Technical Report 1388

Examining Enhanced Suitability Screening for Predicting Drill Sergeant Training and Job Outcomes

Christopher D. Nye
Drasgow Consulting Group

Leonard A. White
U.S. Army Research Institute

Fritz Drasgow
Oleksandr S. Chernyshenko
Stephen Stark
Drasgow Consulting Group

Christopher R. Graves
Human Resources Research Organization

Rabiah S. Muhammad
U.S. Army Research Institute

October 2020

United States Army Research Institute for the Behavioral and Social Sciences

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

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE L. ZBYLUT, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Drasgow Consulting Group

Technical reviews by:

Kristopher Canali, U.S. Army Research Institute
Mark Young, U.S. Army Research Institute
Richard Hoffman III, U.S. Army Research Institute

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
October 2020	Final	November 2015 - May 2020

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Examining Enhanced Suitability Screening for Predicting Drill Sergeant Training and Job Outcomes	W911NF-14-C-0091
	5b. PROGRAM ELEMENT NUMBER 633007
6. AUTHOR(S) Christopher D. Nye, Leonard A. White, Fritz Drasgow, Oleksandr S. Chernyshenko, Stephen Stark, Christopher R. Graves, Rabiah S. Muhammad	5c. PROJECT NUMBER A792
	5d. TASK NUMBER 1031
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Drasgow Consulting Group 3508 N. Highcross Rd. Urbana, IL 61802	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg. 1464/Mail Stop 5586) Fort Belvoir, VA 22060	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1388

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ARI Research POC: Dr. Leonard A. White, Selection and Assignment Research Unit

14. ABSTRACT (Maximum 200 words):
The goal of this U.S. Army Research Institute research was to examine two non-cognitive assessments known as the Noncommissioned Officer Special Assignment Battery (NSAB) and the Assessment of Right Conduct (ARC) as potential predictors of performance in Drill Sergeant assignments. The initial sample consisted of 834 Drill Sergeants in the U.S. Army. The NSAB and ARC were administered to this sample during their first two weeks at the Drill Sergeant Academy (DSA) and then two waves of criterion data were collected first at the end of DSA (approximately six weeks later) and then after approximately 16-24 months on the job. These data were used to examine the validity of the NSAB and ARC for predicting Drill Sergeant outcomes. Results indicated that the both the NSAB and ARC alone have substantial validity for predicting a broad range of outcomes (multiple Rs ranging from .24 to .56) and that combining these two measures in a single regression model can improve the prediction of each outcome. These results suggest that the NSAB and ARC may be useful for identifying Soldiers with high potential to be successful in Drill Sergeant assignments.

15. SUBJECT TERMS
Enlisted Personnel Manpower Personality Measures Selection and Classification

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unclassified Unlimited	64	Tonia S. Heffner 703-545-4408

Technical Report 1388

Examining Enhanced Suitability Screening for Predicting Drill Sergeant Training and Job Outcomes

Christopher D. Nye
Drasgow Consulting Group

Leonard A. White
U.S. Army Research Institute

Fritz Drasgow, Oleksandr S. Chernyshenko, Stephen Stark
Drasgow Consulting Group

Christopher R. Graves
Human Resources Research Organization

Rabiah S. Muhammad
U.S. Army Research Institute

Selection and Assignment Research Unit
Tonia S. Heffner, Chief

October 2020

ACKNOWLEDGMENTS

The authors are especially thankful to Dr. Tonia Heffner for her insights and unwavering support. The authors are also thankful to the U.S. Army Drill Sergeant Academy and the U.S. Army Center for Initial Military Training for supporting this project, and the Soldiers who gave their time and attention to accomplish this task.

EXAMINING ENHANCED SUITABILITY SCREENING FOR PREDICTING DRILL SERGEANT TRAINING AND JOB OUTCOMES

EXECUTIVE SUMMARY

Research Requirement:

For several decades now, the Armed Services Vocational Aptitude Battery (ASVAB) has been used for the selection and classification of Soldiers in the U.S. Army. Although this cognitive test has shown substantial validity for predicting performance, there is growing evidence that other characteristics may also be useful for understanding performance and attrition in the Army (Knapp & Heffner, 2010).

To address this issue, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) was tasked with developing a measure that could be administered online and in an unproctored setting and that would predict success in special duty assignments. The result of this effort was the Noncommissioned Officer Special Assignment Battery (NSAB). The NSAB is an assessment of personality that is based on the Tailored Adaptive Personality Assessment System (TAPAS). It takes advantage of modern psychometric methods and computing technology to offer a new generation of personality measures that (a) are fake-resistant, (b) utilize computer adaptive technology to measure across a broad range of trait continua, and (c) are easily customized to meet the assessment needs of diverse military occupational specialties (MOS). As a result of these benefits, a growing body of evidence suggests that the NSAB is useful for predicting a wide variety of performance criteria (Allen, Cheng, Putka, Hunter, & White, 2010; Horgen, Nye, White, LaPort, Hoffman, Drasgow et al., 2013; Knapp & Heffner, 2012; Nye, Drasgow, Chernyshenko, Stark, Kubisiak, White, & Jose, 2012).

Although a substantial amount of work with the TAPAS, a parallel test to the NSAB, has focused on military accession testing, much less research has been conducted on the use of the NSAB for in-service testing. The goal of this research was to conduct a longitudinal validation of the NSAB and a separate measure known as the Assessment of Right Conduct (ARC) as potential predictors of performance in a sample of Drill Sergeants. Although the NSAB and the ARC have been shown to predict Soldiers' attitudes, performance, and counterproductive work behavior, both of these measures also assess different content that may contribute unique variance to the prediction of these outcomes. In other words, examining a combination of the NSAB and ARC may help to create an enhanced suitability screen (ESS) for Drill Sergeant assignments. This report describes the research done to validate these assessments for this purpose.

Procedure:

The data for this research included NSAB, ARC, and criterion data collected in three phases from an initial sample of 1,047 Drill Sergeants in the U.S. Army. Participants in this research included individuals from 20 Drill Sergeant Academy (DSA) classes. These Soldiers began their involvement as newly arrived Drill Sergeant candidates at the DSA and concluded their participation as experienced Drill Sergeants who had been on the job between 16 and 24

months. After removing potentially unmotivated responders, 834 Drill Sergeants remained for the analyses.

Data collections were conducted in three phases. Phase 1 typically occurred during the first 2 weeks of the Drill Sergeant Academy, Phase 2 occurred during the last 2 weeks of the academy (approximately 6 weeks later), and Phase 3 occurred on the job (between 16 and 24 months after Phase 2). Predictor measures were administered during Phase 1 and criterion measures were administered during Phases 2 and 3. The criteria collected for this research consisted of a broad range of outcomes including perceived fit with the Drill Sergeant role, commitment to the Army, satisfaction, resilience, leadership, and peer and supervisor ratings of performance.

With these criteria, correlation and regression analyses were used to examine the validity of the NSAB and the ARC for predicting each outcome. In addition, due to the large number of criteria measured, we also developed two composites of the outcome variables to reflect both overall training and job criteria, respectively. The predictors of these outcomes were then examined in separate regression analyses. For both training and on-the-job criteria, we first estimated the validity of the NSAB alone by regressing each outcome onto the NSAB scale scores. Next, we examined the incremental validity of the NSAB over the ASVAB General Technical (GT) scores using hierarchical regression. The ASVAB GT scores were used for these analyses because this composite is currently used to screen Drill Sergeants. Finally, we also examined the prediction of individual criteria and the overall performance composites using both the NSAB and the ARC to determine if the ARC scales could add incremental validity.

Findings:

Results showed that the NSAB scales were valid predictors of a broad range of criteria. The multiple Rs for the training outcomes ranged from .18 to .46 (sample sizes ranged from 187 to 834 depending on the outcome). The majority of the multiple Rs were greater than .30 and some of the strongest relationships were with Drill Sergeant fit, Army commitment, satisfaction, resilience, and organizational citizenship behavior (OCB). However, the strongest relationship was observed for predicting the overall training criterion composite (multiple R = .46). In addition, the NSAB scales showed incremental validity over the ASVAB GT scores for predicting these outcomes. The prediction of training outcomes also improved when both NSAB and ARC were included in the model (multiple Rs ranged from .25 to .56).

Similar results were found for predicting Drill Sergeants' attitudes and performance on the job. When predicting criteria after Drill Sergeants had been on the job for 16 to 24 months, multiple Rs ranged from .24 to .46 (N = 295). Again, the NSAB scales provided incremental validity over ASVAB GT scores for predicting these outcomes and a combination of both NSAB and ARC scales improved the prediction over either measure alone. These results suggest that the NSAB and ARC may be useful predictors of Drill Sergeants' attitudes and performance and, therefore, may be useful for identifying high potential Soldiers for Drill Sergeant assignments.

Utilization and Dissemination of Findings:

These results support the use of the NSAB and ARC as screening tools for Drill Sergeant assignments. These measures may be useful for predicting a broad range of outcomes and can add important information to existing screening tools. Importantly, the NSAB and ARC demonstrated validity and incremental validity over a period of approximately 16-24 months, suggesting that these assessments can maintain their validity over time and for predicting both training and job outcomes. The magnitudes of the relationships found in the present research add to previous research on the validity of the NSAB and suggest that this measure will be particularly useful for in-service classification and job assignment decisions. However, to support potential operational applications of these results, more research is needed to examine these findings under operational conditions.

EXAMINING ENHANCED SUITABILITY SCREENING FOR PREDICTING DRILL SERGEANT TRAINING AND JOB OUTCOMES

CONTENTS

Page

List of Tables

List of Figures

EXAMINING ENHANCED SUITABILITY SCREENING FOR PREDICTING DRILL SERGEANT TRAINING AND JOB OUTCOMES

INTRODUCTION

For several decades now, the Armed Services Vocational Aptitude Battery (ASVAB) has been used for the selection and classification of Soldiers in the U.S. Army. Although this cognitive test has shown substantial validity for predicting performance, there is also growing evidence that other characteristics may also be useful for understanding performance and attrition in the Army (Knapp & Heffner, 2010). This is consistent with the broader research literature demonstrating that non-cognitive predictors can be useful for predicting employee outcomes (Barrick & Mount, 1991; Judge, Rodell, Klinger, Simon, & Crawford, 2013).

To address this issue, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducted research from 2000 to 2005 to develop the Noncommissioned Officer Leadership Skills Inventory (NLSI). The NLSI was a paper-and-pencil battery that assessed a number of non-cognitive characteristics. This measure was developed specifically for in-service testing and research demonstrated the validity of this assessment for predicting performance in a number of special duty assignments including both Recruiters (Halstead, 2009; Horgen et al., 2006) and Drill Sergeants (Kubisiak et al., 2005). Based on these initial results, a computerized version of the NLSI began to be administered at Digital Training Facilities (DTF) to screen Soldiers for recruiting duty.

Although the NLSI showed promise in initial testing, this assessment could not be fully implemented due to the limited availability of the DTFs and the insufficient number of Soldiers testing at these facilities. Therefore, the U.S. Army Training and Doctrine Command (TRADOC) and Human Resources Command (HRC) requested that ARI develop a measure that could be administered online and in an unproctored setting. The result of this effort was the Noncommissioned Officer Special Assignment Battery (NSAB). Although the NSAB has been validated for use with Recruiters (Horgen, Nye, White, LaPort, Hoffman, Drasgow et al., 2013; Nye, Muhammad, Graves, Drasgow, Chernyshenko, Stark, & Butt, 2018) and Special Forces (Nye, Beal, Drasgow, Dressel, White, & Stark, 2014), this assessment has not yet been validated in a sample of Drill Sergeants. Therefore, the goal of the current project was to examine the validity of the NSAB for in-service testing with Drill Sergeants.

PERSONALITY ASSESSMENT FOR SOLDIER SCREENING

Given the tasks performed by Drill Sergeants, personality is likely to be a significant predictor of success in this assignment. Interest in personality as a predictor of performance has increased considerably over the past two decades. Much of this interest was galvanized by empirical evidence showing that personality constructs, such as conscientiousness, predict performance across a diverse array of civilian and military occupations (e.g., Barrick & Mount, 1991; Campbell & Knapp, 2001) and provide incremental validity beyond general cognitive ability (e.g., Schmidt & Hunter, 1998).

Despite growing interest and positive empirical support for their validity, a major concern with personality assessments is applicant faking in high-stakes settings. Past research has shown

that test takers can easily identify the correct or socially desirable responses on single statement personality measures and increase or decrease their scores when sufficiently motivated (Mueller-Hanson, Heggestad, & Thornton, 2003; White, Young, & Rumsey, 2001). As a result, faking is a potential threat to the validity of personality measures and may affect their utility in operational selection settings (White, Young, Hunter, & Rumsey, 2008).

To help address issues with faking on personality assessments, the TAPAS was developed under the Army's Small Business Innovation Research (SBIR) grant program. It takes advantage of modern psychometric methods and computing technology to offer a new generation of personality measures that (a) are fake-resistant, (b) utilize computer adaptive technology to measure across a broad range of trait continua, and (c) are easily customized to meet the assessment needs of diverse military occupational specialties (MOS). The TAPAS can measure up to 27 personality dimensions or facets. Of these, 21 cover the behavioral patterns associated with the well-known Big Five personality framework (Goldberg, 1993). The remaining six dimensions cover military-specific temperament traits (Physical Conditioning, Courage, Team-Orientation, Adventure Seeking, Situational Awareness, and Commitment to Serve). As such, the TAPAS is among the most comprehensive measures of personality facets that are currently available (Drasgow, Stark, Chernyshenko, Nye, Hulin, & White, 2012).

TAPAS tests utilize a multidimensional pairwise preference format that is designed to be resistant to faking by asking test-takers to choose the statement out of a pair of statements that best describes them. For each item, the two statements in each pair are matched on both their social desirability and extremity on the dimensions they assess. The purpose of matching statements in this way is to make identifying and selecting the most socially desirable responses more difficult for test-takers. This approach appears to work as research on the operational use of the TAPAS has found no evidence of score inflation, even when compared to other respondents taking the test for "research purposes only" (Drasgow et al., 2012). Due to the measurement approach used by the TAPAS and the broad range of facets assessed, this measure is expected to demonstrate validity even in high-stakes settings where applicants may be motivated to respond dishonestly.

In fact, a growing body of evidence suggests that the TAPAS is useful for predicting a wide variety of performance criteria. For example, research has shown that the TAPAS has validity for predicting outcomes in a Soldier's first term of enlistment including Army Physical Fitness Test (APFT) scores, disciplinary incidents, and attrition from the U.S. Army (Knapp & Heffner, 2012). This research has also demonstrated the utility of the TAPAS for predicting performance within specific MOS. Nye, Drasgow, Chernyshenko, Stark, Kubisiak, White, and Jose (2012) found adjusted (for capitalization on chance) multiple correlations of .32, .25, .36, and .24 for MOS 11B (Infantryman), 31B (Military Police), 68W (Combat Medic Specialist), and 88M (Motor Transport Operator), respectively, for predicting an overall measure of performance. Similar composites were also developed for predicting attrition and job knowledge criteria with validities ranging from .18 to .35 for the same MOS. These results suggest that the TAPAS is useful for predicting performance across a broad range of military specialties.

In addition to demonstrating validity for predicting performance, the TAPAS has also shown incremental validity over other predictors of performance in the Army. In 2006, ARI

initiated a longitudinal research project to examine the validity of non-cognitive measures for predicting Army outcomes. Results showed that the TAPAS provided significant incremental validity over the ASVAB for predicting attrition, end of training criteria, and in-unit performance (Knapp & Heffner, 2009; Knapp, Owens, & Allen, 2011). In addition, this research also showed that the TAPAS provided non-trivial gains in classification efficiency over the ASVAB alone. Additional predictive validity evidence for the TAPAS was collected during the U.S. Army's *Expanded Enlistment Eligibility Metrics* (EEEM) research project from 2007-2009 (Knapp & Heffner, 2010). The results of the EEEM effort indicated that when TAPAS trait scores were added into a regression analysis that already included AFQT, the multiple correlation increased by .26 for the prediction of physical fitness, by .16 for the prediction of disciplinary incidents, and by .20 for the prediction of 6-month attrition (Allen, Cheng, Putka, Hunter, & White, 2010). None of these criteria were predicted well by AFQT alone (predictive validity estimates were consistently below .10).

