly for those criteria tapping motivation-based aspects of Soldier performance, such as Army Fit and Physical Fitness. Thus, TAPAS provides unique and valuable information regarding a recruit's potential success as a Soldier that is not captured elsewhere in the accession process.

While some of the incremental validity gains are small, they are very important for the Army on an aggregate level. Due to the substantial cost associated with training each new recruit, the Army incurs significant losses when Soldiers leave during training or during their first term of enlistment.

The use of non-cognitive screening has consistently shown promise for predicting important outcomes in both the military and civilian sectors (e.g., Campbell & Knapp, 2001; Judge, Higgins, Thoresen, & Barrick, 1999). The results of this investigation are in line with previous research, showing that personality is a valuable predictor of outcomes such as attrition and physical fitness (Stark et al., 2014). This investigation also expands on previous research by looking at the incremental validity of TAPAS over and above cognitive ability screening for in-unit

outcomes, as well as looking at TAPAS in an on-the-job context, showing that personality testing can predict outcomes even several years after administration. The TAPAS results presented here provide further support for the use of non-cognitive measures in selection. The advanced methodology of the TAPAS also reduces the effects of faking, which have long been a concern when using personality measures for high-stakes testing (Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001), making it better-suited for selection testing.

Additional non-cognitive screening tools is important for moving to a "whole-person" approach to assessment and allowing the Army to get a better picture of each applicant. The whole-person approach can make it possible for the Army to be more selective in its screening process, accessing only those demonstrating the potential to perform well as Soldiers, but it is also flexible enough to implement in the compensatory context for which TAPAS was originally designed. For example, because TAPAS predicts performance over and above AFQT, it may be worthwhile to treat lower-AFQT applicants who are in the top TAPAS percentiles more similar to higher-scoring AFQT applicants when it comes to bonuses and eligibility for specific jobs.

The findings presented in this paper lend support to the use of the TAPAS as a tool to identify Soldiers who will perform well in their unit, fit with the Army, and continue as active duty Soldiers. They also lend support to the whole-person applicant assessment concept. By incorporating non-cognitive measures into the assessment process, the Army is getting a more comprehensive view of each applicant and increasing the chances of finding the right Soldiers to maximize its investments and accomplish its mission.

Notes

1. Verbal Expression is a scaled combination of the Word Knowledge (WK) and Paragraph Comprehension (PC) tests.
2. RMSEA = Root mean square error of approximation, acceptable values are < .08; SRMR = Standardized root mean square residual, acceptable values are < .08; CFI = Comparative Fit Index, acceptable values are > .95; TLI = Tucker-Lewis Index, acceptable values are > .95.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Army Research Institute for the Behavioral and Social Sciences [W911NF-16-C-0056].

References

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

Bynum, B. H., & Beatty, A. S. (2014). Description and psychometric properties of criterion measures. In D. J. Knapp & K. A. LaPort (Eds.), *Tier one performance screen initial operational test and evaluation: 2012 annual report* (Tech. Rep. No. 1342) (pp. 15–31). Fort Belvoir, VA: US Army Research Institute for the Behavioral and Social Sciences.

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, *75*, 95–108. doi:10.1037/0021-9010.75.5.581

Ingerick, M., Diaz, T., & Putka, D. (2009). *Investigations into Army enlisted classification systems: Concurrent validation report* (Tech. Rep. No. 1244). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, *52*, 621–652. doi:10.1111/peps.1999.52.issue-3

Knapp, D. J., & Heffner, T. S. (Eds.) (2009). *Predicting future force performance (Army Class): End of training longitudinal validation* (Tech. Rep. No. 1257). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Heffner, T. S. (Eds.) (2010). *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Tech. Rep. No. 1267). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Tremble, T. R. (Eds.) (2007). *Concurrent validation of experimental army enlisted personnel selection and classification measures* (Tech Rep. No. 1205). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Wolters, H. M. (2017). *Tier one performance screen initial operational test and evaluation: 2014 annual report* (Tech. Rep. No. 1358). Fort Belvoir, VA: US Army Research Institute for the Behavioral and Social Sciences.

Laurence, J. H., Ramsberger, P. F., & Arabian, J. M. (1997). *Education credential tier evaluation* (Research Note No. 97–07). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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 No. 2009-11). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual

performance. *Journal of Applied Psychology*, *79*, 475–480. doi:10.1037/0021-9010.79.4.475

Stark, S., Chernyshenko, O. S., Drasgow, F., & Nye, C. D. (2017). The tailored adaptive personality assessment system (TAPAS). In D. J. Knapp & H. M. K. Wolters (Eds.), *Tier one performance screen initial operational test and evaluation: 2014 annual report* Technical Report 1358 (pp. 11–15). Fort Belvoir, VA: US Army Research Institute for the Behavioral and Social Sciences.

Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, *86*, 943–953.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise preference model. *Applied Psychological Measurement*, *29*, 184–201. doi:10.1177/0146621604273988

Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology*, *26*(3), 153–164. doi:10.1037/mil0000044

Strickland, W. J. (2005). *A longitudinal examination of first term attrition and reenlistment among FY1999 enlisted accessions* (Tech. Rep. No. 1172). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

White, L. A., Nord, R. D., Mael, F. A., & Young, M. C. (1993). The assessment of background and life experiences (ABLE). In T. Trent & J. H. Laurence (Eds.), *Adaptability screening for the armed forces* (pp. 101–162). Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel).

White, L. A., Rumsey, M. G., Mullins, H. M., Nye, C. D., & LaPort, K. A. (2014). Toward a new attrition screening paradigm: Latest army advances. *Military Psychology*, *26*(3), 138–152. doi:10.1037/mil0000047

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