Although the TAPAS is typically used for military accession testing, the NSAB is a version of the TAPAS that is used for in-service testing. Nevertheless, despite the substantial amount of work on the validity and incremental validity of the TAPAS, less research has been conducted on the NSAB for predicting performance in special duty assignments. In particular, more research is needed to understand the potential utility of the NSAB for predicting Drill Sergeant performance. Given the validity of the NSAB in other special duty assignments (Horgen et al., 2013; Nye et al., 2014), it is likely that this assessment will predict Drill Sergeant outcomes as well. Nevertheless, the specific dimensions of the NSAB that predict performance in this assignment may differ from those found in previous research.

PURPOSE OF THE CURRENT RESEARCH

In addition to examining the utility of the NSAB for predicting Drill Sergeant performance, a second goal of this effort was to examine additional personality characteristics that might be related to Drill Sergeant success. Although past research has found promising results for the NSAB, there may be other characteristics that are not assessed by the NSAB but that may be relevant for Drill Sergeant performance. Recent research focused on identifying and developing new TAPAS scales that might be useful for predicting performance and deviance in military assignments (Nye, Drasgow, Chernyshenko, Stark, Muhammad, & Wolters, 2017). This work identified six new TAPAS dimensions: Humility, Machiavellianism, Army Self-Efficacy, Persistence, Self-Efficacy, and Virtue. These dimensions were selected by ARI (in consultation with DCG) based on previous research indicating that they could be useful for predicting work outcomes. However, these dimensions have not yet been evaluated to determine their validity for in-service testing in a sample of Drill Sergeants. Therefore, the present work examined the validity of these dimensions as well.

We also examined the Assessment of Right Conduct (ARC) to determine if the characteristics assessed by this measure could add incremental validity over the NSAB scales. The ARC was developed in 2003 to specifically predict counterproductive work behavior in U.S. Army Special Operations Forces (ARSOF; Kilcullen, White, Sanders, & Hazlett, 2003). This measure is comprised of single-statement multiple-choice questions that assess nine broad attributes thought to be related to counterproductive workplace behavior. Although not administered in a forced-choice format like the NSAB, the ARC also includes a Response

Distortion scale that is designed to detect faking. The ARC has been validated in samples of U.S. Army Correctional Specialists, ARSOF candidates, Soldiers entering Initial Military Training (IMT), and U.S. Disciplinary Barracks inmates. In each of these samples, several of the attributes assessed by the ARC were related to various criteria including job performance, disciplinary incidents, and interpersonal skills (Kilcullen et al., 2003).

Importantly, although both the NSAB and the ARC have been shown to predict Soldiers' attitudes, performance, and counterproductive work behavior, both measures also assess different content that may contribute unique variance to the prediction of these outcomes. In other words, examining the combined content in both the NSAB and ARC may help to create an enhanced suitability screen (ESS) for Drill Sergeant assignments. This report describes the research done to validate the NSAB and ARC for this purpose.

METHOD

SAMPLE

The data for this research included NSAB, ARC, and criterion data collected in three phases from an initial sample of 1,047 Drill Sergeants in the U.S. Army. Participants in this research included individuals from 20 Drill Sergeant Academy (DSA) classes. Approximately 79% of this sample (*N* = 799) were male and 69% (*N* = 677) were White. The average individual in this sample was 31 years old and had 12 years of service in the Army. The majority of individuals in this sample were in grades E-5 (20%) or E-6 (68%). In addition, 81% (*N* = 828) of the sample were Regular Army and 14% (*N* = 143) were in the Army Reserve component. These Soldiers began their involvement as newly arrived Drill Sergeant candidates at the DSA and concluded their participation as experienced Drill Sergeants who had been on the job between 16 and 24 months. A total of 315 Drill Sergeants remained in the sample for all three phases of this research.

PREDICTOR MEASURES

Noncommissioned Officer Special Assignment Battery (NSAB). A version of the NSAB that assessed 16 personality dimensions was administered to the Drill Sergeants for this project. These 16 dimensions were chosen for this research from the 32 potential TAPAS dimensions based on a job analysis of the Drill Sergeant position (Muhammad et al., 2018) and previous research with both Drill Sergeants (Kubisiak et al., 2005) and other special duty assignments (Nye et al., 2018). As noted above, five of the six new scales developed for the TAPAS (Nye et al., 2017) were also included in this version of the NSAB. The new Army Self-Efficacy scale was excluded from this research to avoid redundancy with the general Self-Efficacy scale. Table 1 provides descriptions of these 16 dimensions. These dimensions were assessed with 131 items administered in the pairwise preference format described above and respondents were instructed to indicate which statement is “more like you.” Again, the statements for each pair were matched based on both their extremity and social desirability to make the NSAB more fake-resistant (Drasgow et al., 2012). This version of the NSAB was administered as a computerized static assessment on ARI’s Verint system.

Table 1. NSAB Facets

TAPAS Facet Name	Brief Description	"Big Five" Broad Factor
Dominance	High scoring individuals are domineering, "take charge" and are often referred to by their peers as "natural leaders."	Extra-version
Sociability	High scoring individuals tend to seek out and initiate social interactions.	
Humility	High scoring individuals tend to be honest, to put others' needs before their own, and do not focus on being the center of attention.	Agree-ableness
Selflessness	High scoring individuals are generous with their time and resources.	
Achievement	High scoring individuals are seen as hard working, ambitious, confident, and resourceful.	Consci-entiousness
Order	High scoring individuals tend to organize tasks and activities and desire to maintain neat and clean surroundings.	
Non-Delinquency	High scoring individuals tend to comply with rules, customs, norms, and expectations, and they tend not to challenge authority.	
Responsibility	High scoring individuals are dependable, reliable and make every effort to keep their promises.	
Persistence	High scoring individuals tend to focus on tasks and activities until they are completed and are determined to accomplish their goals even in the face of obstacles.	
Virtue	High scoring individuals adhere to standards of honesty, morality, and "good Samaritan" behavior.	
Even Tempered	High scoring individuals tend to be calm and stable. They don't often exhibit anger, hostility, or aggression.	Emotional Stability
Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.	
Intellectual Efficiency	High scoring individuals are able to process information quickly and would be described by others as knowledgeable, astute, and intellectual.	Openness to Experience
Machiavellianism	High scoring individuals generally try to deceive and manipulate others for personal gain.	Other Facets
Physical Conditioning	High scoring individuals tend to engage in activities to maintain their physical fitness and are more likely to participate in vigorous sports or exercise.	
Self-Efficacy	High scoring individuals are confident in their skills and abilities to accomplish any task that they take on.	

Assessment of Right Conduct (ARC). The ARC was also administered to identify the potential contribution of this assessment when combined with the NSAB. The version of the ARC that was administered consisted of 77 single-statement items (i.e., items were not administered in the pairwise preference format used in the NSAB) assessing the nine dimensions shown in Table 2. Each item consisted of five response options (e.g., from 1 = *Very Often* to 5 = *Never*) but the anchors for these options varied for each question.

Table 2. Attributes Assessed by the ARC

Attribute	Definition
Social Maturity	High scoring individuals tend to be law-abiding and respectful of the rights and property of others.
Aggression	High scoring individuals tend to resort to physical force to resolve interpersonal disputes.
Empathy	High scoring individuals tend to feel sympathy and express concern for those who are experiencing misfortune.
Goal Orientation	High scoring individuals tend to set ambitious performance and career advancement goals for their career in the Army and expect to achieve these goals.
Peer Leadership	High scoring individuals tend to desire to obtain positions of authority and influence, are comfortable being in charge of a group and are willing to make tough decisions and accept responsibility for the group's performance.
Hostility to Authority	High scoring individuals tend to be expressively angered by authority figures and may actively disregard their instructions and policies.
Power	High scoring individuals tend to seek control over others for narcissistic and self-serving reasons.
Self-Efficacy	High scoring individuals have felt successful in past undertakings and expect this to continue in the future.
Work Motivation	High scoring individuals tend to give their best effort and work hard toward achieving difficult objectives.

Armed Services Vocational Aptitude Battery (ASVAB). Because of its role in the current screening of Army Drill Sergeants, we used ASVAB scores as the baseline for comparing the validity of the NSAB for predicting Drill Sergeant success. The ASVAB contains 9 subtests that assess multiple aptitudes, which are combined to create composites, and are used as the basis for current selection and classification decisions. For example, the Armed Forces Qualification Test (AFQT), which is a composite of the Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Math Knowledge subtests of the ASVAB, is used for enlistment screening. For MOS classification, the ASVAB subtests are used to form nine Aptitude Area (AA) composites that correspond to the various MOS. Candidates for Drill Sergeant assignments are screened on the General Technical (GT) composite and must receive a minimum score in order to qualify for this assignment. Therefore, we examined the incremental validity of the NSAB over individuals' GT scores.

CRITERION MEASURES

Consistent with recent ARI validation studies (Knapp & Heffner, 2012; Nye, et al., 2018; Nye et al., 2012), the criterion measures for this research included performance rating scales (PRS) provided by supervisors and peers and the Drill Sergeant Life Questionnaire (DLSQ), a self-report attitudinal measure. We developed these instruments in two versions: one for use with Drill Sergeant candidates at the DSA and the other for use with Drill Sergeants on the job at initial military training (IMT) locations. The development of these measures involved modifying and updating scales used in prior research to reflect current U.S. Army Drill Sergeant job documentation and input from U.S. Army Drill Sergeant Leaders at the DSA. We also relied

heavily on the results from a job analysis of the Drill Sergeant position (Muhammad, Rupprecht, & Graves, 2018), which was conducted to facilitate criterion development for this project.

Performance Rating Scales (PRS). The PRS developed for the current research were based on a set of rating scales used with U.S. Army Drill Sergeants in a validation of the NLSI (Kubisiak et al., 2005). To update these scales, we performed a detailed comparison of existing scale content with available documentation for the Drill Sergeant job and existing Army-wide dimensions of performance. The job documentation included content from the U.S. Army Training and Doctrine Command (TRADOC) Regulation 350-6, Initial Entry Training Policies and Administration (2013), the Drill Sergeant Handbook (Center for Army Lessons Learned, 2009), Drill Sergeant performance rubrics from previous research (Kubisiak et al., 2005), and the DSA program of instruction and evaluation forms. We also relied on the knowledge, skills, and abilities (KSAs) and over 30 critical incidents generated during the Drill Sergeant job analysis (Muhammad et al., 2018). Finally, two Drill Sergeant Leaders also performed an initial review of the existing scales to identify missing content and out-of-date requirements.

Using this information, we generated a draft set of scales that reflected current Drill Sergeant requirements as well as several Army-wide dimensions of noncommissioned officer performance. Scales for Drill Sergeant candidates and experienced Drill Sergeants were identical except that the DSA scales referenced performance in relation to their peers in the Academy while the job rating scales referenced performance in relation to other Drill Sergeants at IMT locations. We then presented the draft rating scales to two groups of three Drill Sergeant Leaders at the DSA for review. Input from these six personnel was used to refine the scales and ensure that they accurately reflect important dimensions of Drill Sergeant performance, either in training or on the job. The final set of Drill Sergeant PRS included the following 13 scales:

1. Demonstrating Technical Proficiency
2. Preparing for and Conducting Training
3. Coaching, Mentoring, and Supporting Peers/Trainees
4. Performing Counseling
5. Demonstrating Effort
6. Demonstrating Integrity
7. Maintaining Physical Fitness and Well-Being
8. Showing Consideration and Support for Peers
9. Initiating Structure and Leading Others/Trainees
10. Handling Problems
11. Displaying Tolerance
12. Performing Administrative Duties
13. Overall Drill Sergeant Performance

Drill Sergeant Life Questionnaire (DSLQ). The DSLQ is an adaptation of the Army Life Questionnaire, which is a self-report attitudinal measure currently used in ARI validation research.[1] The DSLQ includes sections on demographic, background, and experience information, as well as assessments of job satisfaction, fit, and commitment. The DSLQ also

[1] The Army Life Questionnaire was initially developed in 2005 (Van Iddekinge, Putka, & Sager, 2005) and has been updated on several occasions to meet the Army's requirements for measuring Soldier outcomes.

includes measures of organizational citizenship behavior (OCB), fit with the Drill Sergeant role, resilience, and counterproductive work behavior (CWB; e.g., waste time on the job). As with the PRS, the DSLQ was developed in two versions, one for administration to Drill Sergeant Candidates and another for administration to Drill Sergeants on the job. Drill Sergeant Leaders reviewed the DSLQ in conjunction with the PRS reviews described earlier in this section to ensure the accuracy of terminology, background, experience, and training item content. Drill Sergeant Leaders recommended numerous changes to tailor the measure for both the candidate and experienced Drill Sergeants.

PROCEDURES

Each participant completed three phases of data collection. Phase 1 typically[2] occurred during the first 2 weeks of the Drill Sergeant Academy, Phase 2 occurred during the last 2 weeks of the academy (approximately 6 weeks later), and Phase 3 occurred on the job (between 16 and 24 months after Phase 2). Predictor measures were administered during Phase 1 and criterion measures were administered during Phases 2 and 3. Criterion data were collected twice to allow for validation against training (Phase 2) and on-the-job (Phase 3) criteria.

The sessions for Phases 1 and 2 were conducted in DSA classes, each of which included approximately 80-100 Drill Sergeant candidates. Candidates were divided into four or five squads (depending on the size of the class) and were supervised (i.e., instructed) by teams of three Drill Sergeant Leaders. Phase 3 sessions were conducted at IMT locations at Fort Jackson, SC; Fort Benning, GA; Fort Leonard Wood, MO; and Fort Sill, OK.[3]

Phase 1. Phase 1 sessions, which focused exclusively on collecting predictor data, were proctored by Drill Sergeant Leaders who had been trained on the data collection procedures. The procedures for these data collections were documented in a manual that described detailed coordination, administration, and data handling protocols. To support Phase 1 data collections, project staff coordinated with the DSA leadership to obtain class rosters and provide unique participant ID numbers for each Drill Sergeant candidate. As each Phase 1 session began, proctors read a prepared script introducing the research and provided pre-assigned ID numbers to participating Drill Sergeant candidates. Phase 1 assessments were administered via ARI's secure online survey website. After each session, proctors provided project staff with a list of Drill Sergeant candidates who had participated in the sessions.

The online assessments for Phase 1 began by asking participants to read information related to the purpose of the research and sign a consent form. After electronically signing the document, participants then entered their names and participant ID numbers, which allowed us to link their responses to subsequently collected predictor and criterion data. Participants then completed a demographic and background information sheet. Next, participants read an instruction page providing detailed information about answering NSAB items before responding to the actual items. Following the NSAB, participants completed the ARC.

Phase 2. Phase 2 sessions, which were conducted toward the end of the DSA, involved collecting criterion data, including the peer and supervisor ratings and DSLQ responses. To prepare for Phase 2 data collections, project staff prepared peer and supervisor rater-assignment cards, which documented (in alphabetical order) the names and participant ID numbers for the Drill Sergeant candidates and instructors (i.e., Drill Sergeant Leaders) who were expected to participate in the Phase 2 data collection. Each Phase 2 session had two parts: one for candidates and one for supervisors.

---

[2] A few Phase 1 data collections were conducted after the first 2 weeks of the course. The interval between Phase 1 and Phase 2 collections ranged from 14 to 46 days, with an average of 33 days.

[3] These four IMT locations were selected to maximize the number of Drill Sergeants who could be assessed in a reasonable number of data collection trips.

For the candidate sessions, teams of ARI and project staff, serving as proctors, delivered a prepared script describing the purpose of the research and the specific tasks to be completed by participants. Proctors then provided the rater-assignment cards to each Drill Sergeant candidate and instructed them on how to identify the peers whose performance they should rate.[4] The process was designed to ensure each candidate received roughly the same number of peer ratings. Each candidate also filled out the DSLQ during these sessions. A total of 578 Drill Sergeant Candidates participated in the Phase 2 data collections and completed the DSLQ.

For the supervisor sessions, proctors delivered a prepared script similar to that delivered to the candidates and distributed cards listing all of the candidates that each supervisor was supposed to rate along with their participant ID numbers. Each supervisor was asked to rate the performance of all candidates in their squad. All Phase 2 assessments were administered via ARI's secure online survey website. Using these procedures, we obtained 679 performance ratings of Drill Sergeant candidates from 1,148 of their peers (multiple raters rated the same individuals). In addition, 67 supervisors provided 206 performance ratings of the Drill Sergeant candidates.

Phase 3. The collection of criterion data in Phase 3 was conducted between 16 and 24 months after the Phase 2 data collection. To recruit participants for Phase 3, ARI coordinated with the U.S. Army Center for Initial Military Training (CIMT) to have the participating IMT units alert all Drill Sergeants who had valid NSAB scores from Phase 1 of the opportunity to participate. In all, 315 Drill Sergeants participated in Phase 3 data collections.

The Phase 3 assessment included the versions of the DSLQ and peer rating scales created for experienced Drill Sergeants.[5] Preparation for Phase 3 involved constructing lists of Drill Sergeants who had valid NSAB scores at each of the four IMT locations. The lists included participants' names and ID numbers. During the sessions, participants were asked to identify (from the lists provided) up to five Drill Sergeants with whom they had worked and whose performance they could rate accurately.

Phase 3 sessions were proctored by teams consisting of an ARI researcher and two project staff members. In each session, proctors delivered a prepared script describing the purpose of the research, informed consent provisions, and the measures to be completed by the participants. Proctors also described the requirement to identify peer ratees from the list of Drill Sergeants provided. Each Drill Sergeant also filled out the DSLQ during these sessions. The Phase 3 assessments were administered via paper-and-pencil on a scannable form due to the

[4] To identify peers, Drill Sergeant candidates were given the following guidance: "Please look at the rater-assignment card you have just been given. First, underline your name and ID number. Then look at the next five names on the list. If you've worked closely enough with these five individuals to be familiar with how they do their job as a Drill Sergeant Candidate and know you can rate their performance accurately, then circle those names. If you are not confident you can rate the performance of any one or more of those candidates, scan down the list until you find a total of five candidates whose performance you can rate. The end-state is for you to have circled the names of five candidates whose performance you can rate accurately."

[5] U.S. Army Drill Sergeants are rated (i.e., supervised) by Company Commanders and Company Training Officers. Due to the difficulty of collecting ratings from these individuals at IMT locations, we were only able to collect peer ratings.

limited availability of computer facilities at IMT locations.[6] A total of 303 experienced Drill Sergeants completed the DSLQ during Phase 3. In addition, 484 experienced Drill Sergeants were rated by 297 of their peers (multiple raters rated the same individuals).

ANALYSES

Before conducting analyses, data from this sample were screened for unmotivated responding. To identify unmotivated responders, we looked for evidence of patterned responding, random responding, and large numbers of missing responses to NSAB items. For patterned responding, we counted the number of times each individual chose either the first or second response options and identified individuals who chose the same option for more than 75% of the 131 NSAB items. Because the specific dimensions assessed by each item were randomly chosen (i.e., the statements in each pair were matched based on social desirability and extremity but the dimensions they assessed were randomly chosen), the probability of selecting more than 75% of just one response option is extremely low. Therefore, individuals who chose a particular response option for more than 75% of the items were flagged as potentially unmotivated. To detect random responding, three items were included in the NSAB that asked Soldiers to select a particular response option. For example, one item might ask Soldiers to "select option A for this item." We then screened out individuals who missed at least one of these items. Finally, we also excluded individuals who did *not* respond to 10% or more of the NSAB items. This was done because large numbers of missing responses would have influenced the IRT trait estimates used to score the NSAB. Therefore, individuals with too many missing responses were excluded to ensure that everyone in the sample had a valid NSAB score. After removing potentially unmotivated responders, 834 of the Drill Sergeants from the Phase 1 data collection remained in the sample. Not surprisingly, the Soldiers who were screened out for potentially unmotivated responses scored significantly lower on the TAPAS facets associated with conscientiousness such as Achievement ($p < .05$, $d = .29$) and Virtue ($p < .05$, $d = .52$) but also on the TAPAS Physical Conditioning facet ($p < .05$, $d = .21$). The Soldiers who were screened out did not differ significantly on sex, age, or grade. This reduced sample of Drill Sergeants was used for all subsequent analyses in this report.

Next, we examined the relationships between the NSAB scales and both training and job criteria using correlation and regression analyses. These analyses were conducted separately for each of the outcomes. In addition, past research has consistently shown that performance ratings in the U.S. military are highly unidimensional (Horgen et al., 2014; Nye et al., 2012) and assess a single overall dimension of performance. Therefore, consistent with this research, we treated the performance ratings as unidimensional indicators of performance rather than examining individual dimensions (e.g., Demonstrating Technical Proficiency, Demonstrating Effort). However, due to differences between peer and supervisor ratings in past research (Horgen et al., 2014; Nye et al., 2018), we examined these two sources of performance ratings separately.

[6] Due to an alignment issue on the Phase 3 assessment forms, data were entered by hand. The data entry procedure included quality checks in which we (a) confirmed the accuracy of all participants' names and ID numbers (raters and ratees), (b) checked for out-of-range values, and (c) reentered 9% of the cases. For the 27 cases reentered, we documented an error rate of 0.35% (18 errors out of ~5200 data points), which we estimated to be less than the analysis error rate and unlikely to change the conclusions of the study.

In addition to examining individual outcomes, we also examined the prediction of two overall criterion composites—one for the training criteria assessed in the DSA and another for the criteria assessed on the job. To calculate these composites, scores for each criterion were first standardized to account for differences in their standard deviations and then summed using unit weights to create either an overall training criterion composite or an overall job criterion composite. Negatively worded scales (i.e., stress, disciplinary incidents, and CWB) were reverse coded before calculating the overall composite scores so that all scales were in a consistent direction. The goal of combining scales in this way was to determine the utility of the NSAB for predicting broader criterion variables and, therefore, for Drill Sergeant selection. All analyses were conducted separately for job and training criteria to identify any differences in the non-cognitive predictors of these criteria.

VALIDATION: RESULTS

Tables 3 and 4 show the descriptive statistics for the NSAB and ARC scales, respectively, collected at the DSA. Table 3 shows the raw scores for the NSAB scales. These scores are in the IRT theta metric and typically range from approximately -2.30 to 2.30. As shown in this table, individuals tended to score highest on the Achievement and Virtue scales and lowest on Intellectual Efficiency. Table 4 reports results for the ARC scales. These scores are also in the raw score metric and can range from 1.00 to 5.00. The NSAB and ARC scores reflected in Tables 3 and 4 were the same scores used to predict both training and job outcomes in subsequent analyses. The intercorrelations between the NSAB and ARC scales are shown in Table 5. As shown in this table, there was strong convergence between many of the NSAB and ARC scales.

Table 3. Descriptive Statistics for the NSAB Dimensions

NSAB Facets	Mean	Standard Deviation
Achievement	0.14	0.58
Dominance	-0.13	0.59
Even Tempered	0.00	0.57
Humility	-0.12	0.91
Intellectual Efficiency	-0.39	0.64
Machiavellianism	-0.22	0.69
Non-Delinquency	-0.11	0.62
Optimism	0.11	0.56
Order	-0.21	0.61
Persistence	-0.22	0.61
Physical Conditioning	0.08	0.65
Responsibility	-0.26	0.58
Self-Efficacy	-0.21	0.75
Selflessness	-0.08	0.52
Sociability	-0.25	0.73
Virtue	0.23	0.76

Note: N = 834.

Table 4. Descriptive Statistics for the ARC Dimensions

ARC Facets	N	Mean	Standard Deviations
Aggression	822	2.34	.74
Empathy	823	3.63	.51
Goal Orientation	823	3.27	.66
Hostility to Authority	823	3.14	.66
Peer Leadership	823	3.56	.54
Power	823	2.72	.62
Self-Efficacy	635	3.92	.53
Social Maturity	823	3.95	.60
Work Motivation	823	3.94	.55

Note: N = 834.

Table 5. Correlations Between the NSAB and ARC Scales

	ARC Scales[a]								
NSAB Facets	Agg.	Emp.	Goal Ori.	Host.	Peer Lead.	Pow.	Self Eff.	Soc. Mat.	Work Mot.
Achievement	**-.14**	**.21**	**.21**	**-.13**	**.19**	**-.09**	**.29**	**.21**	**.33**
Dominance	-.03	**.14**	**.37**	-.02	**.40**	**.17**	**.36**	.04	**.35**
Even Tempered	**-.27**	**.10**	**-.07**	**-.22**	-.05	**-.20**	.05	**.24**	.01
Humility	**-.13**	**.07**	**-.21**	-.02	**-.21**	**-.28**	**-.15**	**.16**	**-.11**
Intellectual Efficiency	-.01	**.07**	**.18**	.01	**.30**	.00	**.29**	.05	**.23**
Machiavellian	**.25**	**-.28**	**.11**	**.16**	.02	**.33**	-.04	**-.29**	**-.10**
Non-Delinquency	**-.21**	**.12**	-.01	**-.16**	-.06	**-.17**	.03	**.35**	.06
Optimism	**-.15**	**.09**	**.12**	**-.22**	**.15**	.00	**.26**	**.11**	**.19**
Order	-.05	.04	.05	-.03	.02	.02	**.09**	**.09**	**.11**
Persistence	**-.13**	**.21**	**.13**	**-.07**	**.21**	**-.09**	**.31**	**.26**	**.31**
Physical Conditioning	-.01	**.10**	**.14**	-.06	**.10**	.02	**.19**	**.07**	**.20**
Responsibility	**-.17**	**.23**	.05	**-.14**	**.12**	**-.15**	**.23**	**.25**	**.22**
Self-Efficacy	**-.08**	**.18**	**.34**	**-.08**	**.30**	.00	**.40**	**.10**	**.35**
Selflessness	**-.27**	**.46**	.05	**-.16**	**.14**	**-.23**	**.16**	**.28**	**.21**
Sociability	-.06	**.14**	**.14**	-.06	**.25**	**.10**	**.23**	.04	**.18**
Virtue	**-.34**	**.35**	.06	**-.20**	**.14**	**-.26**	**.21**	**.40**	**.26**

Note: Sample sizes range from 635 to 823. Bold values are significant, $p < .05$. [a]Agg. = Aggression, Emp. = Empathy, Goal Ori. = Goal Orientation, Host. = Hostility, Peer Lead. = Peer Leadership, Pow. = Power, Self Eff. = Self-Efficacy, Soc. Mat. = Social Maturity, Work Mot. = Work Motivation.

PREDICTING DSA TRAINING OUTCOMES

Table 6 provides the means, standard deviations, coefficient alpha reliability estimates (on the diagonal), and intercorrelations for the peer and supervisor performance ratings and the

DSLQ scale scores collected at the end of DSA. As noted above, in addition to examining each training outcome individually, we also examined the prediction of an overall training criterion composite. Table 6 also provides the descriptive statistics and intercorrelations for this overall composite. Because each of the scales comprising the overall training composite was first standardized to account for differences in their distributions, the mean of this variable was near zero.

Table 6. Descriptive Statistics for the Training Criteria Assessed in the DSA

			Criteria[a]													
Variables	**Mean[b]**	**SD**	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**CWB**	**Dev. Sat.**	**Disc.**	**Peer**	**Sup.**	**APFT**	**OTC**
Fit (DSLQ)	4.03	0.66	*.83*													
Army Commitment (DSLQ)	4.09	0.69	**.52**	*.91*												
DS Satisfaction (DSLQ)	3.99	0.74	**.70**	**.57**	*.95*											
DS Commitment (DSLQ)	4.34	0.80	**.69**	**.58**	**.77**	*.89*										
Resilience (DSLQ)	3.61	0.38	**.46**	**.46**	**.47**	.45	*.71*									
Stress (DSLQ)	3.45	0.86	**-.13**	-.06	**-.11**	-.03	-.02	*.81*								
Organizational Citizenship Behavior (OCB) (DSLQ)	3.61	0.95	**.30**	**.27**	**.35**	**.31**	**.37**	**-.09**	*.94*							
Counterproductive Work Behavior (CWB) (DSLQ)	1.58	0.65	**-.14**	**-.18**	**-.22**	**-.21**	**-.19**	**.14**	**-.18**	*.80*						
Development Satisfaction (DSLQ)	3.97	0.75	**.51**	**.43**	**.61**	**.62**	**.41**	.00	**.24**	**-.09**	*.56*					
Disciplinary Incidents	0.20	0.61	-.01	.06	-.04	-.04	.03	-.05	.05	.01	-.01	--				
Peer Ratings	4.37	0.97	**.13**	.05	.01	.03	.04	-.06	.03	.07	.03	-.06	*.97*			
Supervisor Ratings	4.14	1.05	**.17**	**.11**	**.15**	**.16**	**.09**	**-.26**	-.01	-.02	**.18**	-.01	**.35**	*.95*		
APFT	259.80	24.62	.03	.02	-.01	-.03	.01	.05	-.01	.00	-.04	-.05	**.11**	**.19**	--	
Overall Training Criterion Composite	.01	5.93	**.77**	**.72**	**.83**	**.80**	**.61**	**-.26**	**.51**	**-.39**	**.72**	**-.11**	**.10**	**.14**	**.15**	--

Note: Sample sizes range from 132 (Supervisor Ratings) to 841 (Satisfaction). Bold values are statistically significant, $p < .05$. Alpha reliabilities are shown in italics on the diagonal. Reliabilities were not estimated for disciplinary incidents, APFT scores, or the overall training criterion composite because these outcomes were either not assessed with multi-item scales or did not reflect unidimensional latent constructs. Because each of the scales comprising the overall training composite was first standardized to account for differences in their distributions, the mean of this variable was near zero. DSLQ = Drill Sergeant Life Questionnaire.

[a] DS = Drill Sergeant; Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; Sup. = Supervisor Ratings; OTC = Overall Training Criterion.

[b] Scaling of most variables was on a 1-5 Likert scale except for Disciplinarily Incidents, which was a frequency count, APFT scores, and the Overall Training Criterion Composite.

Tables 7 and 8 show the correlations between the predictor scales and the training criteria. Table 7 shows that a number of NSAB facets were significant predictors of these criteria. For example, the Self-Efficacy facet had the strongest relationships with several training criteria and was positively related to outcomes like DS fit (r = .28), commitment (r = .29), satisfaction (r = .28), and the overall training criterion (r = .33). Similarly, Optimism, Achievement, and Dominance were related to a number of criteria as well. Other facets had moderate relationships with specific criteria. For example, Virtue was related to CWB (r = -.25) while Responsibility was more closely related to OCB (r = .26). Table 8 shows that the ARC scales were also significantly correlated with a number of criteria. Empathy, Goal Orientation, Peer Leadership, Self-Efficacy, and Work Motivation had some of the largest correlations in this sample across a broad range of criteria. Based on these results, it appears that the ARC scales may contribute to the prediction of these outcomes and may be useful to consider in combination with the NSAB.

Table 7. Correlations between the NSAB Facets and the Training Criteria Assessed in the DSA

	Criteria[a]													
Facets	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**CWB**	**Dev. Sat.**	**Disc.**	**Peer**	**Sup.**	**APFT**	**OTC**
Achievement	**.14**	**.14**	**.18**	**.18**	.07	**-.10**	**.13**	**-.12**	**.15**	-.03	.05	.07	.04	**.19**
Dominance	**.22**	**.21**	**.18**	**.18**	**.10**	**-.11**	.15	-.05	**.13**	.01	.00	.06	.05	**.24**
Even Tempered	.00	.02	.05	.01	.04	**-.11**	.06	-.04	.00	.00	-.05	.04	-.06	.03
Humility	-.08	**-.10**	**-.12**	**-.12**	**-.09**	.01	-.07	-.05	**-.14**	.05	.03	-.02	-.04	**-.11**
Intellectual Efficiency	**.10**	.05	.01	.02	.09	-.09	**.13**	-.05	-.02	-.02	.06	.05	-.03	.09
Machiavellianism	.06	.01	.03	.06	.04	-.01	**.11**	**-.15**	.04	.02	-.05	.01	-.03	.07
Non-Delinquency	.03	.06	.03	.01	.01	-.01	.09	**-.18**	-.02	-.07	-.04	.01	.03	.05
Optimism	**.21**	**.15**	**.19**	**.20**	**.17**	**-.13**	**.11**	**-.11**	**.17**	.00	-.03	.06	.00	**.27**
Order	.04	.07	**.10**	.09	.07	-.01	.05	-.08	.04	.00	-.02	-.01	.07	**.12**
Persistence	**.17**	**.16**	**.11**	.08	**.13**	-.03	**.21**	-.08	.05	.01	-.03	.09	.07	**.17**
Physical Conditioning	.08	.09	.09	.07	.08	-.05	.07	-.04	.07	-.01	-.02	.11	**.26**	**.13**
Responsibility	**.13**	**.11**	.08	.06	**.17**	**-.10**	**.26**	**-.12**	.06	.03	-.04	.03	.02	**.21**
Self-Efficacy	**.28**	**.29**	**.28**	**.26**	**.23**	-.05	**.22**	**-.10**	**.24**	.01	.08	.05	.08	**.33**
Selflessness	.04	**.12**	.08	.08	.07	-.05	**.20**	-.07	.08	.05	-.07	-.08	.02	**.12**
Sociability	**.13**	**.12**	**.12**	**.16**	.06	**-.12**	**.15**	.00	**.21**	.05	**-.16**	.02	-.05	**.16**
Virtue	**.12**	**.14**	**.14**	**.11**	**.12**	-.04	**.19**	**-.25**	.06	.05	-.03	.02	.03	**.19**

Note: Sample sizes for these analyses ranged from 167 (Supervisor Ratings) to 834 (APFT). Bold values are statistically significant, $p < .05$.

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; Sup. = Supervisor Ratings; OTC = Overall Training Criterion.

Table 8. Correlations Between the ARC Scales and the Training Criteria Assessed in the DSA

	Criteria[a]													
Scales	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**CWB**	**Dev. Sat.**	**Disc.**	**Peer**	**Sup.**	**APFT**	**OTC**
Aggression	**-.14**	**-.11**	**-.11**	**-.14**	**-.11**	.01	-.10	**.22**	**-.13**	-.06	.05	.09	-.05	**-.20**
Empathy	**.14**	**.16**	**.16**	**.14**	**.21**	-.07	**.23**	**-.19**	**.22**	.08	-.08	-.05	.04	**.25**
Goal Orientation	**.19**	**.30**	**.21**	**.18**	**.18**	-.03	**.18**	-.02	.20	-.05	-.01	.06	**.13**	**.26**
Hostility	-.10	**-.11**	-.07	**-.11**	-.09	.09	-.04	**.14**	-.05	-.02	.02	-.10	-.08	**-.12**
Peer Leadership	**.19**	**.24**	**.19**	**.17**	**.24**	-.09	**.29**	**-.11**	**.20**	.06	-.04	.03	.06	**.28**
Power	-.01	-.02	.02	-.01	-.09	.05	-.03	**.15**	.00	-.01	-.05	-.03	-.02	-.02
Self-Efficacy	**.32**	**.30**	**.33**	**.28**	**.35**	**-.18**	**.26**	**-.16**	**.32**	-.01	-.04	**.16**	.06	**.41**
Social Maturity	**.18**	.10	**.15**	**.12**	**.16**	-.05	**.18**	**-.27**	**.14**	-.02	-.06	-.10	.02	**.23**
Work Motivation	**.36**	**.30**	**.36**	**.32**	**.37**	-.07	**.32**	**-.13**	**.38**	.00	-.05	.13	.09	**.44**

Note: Sample sizes for these analyses ranged from 167 (Supervisor Ratings) to 822 (APFT). Bold values are statistically significant, $p < .05$.

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; Sup. = Supervisor Ratings; OTC = Overall Training Criteria.

Next we examined the validity of the NSAB scales using regression analysis. Table 9 shows the NSAB scales that were significant predictors of each training criterion. Again, across all criteria, there were a number of NSAB scales that were consistent predictors of these outcomes. For example, the Self-Efficacy facet was significantly related to nearly all of the outcomes examined. The Optimism and Virtue dimensions were also related to most criteria.

Although a number of NSAB facets were significantly related to various criteria, the multiple Rs indicated that a combination of facets was a much better predictor of these outcomes than was any individual dimension. The multiple Rs ranged from .18 to .39 across each of the individual criteria in this sample. Some of the strongest relationships were with attitudinal and motivational outcomes like Army commitment (R = .36), satisfaction (R = .38), and DS fit (R = .37). In addition, the NSAB facets were also important predictors of OCB (R = .39) and the overall training criterion composite (R = .46). These results indicate that the NSAB scales have utility for identifying Soldiers who will be successful in training at the DSA.

Table 9. Standardized Regression Weights for the NSAB Predicting Training Outcomes in the DSA

	Criteria[a]													
Facets	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**CWB**	**Dev. Sat.**	**Disc.[b]**	**Peer**	**Sup.**	**APFT**	**OTC**
Achievement			.06	.07	-.07	-.07			.07	-.20	**.10**			
Dominance	.09	**.13**				-.06								.08
Even Tempered	-.05			-.05		-.08			-.05			.06		-.05
Humility		-.07	-.06	-.06	-.05		-.07		-.06					
Intellectual Efficiency	-.05	**-.11**	**-.15**	**-.13**		-.05			**-.17**	-.37	.05		-.07	-.09
Machiavellianism				-.06			-.05	.08	-.05	.15			.05	
Non-Delinquency								**-.12**		**-.62**				
Optimism	**.14**	.05	.08	**.11**	**.12**	-.07		-.09	.07	-.31				**.15**
Order			.08	.07	.05			-.07						.09
Persistence	.08	.06			.05		.06	.06	-.05	.32		.09		
Physical Conditioning												.07	**.26**	.05
Responsibility					.10	-.09	**.16**							**.10**
Self-Efficacy	**.20**	**.19**	**.26**	**.24**	**.20**		**.12**	-.06	**.25**	.15	.08		.07	**.25**
Selflessness		.05					**.11**				-.05	-.11		
Sociability				.08		-.05	.08		**.17**	.28	**-.18**		-.07	.07
Virtue	.07	.08	.11	.07	.09		.06	**-.16**		.37				**.12**
Multiple R	.37	.36	.38	.37	.32	.22	.39	.32	.37	.23	.24	.18	.30	.46
Adjusted Multiple R	.32	.31	.33	.32	.25	.10	.34	.25	.31	--	.16	.00	.27	.42

Note: N = 167 (Supervisor Ratings)-834 (APFT). The values shown in this table represent the standardized regression weights in each model. Regression weights less than .05 have been removed from the table to increase clarity. Bold values are statistically significant, $p < .05$.

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; Sup. = Supervisor Ratings; OTC = Overall Training Criterion.

[b]These results are based on a logistic regression analysis. Therefore, the regression weights reported here are the unstandardized regression weights and an adjusted multiple R is not reported.

Figures 1 and 2 illustrate the practical importance of the relationships between the NSAB and several of the criteria assessed at the DSA. We used the standardized regression weights from the analyses shown in Table 9 for predicting the overall training criterion to calculate NSAB training composite scores for each individual. We then used these scores to plot the relationships between this NSAB composite and several criteria. First, Figure 1 illustrates the relationship between the NSAB composite and the overall training criterion. On the X-axis of this plot are the quintiles for the predicted scores from the NSAB composite. On the Y-axis are the overall training scores. To standardize this graph and the results in Figure 2, the outcomes were scaled to have a mean of 100 and a standard deviation of 20 and the Y-axes for these figures are scaled to range from the mean of the outcome variable +/- 1 standard deviation.

As shown in Figure 1, the NSAB was a strong predictor of the overall training criterion. Individuals scoring in the bottom 20% on the NSAB composite had an average overall training score of 83 compared to an average score of 112 in the highest scoring group. In other words, individuals in the highest scoring group on the NSAB composite performed nearly 1.5 standard deviations higher than individuals in the lowest scoring group. Looking at the distribution of the criterion, this means that individuals in the lowest scoring group on the NSAB composite scored, on average, at about the 22nd percentile on the overall training criterion. In contrast, individuals in the highest scoring group on the NSAB composite had average scores in the 70th percentile on the overall training criterion. These results provide evidence of the validity of the NSAB for predicting criteria during training at the DSA.

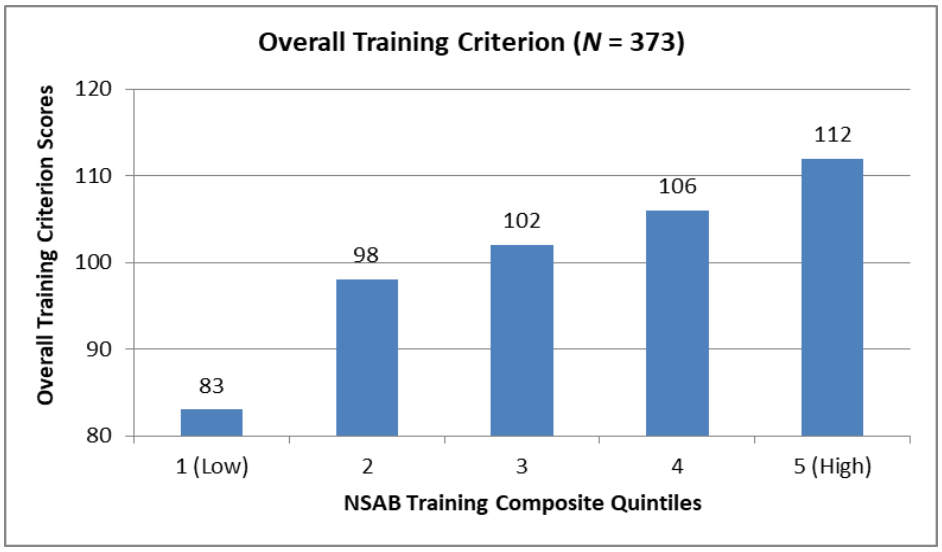

Figure 1. Quintile Plot of the Relationship between the NSAB Composite and the Overall Training Criterion

Figure 2 shows the magnitudes of the relationships between the NSAB composite and perceptions of fit with the Drill Sergeant role, satisfaction with the Drill Sergeant role, resilience, and OCB. Again, the X-axes for these plots represent the same quintiles on the NSAB training composite shown in Figure 1 and the Y-axes provide the average scores on the criteria (scaled with a mean of 100 and an SD of 20). The results indicated that individuals who scored higher on the NSAB composite had better fit with their Drill Sergeant assignment and were more satisfied

with their role, more resilient, and more likely to engage in OCB. In addition, for most of these outcomes, there was nearly a full standard deviation difference between the highest and lowest scoring groups on the NSAB composite, indicating that the effects were substantial.

Next, we also examined the incremental validity of the NSAB over individuals' ASVAB GT scores (Average GT score = 109.15, SD = 9.84). Again, ASVAB GT scores were used for these analyses because this composite is currently used to screen Drill Sergeants. Figure 3 illustrates the results from hierarchical regression analyses using both the ASVAB GT scores and all of the NSAB facets. For these analyses, the ASVAB GT score was included in Step 1 and the NSAB facets were added in Step 2. As expected, the NSAB facets contributed substantial incremental validity for all of the training criteria assessed at the DSA. For nearly all of these outcomes, the validity of GT scores alone was around .10 or below. However, after adding the NSAB scales to the regression equations, the multiple Rs increased substantially. For the majority of these criteria, adding the NSAB scales to the model increased the multiple R by more than .20. When predicting both satisfaction and OCB, the multiple R increased by nearly .30. Thus, the NSAB scales strongly contributed to the prediction of a broad range of training criteria even after accounting for GT scores.

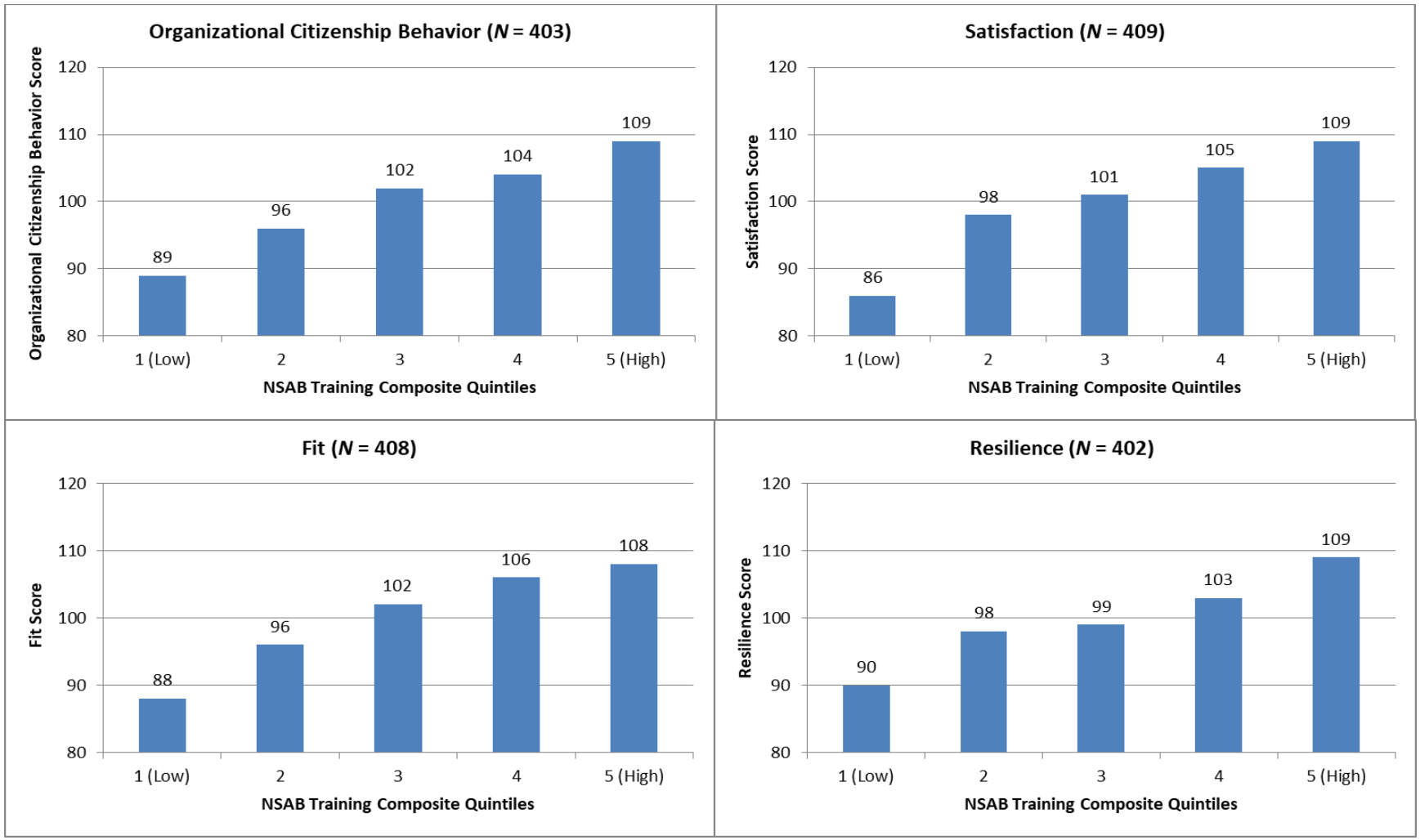

Figure 2. NSAB Training Composite Quintile Plots for OCB, Satisfaction, Resilience, and Fit Assessed in the DSA

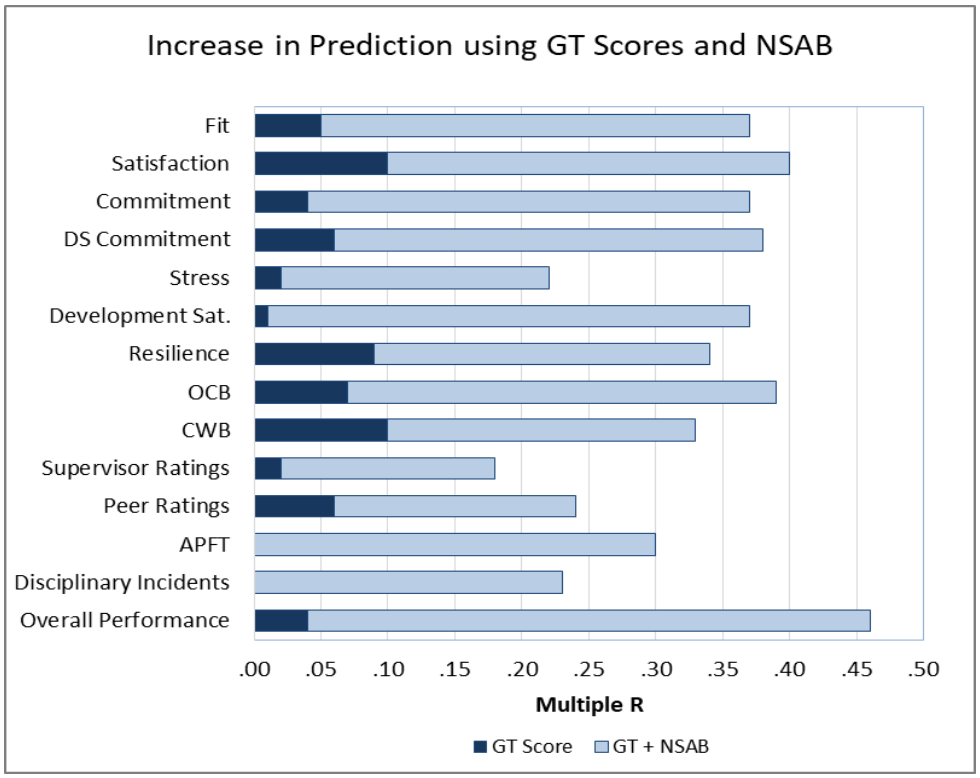

Figure 3. Incremental Validity of the NSAB over the ASVAB GT Scores for Predicting Training Outcomes

We next examined the validity of the combined NSAB and ARC scales. First, we examined the validity of the ARC alone and Table 10 shows the ARC scales that were significant predictors of each criterion. As shown in this table, a number of ARC scales were related to the criteria assessed in this project. In addition, Figure 4 shows the incremental validity of the ARC over individuals' ASVAB GT scores. As with the analyses conducted on the NSAB, the ASVAB GT scores were entered in Step 1 and the ARC scales were entered in Step 2. Similar to the NSAB results shown in Figure 3, the ARC added incremental validity over GT scores for predicting all of the training criteria assessed in the DSA. Therefore, the ARC clearly adds incremental validity over the ASVAB.

Next, we examined combined regression models with both the NSAB and the ARC included. The results of these analyses are shown in Table 11. In this table, the multiple Rs for the NSAB scales are from the model including only the NSAB. In contrast, the multiple Rs and adjusted multiple Rs in the bottom row of the table represent the overall validity when both NSAB and ARC are included in the model. As shown in Table 11, both NSAB and ARC contributed to the prediction of these training criteria. Although the NSAB and ARC provided sizeable validity alone, the regression models with both included increased prediction for each criterion. In some cases, the overall validity with both NSAB and ARC was substantial (e.g., the multiple Rs for several outcomes were above .40). Again, these results indicate that the combination of the NSAB and ARC can provide stronger prediction than either assessment alone.

Table 10. Standardized Regression Weights for the ARC Scales Predicting Training Outcomes in the DSA

	Criteria[a]													
Scales	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**CWB**	**Dev. Sat.**	**Disc.[b]**	**Peer**	**Sup.**	**APFT**	**OTC**
Aggression	-.05	-.06		-.09		**-.11**		.07	-.08	-.34		.05	-.08	-.08
Empathy	-.10	-.08	-.05	-.06				-.07		.47	-.08	-.14		
Goal Orientation	.05	**.21**	.05	.05			.06			-.19			**.14**	.07
Hostility	-.08	-.11	-.07	-.11	-.07	.12						**-.18**	**-.09**	-.10
Peer Leadership	-.11		**-.17**	-.12			**.19**	-.06	**-.14**	.59		-.16		-.08
Power		-.05	.06		-.09	.06			.05		**-.13**	-.11		.07
Self-Efficacy	.17	.11	**.21**	.14	.19	**-.29**	-.05	-.09	**.15**	-.32		**.27**	-.07	.20
Social Maturity	.06						.12	**-.18**		-.40	-.06	**-.25**		.08
Work Motivation	**.32**	.15	**.30**	**.31**	**.28**	.12	**.16**	**.05**	**.32**	-.38		.18	.09	**.29**
Multiple R	.41	.39	.41	.37	.41	.24	.37	.31	.41	.18	.13	.34	.19	.50
Adjusted Multiple R	.38	.37	.38	.34	.38	.19	.34	.27	.38	--	.00	.26	.15	.48

Note: Sample sizes for these analyses ranged from 167 (Supervisor Ratings) to 822 (APFT). The values shown in this table represent the standardized regression weights for the ARC scales. Regression weights less than +\- .05 have been removed from the table to increase clarity. Bold values are statistically significant, $p < .05$.

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; Sup. = Supervisor Ratings; OTC = Overall Training Criterion.

[b] These results are based on a logistic regression analysis. Therefore, the regression weights reported here are the unstandardized regression weights and an adjusted multiple R cannot be reported.

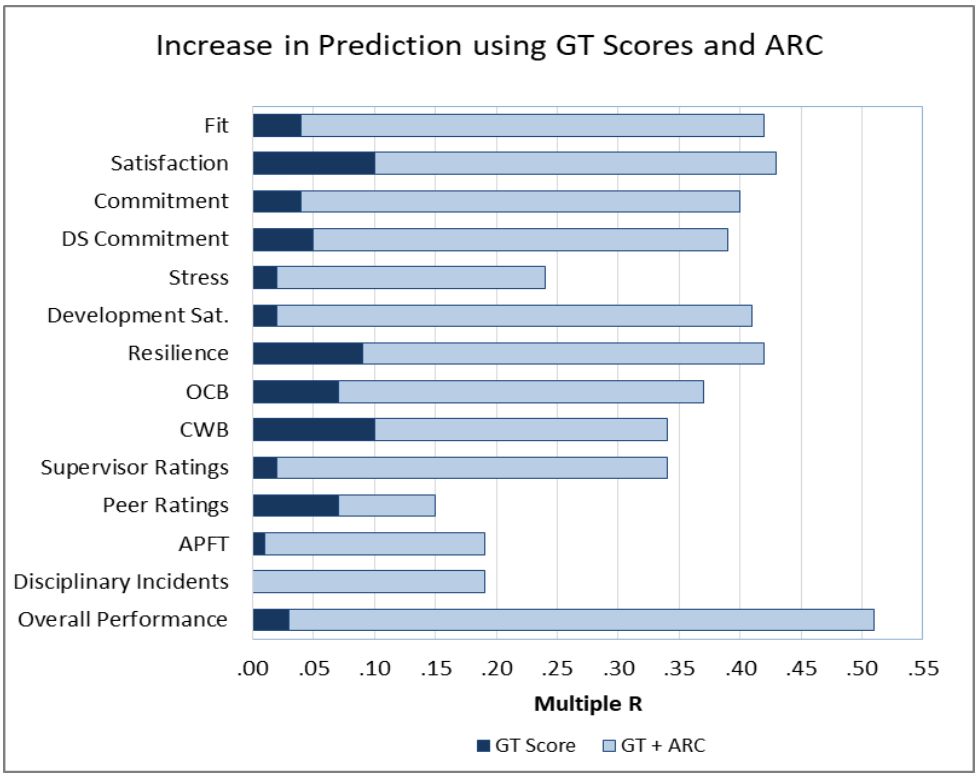

Figure 4. Incremental Validity of the ARC over the ASVAB GT Scores for Predicting Training Outcomes

Table 11. Standardized Regression Weights for the NSAB and ARC Scales Predicting Training Outcomes in the DSA

	Criteria[a]													
Variables	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**CWB**	**Dev. Sat.**	**Disc.[b]**	**Peer**	**Sup.**	**APFT**	**OTC**
Achievement	-.05			.05	-.10	-.06			.05	-.10	.09			
Dominance	.07					-.07			-.05	.09			.05	
Even Tempered	-.07			-.05		-.10		.08	-.06	.11		.06		-.05
Humility			-.05	-.05					-.06	.09			.05	
Intellectual Efficiency	-.05	**-.12**	**-.15**	**-.12**					**-.18**	-.42	.05		-.05	-.10
Machiavellianism				.06			.06	-.06		-.23			**-.11**	
Non-Delinquency					-.05			-.08	-.07	**-.55**			.06	
Optimism	**.12**		.07	.09	.09			-.08	.06	-.29				**.14**
Order			.05					-.06		-.08			.06	
Persistence	.05			-.06				.07	-.09	.43		.09	.05	
Physical Conditioning												.07	**.24**	
Responsibility					.07	-.08	**.15**							.09
Self-Efficacy	**.14**	**.17**	**.19**	**.18**	.11	.08	.08	-.05	**.18**	.24	.10			**.16**
Selflessness							.10	.08		-.17		-.11		
Sociability				.06		-.06	.05		**.13**	.26	**-.16**			
Virtue			.08		.05			-.12		.40				.07
Multiple R	.36	.36	.38	.36	.32	.22	.39	.32	.36	.23	.23	.18	.31	.46
Aggression	-.05	-.05		-.08		**-.14**		.08	-.08	-.33			-.09	-.07
Empathy	-.05	-.08	-.06	-.08				-.09		.50		-.14		
Goal Orientation		**.18**					.06			-.28			**.10**	
Hostility	-.05	-.08		-.06	-.05	.09						-.16	-.07	-.05
Peer Leadership	-.14		**-.14**	-.10	-.05		**.16**	-.09	-.13	.51		-.16		-.08
Power		-.07			-.09	.07				-.12	-.09	-.11		.06
Self-Efficacy	.09	.07	**.17**	.08	.15	**-.28**	-.09		.12	-.33		.28	-.09	.13
Social Maturity	.06	-.06					.08	**-.14**	.05	-.47	**-.06**	**-.31**	-.07	.07
Work Motivation	**.30**	.13	**.27**	**.28**	**.27**	.12	.14		**.30**	-.42		.18		**.26**
Multiple R	.46	.44	.47	.44	.45	.30	.44	.38	.48	.29	.25	.38	.34	.56
Adjusted Multiple R	.40	.37	.41	.37	.38	.17	.37	.29	.43	--	.13	.00	.28	.51

Note: Sample sizes for these analyses ranged from 167 (Supervisor Ratings) to 822 (APFT). The values shown in this table represent the standardized regression weights in each model with both NSAB and ARC included. Regression weights less than +\- .05 have been removed from the table to increase clarity. Bold values are statistically significant, $p < .05$. [a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; Sup. = Supervisor Ratings; OTC = Overall Training Criterion. [b]These results are based on a logistic regression analysis. Therefore, the regression weights reported here are the unstandardized regression weights and an adjusted multiple R cannot be reported.

Figures 5 and 6 illustrate the practical importance of the relationships shown for the combined NSAB and ARC. Consistent with the previous figures, the X-axes for these plots represent the quintiles on the combined composite of NSAB and ARC scales for predicting the overall training criterion. In addition, the Y-axes provide the average scores on each criterion (scaled with a mean of 100 and an SD of 20). As shown in Figure 5, the combined NSAB and ARC had a strong relationship with the overall training criterion. Again, these results suggest that individuals in the lowest scoring group on the NSAB composite scored, on average, in the 22nd percentile on the overall training criterion compared to the 76th percentile for individuals in the highest scoring group. For comparison with Figure 2, the relationships with the same criteria are illustrated in Figure 6 for the combined NSAB and ARC. Figure 6 shows even stronger relations than for the NSAB alone. These figures confirm the potential utility of combining both the NSAB and the ARC to predict Drill Sergeant training outcomes.

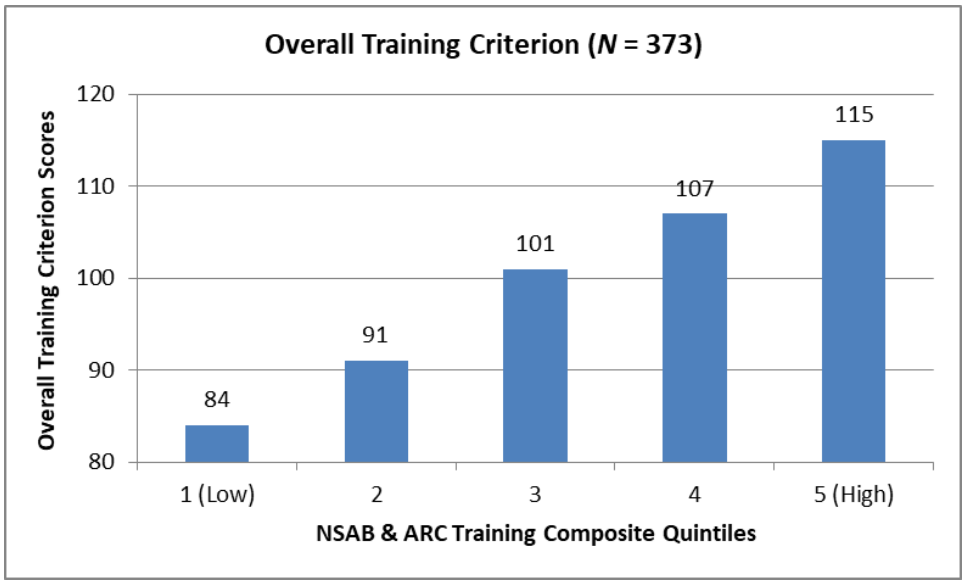

Figure 5. Quintile Plot of the Relationship between a Composite of NSAB and ARC Scales and the Overall Training Criterion

Figure 6. NSAB and ARC Training Composite Quintile Plots for OCB, Satisfaction, Fit, and Resilience Assessed in the DSA

Finally, Figure 7 shows the incremental validity of both the NSAB and the ARC over ASVAB GT scores for predicting each outcome examined at the DSA. As shown in this figure, adding the NSAB and the ARC to a regression model that included ASVAB GT scores increased the prediction over any of these measures alone. Overall, the most substantial increase in prediction was for the overall training criterion. For this outcome, the validity of the GT scores alone was only .03. However, the multiple R with all three measures included was .57. Again, these results demonstrate that a combination of both NSAB and ARC can add substantially to the prediction of criteria in the DSA, even after controlling for ASVAB GT scores.

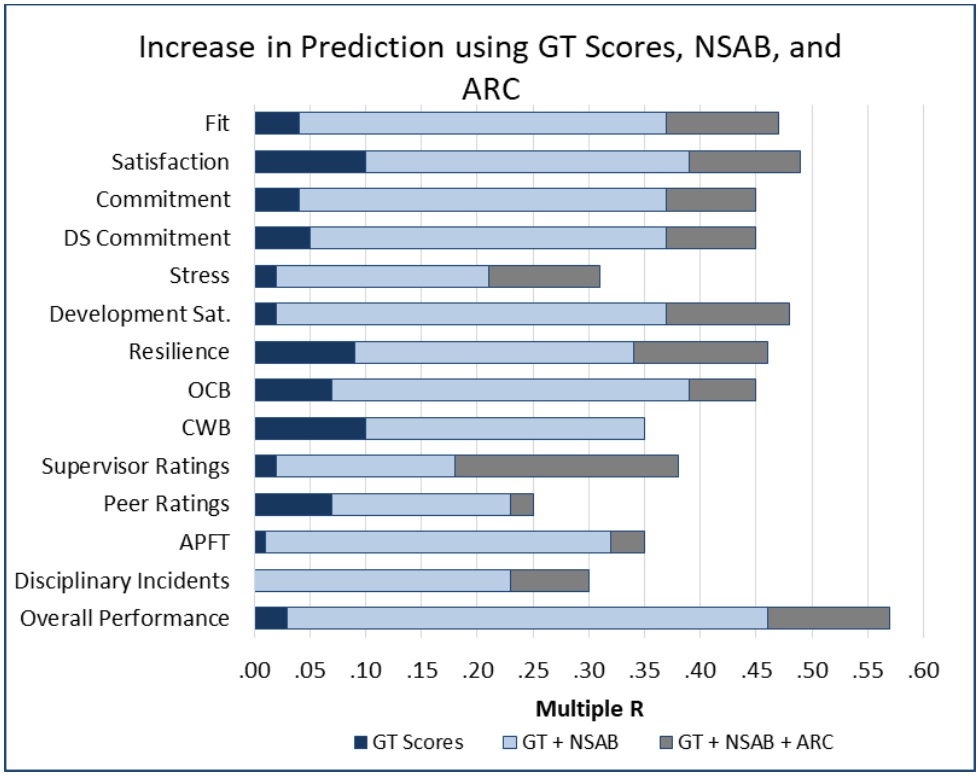

Figure 7. Incremental Validity of the NSAB and ARC over the ASVAB GT Scores for Predicting Training Outcomes

PREDICTING DRILL SERGEANT JOB OUTCOMES

Table 12 provides the means, standard deviations, coefficient alpha reliability estimates (on the diagonal), and intercorrelations for the peer performance ratings, the DSLQ scale scores, and the overall job criterion composite assessed after 16-24 months on the job. In addition, Tables 13 and 14 show the correlations of the NSAB and ARC scales with on-the-job criteria. The correlations shown in these tables indicate that some of the strongest relationships with these criteria were found for the NSAB Self-Efficacy facet. In addition, the NSAB Achievement, Dominance, Optimism, and Virtue facets also showed significant correlations with a number of criteria. Overall, the NSAB correlation results in Table 13 for on-the-job criteria were largely consistent with the results for training criteria. In addition, correlations with the ARC were also consistent across samples with several scales, such as Empathy, Goal Orientation, Peer Leadership, and Work Motivation, showing significant correlations with a number of criteria.

Table 12. Descriptive Statistics for the Criteria Assessed on the Job

			Criteria[a]													
Variables	**Mean[b]**	**SD**	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**Lead.**	**Dev. Sat.**	**Disc.**	**Medals**	**Peer**	**APFT**	**OJC**
Fit (DSLQ)	3.90	.68	*.81*													
Army Commitment (DSLQ)	4.06	.67	**.59**	*.88*												
DS Satisfaction (DSLQ)	3.64	.88	**.74**	**.61**	*.94*											
DS Commitment (DSLQ)	4.05	.91	**.68**	**.62**	**.80**	*.85*										
Resilience (DSLQ)	3.85	.43	**.51**	**.55**	**.45**	**.41**	*.82*									
Stress (DSLQ)	4.10	.83	**-.28**	**-.18**	**-.32**	**-.18**	**-.25**	*.83*								
Organizational Citizenship Behavior (OCB) (DSLQ)	3.90	.75	**.37**	**.45**	**.44**	**.37**	**.51**	**-.20**	*.88*							
Leadership Motivation (DSLQ)	3.85	.60	**.25**	**.33**	**.30**	**.28**	**.36**	-.07	**.42**	*.92*						
Development Satisfaction (DSLQ)	4.01	.85	**.60**	**.52**	**.61**	**.62**	**.37**	-.06	**.29**	**.14**	*.82*					
Disciplinary Incidents	.18	.55	**-.14**	**-.19**	**-.12**	-.11	**-.13**	.01	-.11	.00	-.13	--				
Medals	3.53	.84	-.02	.10	-.03	-.02	.11	.09	.09	**.12**	-.01	-.09	--			
Peer Ratings	4.77	.95	.10	-.01	-.02	-.02	**.12**	-.05	.05	.10	.04	-.07	-.16	*.97*		
APFT	266.81	22.16	.04	.10	.07	.03	**.16**	-.08	**.14**	**.16**	.08	.01	-.02	.06	--	
Overall Job Criterion Composite	-.07	6.66	**.78**	**.78**	**.83**	**.80**	**.71**	**-.39**	**.64**	**.49**	**.67**	**-.27**	.04	.10	**.26**	--

Note: Sample sizes range from 284 (Overall Job Criterion) to 295 (Fit). Bold values are statistically significant, $p < .05$. Alpha reliabilities are shown in italics on the diagonal. Reliabilities were not estimated for disciplinary incidents, medals, APFT scores, or the overall job criterion composite because these outcomes were either not assessed with multi-item scales or did not reflect unidimensional latent constructs. Because each of the scales comprising the overall job criterion was first standardized to account for differences in their distributions, the mean of this variable was near zero. DSLQ = Drill Sergeant Life Questionnaire.

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Lead. = Leadership Motivation; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; OJC = Overall Job Criterion.

[b] Scaling of most variables was on a 1-5 Likert scale except for Disciplinarily Incidents and Medals, which were frequency counts, APFT scores, and the Overall Job Criterion.

Table 13. Correlations between the NSAB Facets and Each Criterion Assessed on the Job

	Criteria[a]													
Facets	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**Lead.**	**Dev. Sat.**	**Disc.**	**Medals**	**Peer**	**APFT**	**OJC**
Achievement	**.17**	**.16**	**.19**	**.17**	**.17**	-.11	**.26**	**.23**	.06	-.05	-.01	.02	**.25**	**.26**
Dominance	**.20**	**.17**	**.14**	**.19**	**.15**	**-.13**	**.19**	**.18**	**.17**	.01	-.01	-.05	.08	**.25**
Even Tempered	.03	.03	.04	.02	**.13**	-.10	.07	-.01	-.02	-.11	.05	.11	-.07	.05
Humility	-.11	-.05	-.05	-.02	-.06	.04	-.09	**-.16**	**-.15**	-.02	.00	.02	-.03	-.11
Intellectual Efficiency	.05	.08	-.05	-.03	.11	-.02	**.17**	.06	-.01	-.06	.03	.01	.09	.08
Machiavellianism	.08	-.01	.05	.07	-.03	.03	.01	**-.15**	.05	-.11	.01	-.03	-.11	.01
Non-Delinquency	.05	.08	.10	.09	.02	.00	.09	.05	.06	-.08	.11	-.09	-.02	.11
Optimism	**.26**	**.21**	**.27**	**.21**	**.32**	**-.23**	**.19**	**.16**	**.25**	.06	-.06	.02	.02	**.30**
Order	.02	.10	.09	.10	.01	-.02	.10	**.15**	**.12**	-.06	-.04	-.03	**.12**	**.13**
Persistence	.09	**.15**	.09	.07	**.13**	-.09	**.16**	**.16**	.03	-.04	.06	-.10	**.16**	**.17**
Physical Conditioning	.02	.04	.04	-.01	**.12**	.03	.06	**.12**	.05	.10	-.06	-.07	**.25**	.08
Responsibility	.11	.09	**.12**	.11	.11	-.06	**.17**	.11	.06	-.01	.08	-.09	**.14**	**.15**
Self-Efficacy	**.23**	**.20**	**.20**	**.20**	**.13**	**-.12**	**.26**	**.22**	**.13**	-.03	.07	-.09	.07	**.28**
Selflessness	**.12**	**.19**	.10	**.13**	.10	-.06	**.21**	-.01	.02	-.05	**.14**	-.07	.10	**.18**
Sociability	**.22**	**.17**	**.14**	**.16**	**.16**	-.05	**.13**	.11	**.28**	-.08	.10	-.02	.00	**.22**
Virtue	**.16**	**.17**	**.21**	**.21**	**.14**	-.02	**.27**	.08	**.12**	**-.13**	**.13**	-.11	.06	**.25**

Note: Bold values are statistically significant, $p < .05$. Sample sizes for these analyses ranged from 284 (Overall Job Criterion) to 303 (Fit).

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Lead. = Leadership Motivation; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; OJC = Overall Job Criterion.

Table 14. Correlations Between the ARC Scales and Each Criterion Assessed on the Job

	Criteria[a]													
Scales	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**Lead.**	**Dev. Sat.**	**Disc.**	**Medals**	**Peer**	**APFT**	**OJC**
Aggression	-.08	-.07	-.11	**-.17**	-.03	-.06	-.08	-.01	-.09	.06	-.12	-.05	.02	**-.12**
Empathy	**.22**	**.20**	**.20**	**.26**	**.24**	-.04	**.30**	.12	**.15**	-.09	.04	-.02	.07	**.27**
Goal Orientation	**.24**	**.16**	**.17**	**.23**	**.16**	**-.13**	**.26**	**.16**	**.15**	.03	-.01	.01	**.17**	**.26**
Hostility	**-.14**	**-.22**	**-.21**	**-.21**	**-.23**	.18	-.16	-.06	-.10	.06	-.03	-.03	.04	**-.23**
Peer Leadership	**.19**	**.23**	**.17**	**.21**	**.26**	-.10	**.27**	**.25**	**.18**	.01	.12	-.09	.10	**.29**
Power	.10	**.14**	**.13**	.09	.03	-.01	.03	**.14**	**.14**	.04	.04	**-.15**	.08	**.13**
Self-Efficacy	**.31**	**.36**	**.28**	**.36**	**.37**	-.10	**.37**	**.29**	**.21**	-.09	.09	-.08	.10	**.43**
Social Maturity	.04	.02	.04	.09	.09	.12	**.18**	.11	.01	-.11	.12	-.01	-.05	.09
Work Motivation	**.25**	**.24**	**.23**	**.32**	**.29**	-.09	**.30**	**.22**	**.18**	-.05	-.01	-.03	**.15**	**.36**

Note: Bold values are statistically significant, $p < .05$. Sample sizes for these analyses ranged from 216 (Peer Ratings) to 301 (Fit).

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Lead. = Leadership Motivation; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; OJC = Overall Job Criterion.

Next, we examined the validity of the NSAB facets for predicting criteria assessed on the job using regression analysis. Table 15 shows the standardized regression weights for the NSAB facets that predicted on-the-job criteria. Although not all of the same outcomes were collected on the job and in the DSA, the results presented in Table 15 are largely consistent with the results for the training outcomes. Again, a number of NSAB facets were useful predictors of on-the-job outcomes with Optimism, Self-Efficacy, and Virtue as some of the most consistent predictors across all outcomes. The multiple Rs ranged from .24 to .46 while the adjusted multiple Rs ranged from .04 to .41. The strongest relationship was with the overall job criterion (R = .47), though the multiple Rs were .40 or greater for five other criteria as well. Overall, these results indicate that composites of the NSAB scales were the best predictors of each outcome and have utility for identifying Soldiers who will be successful Drill Sergeants.

Table 15. Standardized Regression Weights for the NSAB Facets Predicting Drill Sergeant Outcomes Assessed on the Job

	Criteria[a]													
Facets	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**Lead.**	**Dev. Sat.**	**Disc.[b]**	**Medals**	**Peer**	**APFT**	**OJC**
Achievement			.07			-.05	.11	**.12**	-.10	-.21	-.08	**.14**	**.20**	.06
Dominance	.07	.06	.05	**.13**		-.10	.05		.06	-.37	-.05			.09
Even Tempered			-.05	-.05	.08	-.08	.06		**-.12**	-.66		**.13**		
Humility	-.06						-.10	-.10	-.05	.09				-.05
Intellectual Efficiency	-.10		**-.19**	**-.19**		.07		-.07	**-.12**	.15		.07		-.10
Machiavellianism	.05	-.09			**-.11**		-.07	**-.16**	.05	-.33	-.07		**-.17**	-.07
Non-Delinquency		.06	.05	.05					.07	-.10	.07	-.08		.07
Optimism	**.18**	**.15**	**.21**	**.13**	**.28**	**-.20**	.09	.05	**.17**	**.88**	-.10			**.19**
Order				.05	**-.11**	.05		.07	.09	-.30				
Persistence	-.05		-.06	-.09		-.06		.05	-.08	-.20				
Physical Conditioning				-.05	**.11**	.06		.06		**.56**		-.06	**.21**	
Responsibility			.05					.06	.05				.07	
Self-Efficacy	**.13**	.07	**.13**	**.14**	-.05		.08	**.11**	.07	.09	.05	-.08	-.07	.11
Selflessness		**.14**			.07	-.08	.11	-.07	-.07	.12	.09		.08	.07
Sociability	**.12**	.07	.08	.10		.09			**.27**	-.41	**.14**	-.07		.10
Virtue	.10	.11	**.16**	**.15**	.11	.06	**.18**	.07	**.15**	-.38	.10	-.10		**.17**
Multiple R	.39	.35	.40	.39	.40	.29	.42	.38	.43	.37	.26	.24	.38	.47
Adjusted Multiple R	.32	.26	.33	.32	.33	.18	.36	.31	.36	--	.11	.04	.31	.41

Note: Sample sizes for these analyses ranged from 284 (Overall Job Criterion) to 295 (Fit). The values shown in this table represent the standardized regression weights in each model. Regression weights less than .05 have been removed from the table to increase clarity. Bold values are statistically significant, $p < .05$.

[a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Lead. = Leadership Motivation; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; OJC = Overall Job Criterion.

[b] These results are based on a logistic regression analysis. Therefore, the regression weights reported here are the unstandardized regression weights and an adjusted multiple R is not reported. In addition, many of these unstandardized weights are larger in magnitude than the standardized weights. However, this is due to the different metrics of the standardized and unstandardized weights so even relatively large values compared to results for other outcomes may not be statistically significant.

Figures 8 and 9 show the practical importance of the results reported in Table 15. Figure 8 shows the relationship between a composite of the NSAB facets and the overall job criterion. As with the figures illustrating the DSA results, the X-axis of this plot shows the quintiles for scores on an NSAB composite that was developed to predict the overall job criterion. On the Y-axis are the actual scores on the overall job criterion. Again, the outcomes in Figures 8 and 9 were all scaled to have a mean of 100 and a standard deviation of 20 and the Y-axes are scaled to range from the mean of the outcome variable +/- 1 standard deviation.

The results in Figure 8 indicate that the relationship between the NSAB composite and the overall job criterion had substantial practical utility. Individuals scoring in the bottom 20% on the NSAB composite scored more than a full standard deviation lower on the overall criterion than did individuals scoring in the top 20%. Looking at the distribution of the criterion, this means that individuals in the lowest scoring group on the NSAB composite scored, on average, in the 25th percentile on the overall job criterion compared to individuals in the highest scoring group on the NSAB composite who had average criterion scores in the 74th percentile. It is important to note that the individuals in this sample are the same individuals represented in both Figures 1 and 2. In addition, the NSAB composite shown in Figure 8 was correlated .90 with the NSAB Training Composite. Therefore, many of the individuals who scored well on the NSAB Training Composite also tended to score well on the NSAB Job Composite, suggesting that they will perform well both in the DSA and on the job.

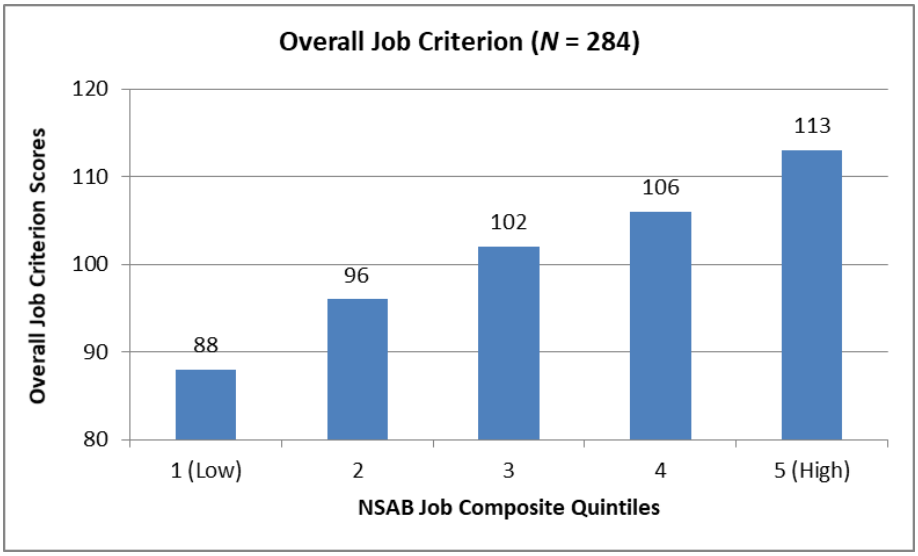

Figure 8. NSAB Composite Quintile Plot for Predicting the Overall Job Criterion

Figure 9 shows the relationships with a number of other Drill Sergeant outcomes assessed on the job. Consistent with the results in the DSA, the NSAB scales demonstrated meaningful relationships with OCB, satisfaction, fit with the DS role, and resilience. In all of these cases, individuals scoring low on the NSAB composite also tended to score nearly a full standard deviation lower on each of these outcomes compared to the highest scoring group. As with Figure 8, these individuals are the same individuals shown in Figure 2, suggesting that the NSAB is related to these outcomes both in the DSA and on the job.

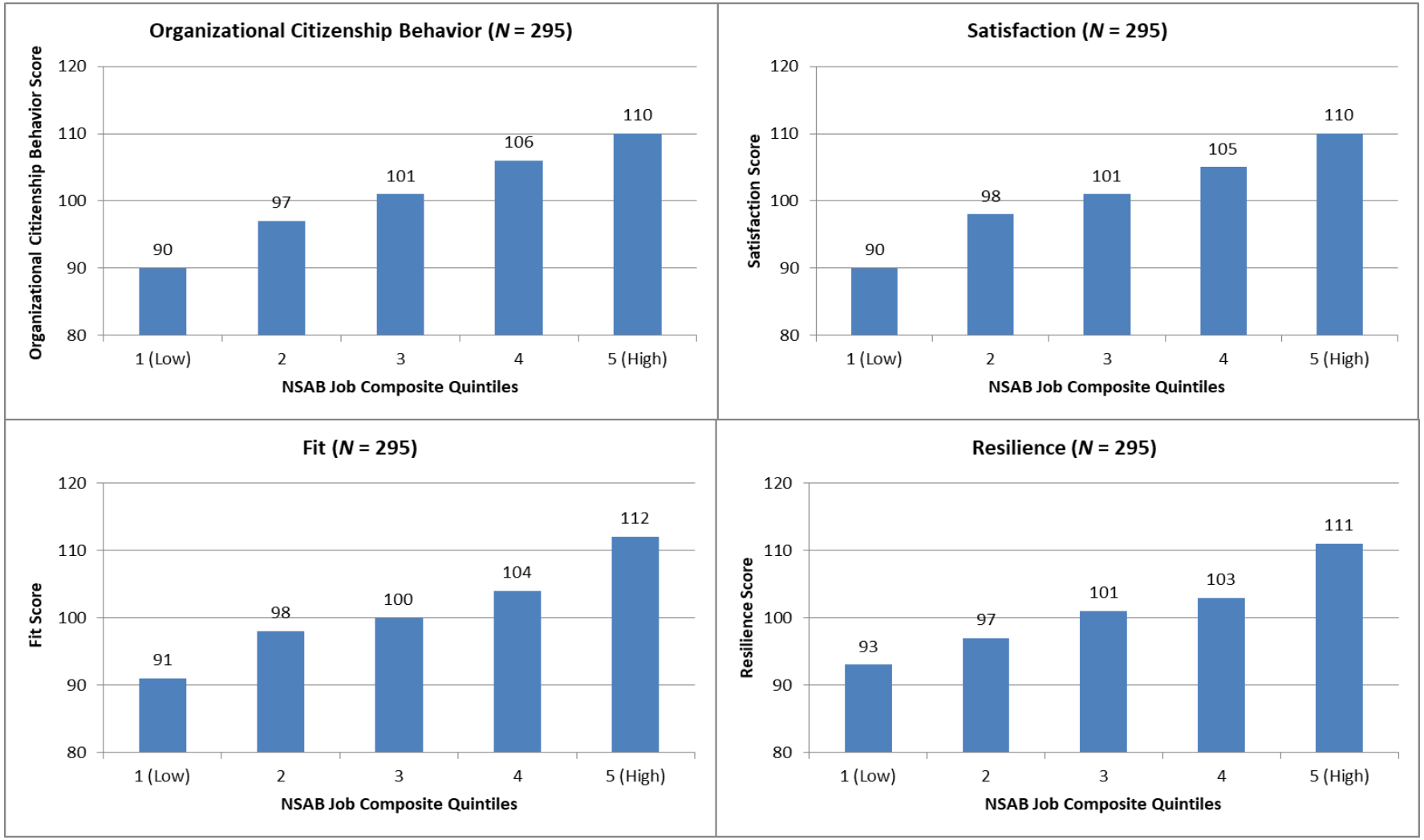

Figure 9. NSAB Composite Quintile Plots for OCB, Satisfaction, Fit, and Resilience Assessed on the Job

We also looked at the incremental validity of the NSAB facets for predicting job outcomes over the ASVAB GT composite. The results of these analyses are shown in Figure 10. Consistent with the results in the DSA, the NSAB added substantial incremental validity for predicting these outcomes. For most outcomes, adding the NSAB to a regression model that already included ASVAB GT scores increased the multiple R by .20 or .30. Therefore, the NSAB scores may be able to contribute to the prediction of Drill Sergeant outcomes on the job above and beyond existing predictors.

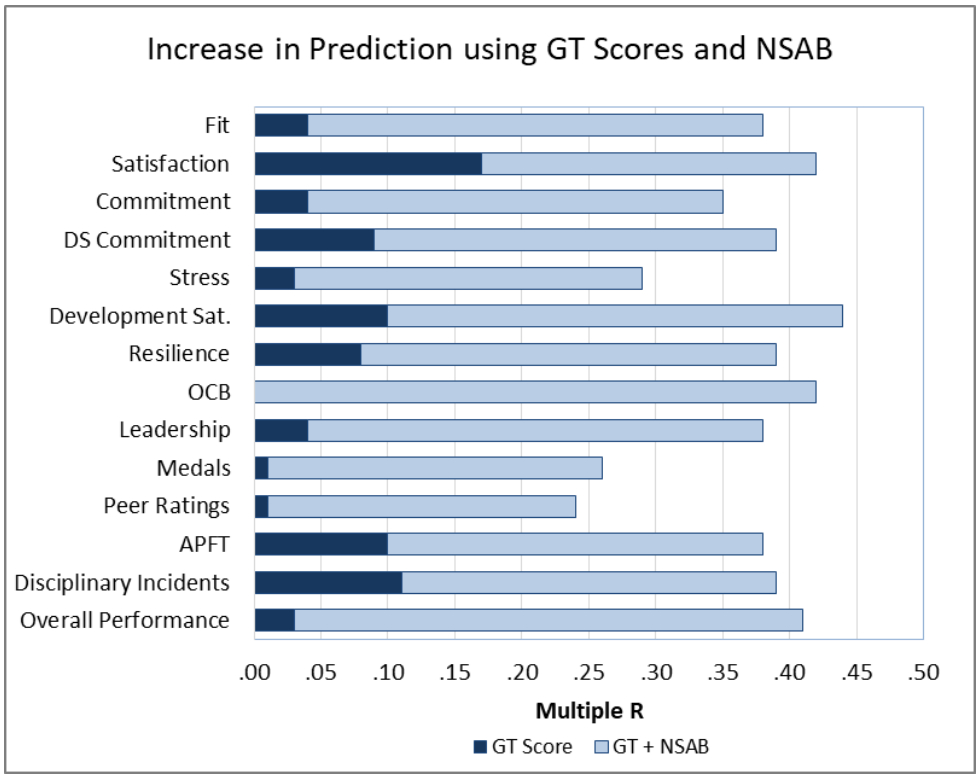

Figure 10. Incremental Validity of the NSAB for Predicting Job Outcomes over the ASVAB GT Scores

Next, we examined the combination of the NSAB and ARC for predicting the same outcomes reported in Table 15. First, we examined the validity of the ARC alone and Table 16 shows the ARC scales that were significant predictors of on-the-job outcomes. Consistent with the results in the DSA, a number of ARC scales were significant predictors of these outcomes. In particular, the Hostility scale was significantly negatively related to several of the criteria. In addition, Figure 11 shows the incremental validity of the ARC scales over the ASVAB GT scores. As with previous analyses, the ARC added incremental validity over GT scores for predicting all of the on-the-job criteria measured in this sample.

Table 17 shows the results of the regression models including both NSAB and ARC. Again, the multiple Rs for the NSAB scales are from the model including only the NSAB. In contrast, the multiple Rs and adjusted multiple Rs in the bottom rows of the table represent the overall validity when both NSAB and ARC are included in the model. Consistent with the DSA results and despite the validity of both the NSAB and ARC alone, combining both of these

measures improved the prediction even further. The combined validities of these two assessments were substantial, with validities greater than .40 for seven of the outcomes. However, there were also some differences between the combined NSAB and ARC results for predicting training and on-the-job outcomes. For example, the magnitude and frequency of the regression weights suggested that the NSAB Optimism and Virtue dimensions were more predictive on the job than in the DSA. In contrast, NSAB Self-Efficacy seemed to play a stronger role in the DSA than on the job. For the ARC, Hostility and Power were stronger predictors on the job than in the DSA while Work Motivation was more predictive in the DSA. These results suggest that Drill Sergeants' roles may change slightly as they move from training to the job. This is not surprising given that Drill Sergeants are more focused on learning at the DSA while focusing on applying this knowledge on the job. Nevertheless, these differences were often small, indicating that the differences between the roles of candidates and experienced Drill Sergeants were generally negligible.

Table 16. Standardized Regression Weights for the ARC Scales Predicting Outcomes Assessed on the Job

	Criteria[a]													
Scales	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**Lead.**	**Dev. Sat.**	**Disc.[b]**	**Medals**	**Peer**	**APFT**	**OJC**
Aggression	-.06		-.06	-.12	.13		.05		-.08	.17	-.11	-.09	-.07	
Empathy	.15	.14	.10	.09	.08		.10	-.11	.14	-.27		-.09	.08	.11
Goal Orientation	.12			.07	-.08		.11	-.05	.07	-.21		.15	.08	
Hostility	**-.15**	**-.26**	**-.23**	**-.17**	**-.28**	**.27**	**-.17**	-.11	-.12	-.82		.05		**-.23**
Peer Leadership	-.05				.07	-.18	.13	.20	.08	.84	.23	-.11	-.13	.05
Power	**.21**	**.19**	**.26**	**.23**	.10		.12	.14	**.25**		.06	**-.17**	.15	**.24**
Self-Efficacy	.21	**.27**	.05		.22		.10	.11		-.94				.17
Social Maturity		-.05	-.05			**.18**	**.15**	.14		**-1.32**	.12		-.11	
Work Motivation	-.05		.14	**.23**	.11		.05	.09		-.52	-.11		.15	.12
Multiple R	.41	.45	.42	.48	.47	.31	.47	.38	.36	.41	.25	.24	.24	.54
Adjusted Multiple R	.36	.17	.37	.45	.42	.24	.43	.33	.30	--	.16	.12	.13	.50

Note: Sample sizes for these analyses ranged from 216 (Peer Ratings) to 301 (Fit). Regression weights less than .05 have been removed from the table to increase clarity. Bold values are statistically significant, $p < .05$. [a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Lead. = Leadership Motivation; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; OJC = Overall Job Criterion. [b]These results are based on a logistic regression analysis. Therefore, the regression weights reported here are the unstandardized regression weights and an adjusted multiple R cannot be reported.

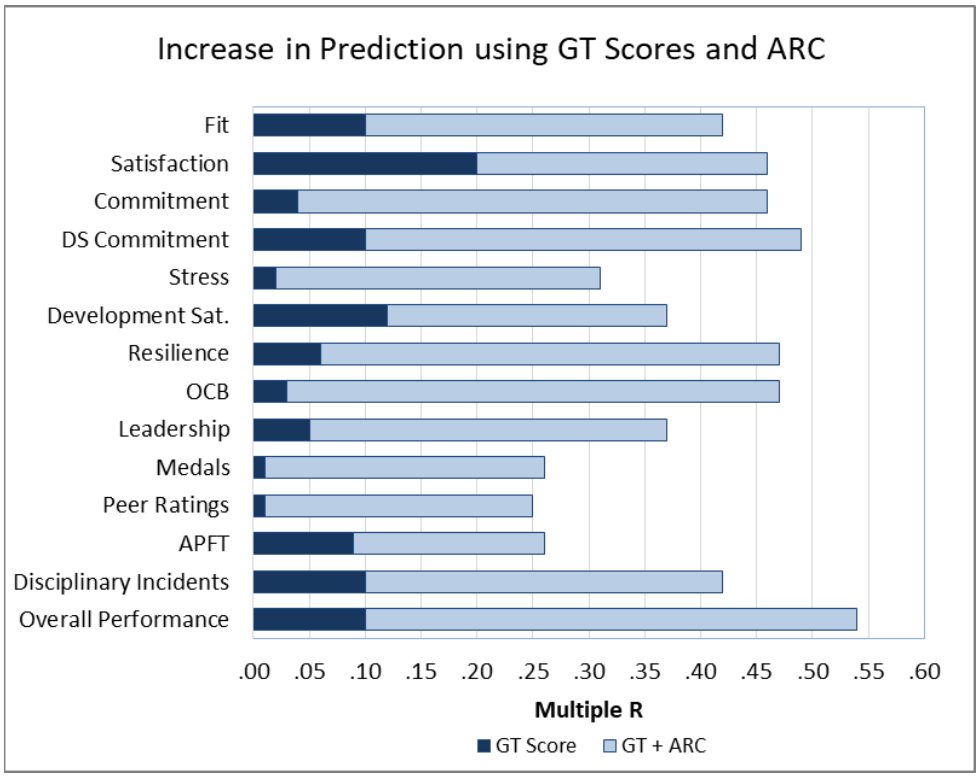

Figure 11. Incremental Validity of the ARC for Predicting Job Outcomes over ASVAB GT Scores

Table 17. Standardized Regression Weights for the NSAB and ARC Scales Predicting Outcomes Assessed on the Job

	Criteria[a]													
Variables	**Fit**	**Com.**	**Sat.**	**DS Com.**	**Res.**	**Stress**	**OCB**	**Lead.**	**Dev. Sat.**	**Disc.[b]**	**Medals**	**Peer**	**APFT**	**OJC**
Achievement						-.06	.13	.15	-.12	-.72	-.06	**.20**	**.17**	.07
Dominance		-.06		.06	-.06	-.13				.17	-.12	-.11	-.11	
Even Tempered	-.07				.05	-.06	.08		**-.16**	-.85		.15	-.07	
Humility			.09	.08			-.10	-.11		.19		-.05	-.07	
Intellectual Efficiency	-.08		**-.15**	**-.16**		.13		**-.20**	-.13	-.33		.08	.08	-.11
Machiavellianism	.07		.06	.06	-.05		-.08	-.09	.08	.54			**-.15**	
Non-Delinquency	.08	.07	.08	.09		-.06		.07	.09	**-.11**		.06	.05	.09
Optimism	**.26**	.12	**.24**	.13	**.27**	**-.20**	.12		**.25**	1.04	-.15			**.21**
Order	-.05				-.10	.05			.07	-.62		-.06	.06	
Persistence			-.06	-.11		-.05	-.06		-.10	-.51			.08	
Physical Conditioning	-.09			-.08	**.14**	.07		.10	.05	.69		.08	**.15**	
Responsibility	.06					-.05			.07	.05		-.07	.06	.05
Self-Efficacy	.06		.07		-.07		.07	.12	.08	.98	.12	-.13		**.05**
Selflessness		.09				-.06	.10		-.10	.91	.09		.12	
Sociability	.06					.09	-.05		**.18**	-.79	.14		-.05	
Virtue	.07	.13	**.17**	.08	.05		**.10**		.11		.13	**-.20**		.11
Multiple R	.44	.41	.45	.42	.47	.33	.48	.38	.49	.43	.31	.31	.38	.53
Aggression	-.05		-.05	-.09	.13		.05		-.07	-.06	-.12	-.08	-.13	
Empathy	.13	.07			.12		.07	-.08	.15	-1.06	-.06		.07	.08
Goal Orientation	.13			.07	-.08		.08	-.11	.05	-.52		.15	.07	
Hostility	-.09	**-.21**	**-.17**	**-.15**	**-.21**	**.21**	-.10	-.08	-.07	-.88		.08		**-.16**
Peer Leadership			.10		.06	-.20	.13	**.25**	.07	1.60	.22	-.06	-.13	.09
Power	**.22**	**.20**	**.29**	**.24**	.08		.10		**.17**		.07	-.17	.10	**.22**
Self-Efficacy	.11	.23		.05	.14	.09		.18	-.10	-1.54		.09		.10
Social Maturity	-.06	-.10	-.11			**.23**	.11	.13		-1.17	.11		-.17	
Work Motivation			.11	**.23**	.12	.06				-1.00	-.12			.08
Multiple R	.50	.50	.54	.55	.56	.42	.54	.47	.53	.59	.37	.36	.42	.61
Adjusted Multiple R	.40	.39	.45	.46	.47	.27	.45	.35	.43	--	.17	.09	.27	.53

Note: Sample sizes for these analyses ranged from 214 (Overall Job Criterion) to 223 (Fit). The multiple R for the NSAB alone is slightly different than in Table 15 because these results are based on only those individuals with both NSAB and ARC scores. Regression weights less than .05 have been removed from the table to increase clarity. Bold values are statistically significant, $p < .05$. [a] Fit = Drill Sergeant Fit, Com. = Army Commitment; Sat. = Drill Sergeant Satisfaction; DS Com. = Commitment to the Drill Sergeant Role; Res. = Resilience; Lead. = Leadership Motivation; Dev. Sat. = Development Satisfaction; Disc. = Disciplinary Incidents; Peer = Peer Ratings; OJC = Overall Job Criterion. [b]These results are based on a logistic regression analysis. Therefore, the regression weights reported here are the unstandardized regression weights and an adjusted multiple R cannot be reported.

Figures 12 and 13 illustrate the practical importance of the relationships for both NSAB and ARC combined. Consistent with the previous figures, the X-axes for these plots represent the quintiles for a combined composite of NSAB and ARC scales for predicting the overall job criterion. In addition, the Y-axes provide the average scores on each criterion (scaled with a mean of 100 and an SD of 20). As shown in Figure 12, the combined NSAB and ARC had a strong relationship with the overall job criterion (adjusted multiple R = .53, see Table 17). The results presented in this figure suggest that individuals in the lowest scoring group on the NSAB and ARC composite scored, on average, in the 19th percentile on the overall job criterion while individuals in the highest scoring group on the NSAB and ARC composite had average criterion scores in the 76th percentile. In addition, Figure 13 shows even stronger relationships than were observed in Figure 9 for the NSAB alone. These figures provide additional support for the potential utility of combining both the NSAB and the ARC to predict Drill Sergeant outcomes on the job.

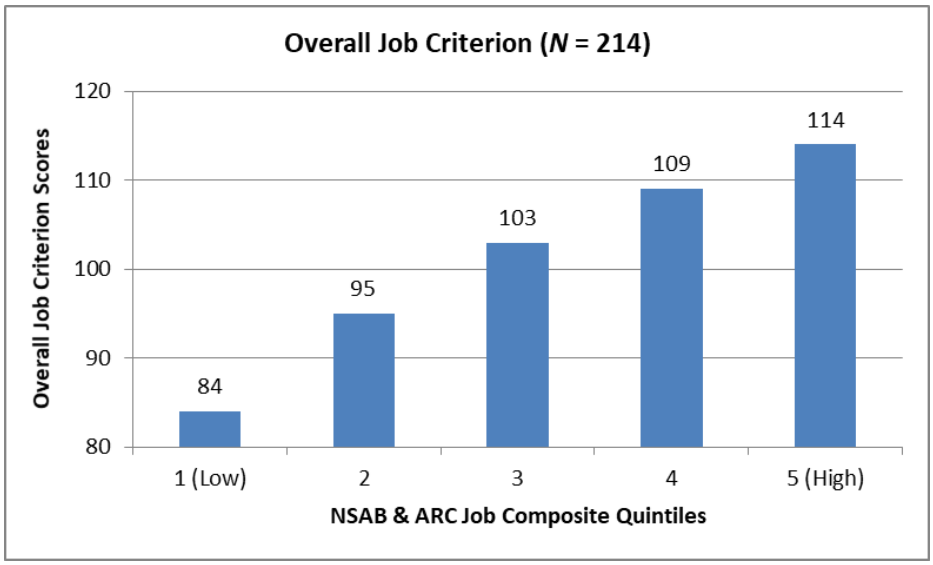

Figure 12. Quintile Plot of the Relationship between a Composite of NSAB and ARC Scales and the Overall Job Criterion

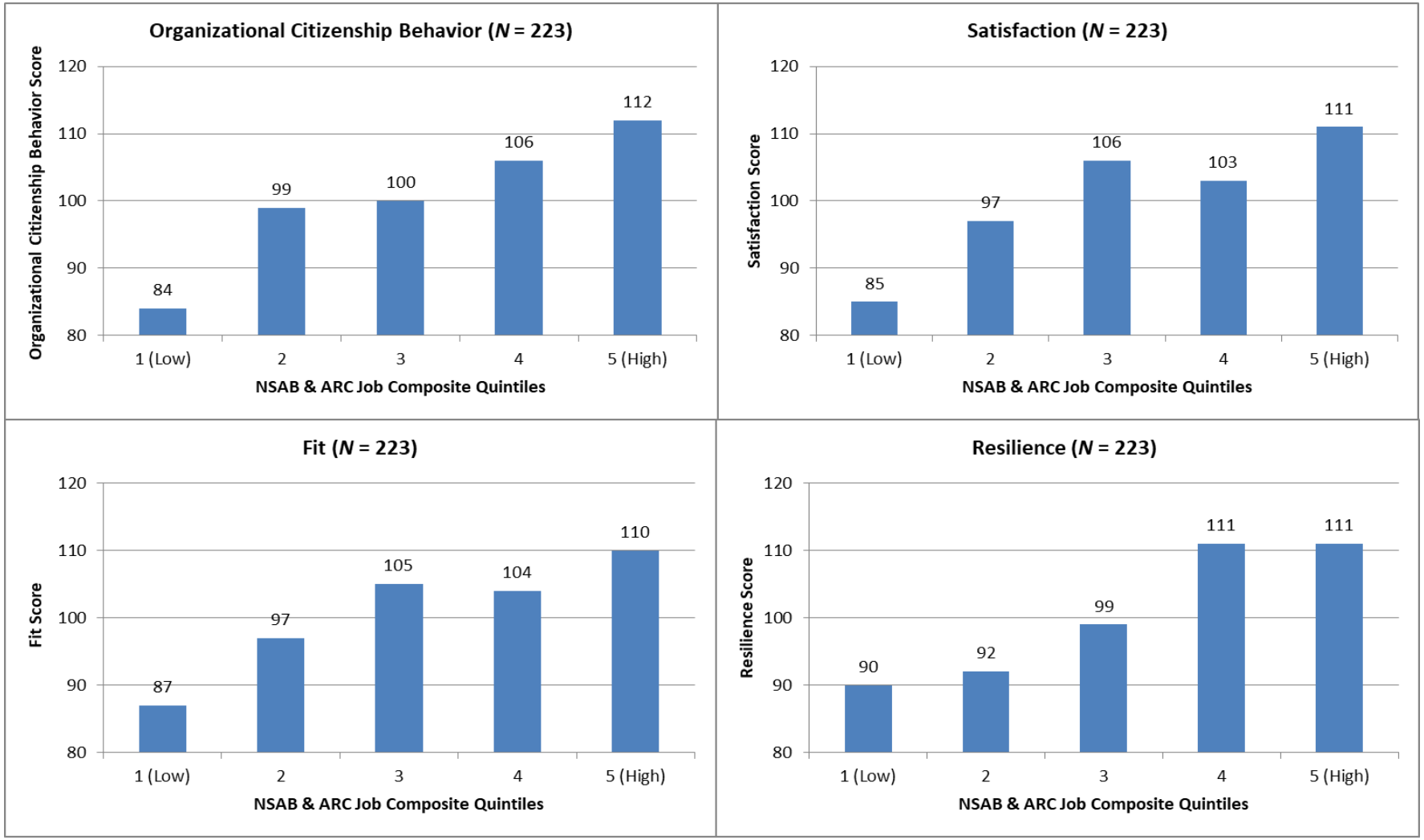

Figure 13. NSAB and ARC Composite Quintile Plots for OCB, Satisfaction, Fit, and Resilience Assessed on the Job

Finally, Figure 14 shows the incremental validity of both the NSAB and the ARC over ASVAB GT scores for predicting each outcome assessed on the job. Consistent with the results in the DSA sample, adding the NSAB and the ARC to a regression model that included ASVAB GT scores increased the prediction over any of these measures alone. In fact, for nearly all of the outcomes, the increase in prediction was substantial suggesting that both the NSAB and ARC can account for unique variance in Drill Sergeant performance on the job even after controlling for ASVAB GT scores.

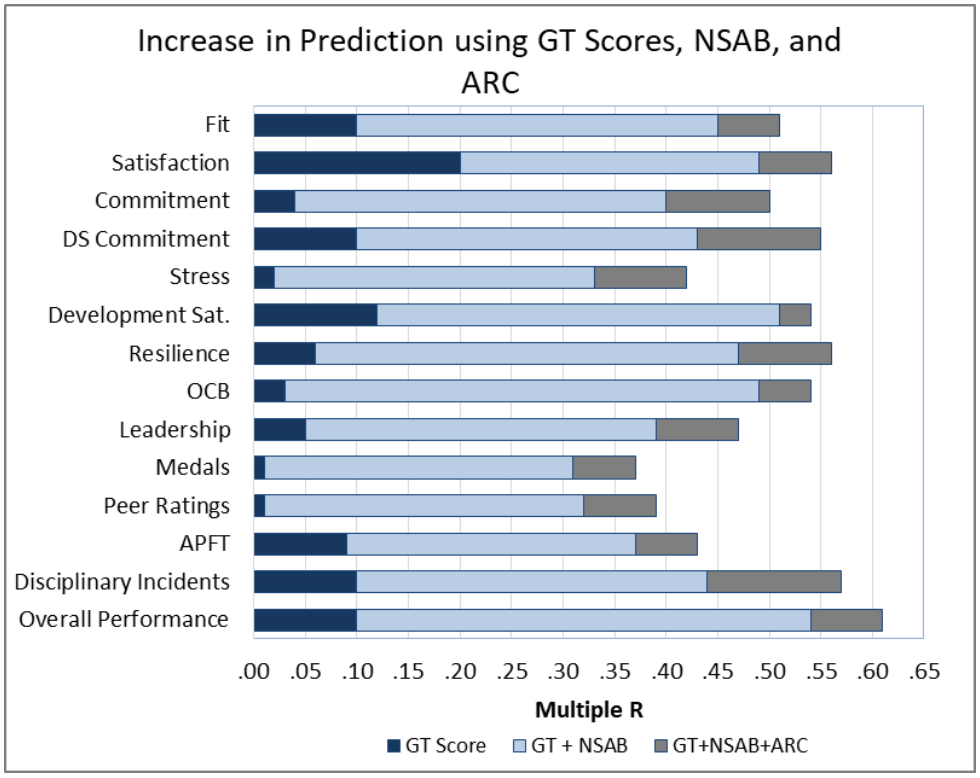

Figure 14. Incremental Validity of the NSAB and ARC for Predicting Job Outcomes over ASVAB GT Scores

CONCLUSIONS AND NEXT STEPS

The results presented here suggest that the NSAB has substantial validity for predicting a number of Drill Sergeant outcomes and can add incremental validity to the prediction of these outcomes over ASVAB GT scores. The NSAB also contributed substantially to the prediction of overall criterion composites, with multiple Rs of .46 and .47 in training and on the job, respectively. Consequently, these results suggest that NSAB composites may be useful for identifying high potential Soldiers for Drill Sergeant assignments. These results are consistent with past research demonstrating that the NSAB can be a useful predictor of important outcomes in a broad range of military occupations (Nye et al., 2012) and assignments (Horgen et al., 2014; Nye et al., 2014; Nye et al., 2018). Therefore, the results of this research expand that work and indicate that the NSAB scales are strong predictors of Drill Sergeants' attitudes and performance as well.

Similar results were found for predicting training and on-the-job outcomes. Importantly, these outcomes were assessed in the same sample of individuals over time, suggesting that NSAB scores obtained at the beginning of DSA could predict both short-term training criteria and long-term outcomes after these individuals have gained more experience on the job. The DSA criteria were collected six weeks after the NSAB was administered while on-the-job outcomes were collected approximately 16-24 months after the DSA. These results demonstrate the longitudinal validity of the NSAB.

The magnitudes of the relationships found in the present research add to previous research on the validity of the NSAB and suggest that this measure will be useful for in-service testing. In fact, the relationships between the NSAB and the outcomes assessed in this sample were generally larger than those found in other entry-level military occupations. Other research examining the prediction of similar performance criteria has generally found validities ranging from .24 to .36 across multiple jobs (Nye et al., 2012). In contrast, the multiple Rs found in the present research ranged from .24 to .46 for the NSAB alone. In addition, the multiple Rs for predicting overall criteria (both in training and on-the-job) were also around .46, which is substantially larger than in previous research. This finding is consistent with other research examining in-service testing (Horgen et al., 2013; Nye et al., 2014). In particular, these results are closely related to recent research with Recruiters, which found a multiple R closer to .55 for predicting a similar overall criterion on the job (Nye et al., 2018). These findings suggest that the NSAB may be particularly useful for selecting individuals who are already serving in the military for special duty assignments.

Despite the strong validity of the NSAB, the results also indicated that combining the NSAB and ARC could improve the prediction of Drill Sergeant outcomes over either scale alone and over ASVAB GT scores. These results suggest that although the NSAB assesses a wide range of characteristics, there are other characteristics that are not assessed well by this measure but that are related to Drill Sergeants' attitudes and performance. The ARC was developed specifically to predict counterproductive work behavior (Kilcullen et al., 2003) but the results of the present research suggest that these scales can predict broader attitudes and performance criteria as well. As a result, the combination of the ARC and the NSAB provided the best overall prediction for virtually all criteria. Again, this finding is consistent with other recent research on in-service testing for Recruiters (Nye et al., 2018).

Overall, the results suggest that the NSAB and ARC may be useful as in-service screening tools for Drill Sergeants. However, future research is necessary to provide further evidence of the validity and utility of these measures for this purpose. First, the present data were collected for research purposes only. As a result, it would be useful to examine the validity of the NSAB and ARC under high-stakes operational conditions. As noted above, a key concern with personality measures is faking in high-stakes settings. Therefore, it is important to demonstrate the validity of the NSAB even when individuals are motivated to distort their responses and inflate (or deflate) their scores. Again, the NSAB was designed specifically to be resistant to faking. The NSAB items are administered in a pairwise preference format that has demonstrated validity and negligible score inflation under operational conditions (Stark, Chernyshenko, Drasgow, Nye, White, Heffner, & Farmer, 2014). Although there is evidence that the NSAB format is fake resistant (Drasgow et al., 2012), this research has primarily examined resistance to faking good (i.e., individuals are responding in a way that makes them appear like a strong candidate) but has yet to explore conditions in which candidates may be faking bad (i.e., individuals are responding in a way that makes them look like a poor candidate). The latter situation may be more common for in-service testing where Soldiers may be motivated to distort their responses so that they are not recruited for particular assignments. To address these issues, it would be useful to examine both the NSAB and the ARC under operational conditions for in-service testing to demonstrate that these measures maintain their validity in these settings.

In addition, future research on the NSAB and ARC should also examine the prediction of more objective criteria. In the present research, these assessments predicted a broad range of criteria that included both self-ratings of attitudes and behavior as well as peer and supervisor ratings of Drill Sergeants' performance. Although these outcomes provide useful information about the utility of these measures, it would also be useful to examine the validities of the NSAB and ARC for predicting more objective outcomes. Self-reports can sometimes be inflated due to socially desirable responding and both peer and supervisor ratings can suffer from rater biases. Therefore, examining the prediction of objective criteria will provide an additional source of evidence for the validities of the NSAB and ARC. Past research has found that the NSAB has validity for predicting objective criteria like attrition, training success, or training failure (Nye et al., 2012; Nye et al., 2014). As such, similar research would be beneficial with a sample of Drill Sergeants.

The results of the current research suggest that both the NSAB and ARC are promising predictors of Drill Sergeant success and can predict a broad range of outcomes. Importantly, these measures predicted these outcomes both in training and on the job, indicating that both assessments may be useful for identifying high potential individuals for Drill Sergeant assignments.

REFERENCES

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

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

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

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 selection and classification decisions* (Technical Report 1311). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26-34.

Halstead, J. B. (2009). Recruiter selection model and implementation within the U.S. Army. *IEEE Transactions on Systems, Man, and Cybernetics, 39*, 93-100.

Horgen, K. E., Kubisiak, U. C., Lee, V. B., Connell, P. W., Penney, L. M., Borman, W. C., Pace, V., Lentz, E., White, L. A., Young, M. C., & Bowles, S. V. (2006). *Evaluation and refinement of a screening tool for U.S. Army Recruiters: Noncommissioned Officer Leadership Skills Inventory* (Technical Report 1177). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Horgen, K. E., Nye, C. D., White, L. A., Laporte, K. A., Hoffman, R. R., Drasgow, F., Chernyshenko, O. S., Stark, S., & Conway, J. S. (2013). *Validation of the Non-Commissioned Officer Special Assignment Battery (NSAB)* (Technical Report 1328). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*, 875-925.

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

Kubisiak, U. C., Horgen, K. E., Connell, P. W., Lentz, E., Xu, X., Borman, W. C., White, L. A., & Young, M. C. (2005). *Concurrent validation for the NLSI for U.S. Army drill sergeants* (Research Note 2006-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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

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

Muhammad, R. S., Rupprecht, E. A., & Graves, C. R. (2018). *Job analysis of U.S. Army Drill Sergeants* (Research Report 2019). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mueller-Hanson, R. Heggestad, E. D., & Thornton, G. C. (2003). Faking and Selection: Considering the Use of Personality from a Select-In and Select-Out Perspective. *Journal of Applied Psychology, 88*, 348-355.

Nye, C. D., Beal, S. A., Chernyshenko, O. S., Stark, S., Drasgow, F., White, L. A., & Dressel, J. D. (2014). *Assessing the Tailored Adaptive Personality Assessment System (TAPAS) for Special Forces personnel* (Research Report 1971). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Nye, C. D., Drasgow, F., Chernyshenko, O. S., Stark, S., Kubisiak, C., White, L. A., & Jose, I. (2012). *Assessing the Tailored Adaptive Personality Assessment System (TAPAS) as an MOS qualification instrument* (Technical Report 1312). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Nye, C. D., Drasgow, F., Chernyshenko, O. S., Stark, S., Muhammad, R. S., & Wolters, H. M. (2017). *New scale development for enhanced suitability screening (ESS)* (Research Note). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Nye, C. D., Muhammad, R. S., Graves, C., Drasgow, F., Chernyshenko, O. S., Stark, S., & Butt, S. (2018). *Examining enhanced suitability screening for predicting performance in recruiting duty assignments* (Technical Report 1366). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

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*, 138-152.

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

White, L. A., Young, M. C., Hunter, A. E., & Rumsey, M. G. (2008). Lessons learned in transitioning personality measures from research to operational settings. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*(*3*), 291-295.

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