

Technical Report 1419

Development of a Computer Adaptive Non-Cognitive Test for Officer Assessment Applications: Research Plan

Christopher Nye
Drasgow Consulting Group

Colin L. Omori
U.S. Army Research Institute

Fritz Drasgow
Oleksandr Chernyshenko
Stephen Stark
Nathan Baker
William Scott
Drasgow Consulting Group



December 2022

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

Approved for public release; distribution is unlimited.

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

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

Authorized and approved:

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

Research accomplished under contract
for the Department of the Army by

Drasgow Consulting Group

Technical review by:

Krystal Roach, U.S. Army Research Institute
Elyssa Johnson, U.S. Army Research Institute

DISPOSITION

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

REPORT DOCUMENTATION PAGE					
1. REPORT DATE (dd-mm-yy) December 2022		2. REPORT TYPE Final		3. DATES COVERED (from... to) April 2021 – July 2022	
4. TITLE AND SUBTITLE Development of a Computer Adaptive Non-Cognitive Test for Officer Assessment Applications: Research Plan				5a. CONTRACT OR GRANT NUMBER W911NF-21-C-0039	
				5b. PROGRAM ELEMENT NUMBER 622785	
6. AUTHOR(S) Christopher D. Nye, Colin L. Omori, Fritz Drasgow, Oleksandr Chernyshenko, Stephen Stark, Nathan Baker, & William Scott				5c. PROJECT NUMBER A790	
				5d. TASK NUMBER 1031	
				5e. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Drasgow Consulting Group 3508 N High Cross Rd Urbana, IL 61802-9759				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg 1464/Mail Stop 5610) Fort Belvoir, VA 22060				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER Technical Report 1419	
12. DISTRIBUTION/AVAILABILITY STATEMENT, Distribution Statement A: Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES ARI Research POC: Dr. Colin L. Omori, Selection and Assignment Research Unit					
14. ABSTRACT (Maximum 200 words): Officers play an important role in the U.S. Army and have a direct impact on the Soldiers that they lead. As a result, it is important to be able to identify high-potential individuals for these roles. The goal of the current research was to identify the most promising characteristics of an assessment that can support officer selection and assessment requirements. First, a meta-analysis was conducted to identify the personality characteristics that have the strongest relationships with leader potential and performance outcomes. The results of this meta-analysis were then supplemented with a qualitative review of Army doctrine and subsequent focus groups with subject matter experts to develop and refine a comprehensive list of personality characteristics for officer assessment. In addition, a review of the psychometric literature was also conducted to identify new models and administration formats that may enhance the validity of the assessment. This research provided a clearer picture of the characteristics of non-cognitive assessments that can be successfully used for officer selection and assignment requirements. Therefore, this report concludes with a proposed research plan for the Army Research Institute for the Behavioral and Social Sciences (ARI) to develop an officer assessment that incorporates the findings from this project.					
15. SUBJECT TERMS: Officers, Manpower, Personality Measures, Selection and Assignment, Army Personnel					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited Unclassified	85	Tonia S. Heffner 703-545-4408

Technical Report 1419

**Development of a Computer Adaptive Non-Cognitive
Test for Officer Assessment Applications:
Research Plan**

Christopher D. Nye
Drasgow Consulting Group

Colin L. Omori
U.S. Army Research Institute

Fritz Drasgow
Oleksandr Chernyshenko
Stephen Stark
Nathan Baker
William Scott
Drasgow Consulting Group

Selection and Assignment Research Unit
Tonia S. Heffner, Chief

December 2022

Approved for public release; distribution is unlimited

ACKNOWLEDGEMENT

The authors are especially thankful to Drs. Tonia Heffner and Elizabeth Salmon for their insights and unwavering support. The authors are also thankful to the officers and other subject matter experts who gave their time and attention to support this research and accomplish this task.

DEVELOPMENT OF A COMPUTER ADAPTIVE NON-COGNITIVE TEST FOR OFFICER ASSESSMENT APPLICATIONS: RESEARCH PLAN

EXECUTIVE SUMMARY

Research Requirement:

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 temperament constructs 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 (Schmidt & Hunter, 1998). However, due to the limitations of traditional personality measures, some research has called into question their use in high-stakes settings (Morgeson et al., 2007). Although traditional personality measures are useful in research and counseling settings where respondents have little motivation to respond dishonestly, they are unsuitable for making important personnel decisions for several reasons. In particular, research shows that faking is an important concern in traditional personality assessments (White et al., 2008).

Given the limitations of traditional personality measures, a new approach is needed to support officer selection and assessment requirements in the U.S. Army. To take advantage of recent psychometric advances, this report describes a series of tasks that were designed to identify the characteristics of successful officers in the Army and the most promising administration formats for high-stakes non-cognitive assessments. Based on the findings from these tasks, a rigorous research plan is proposed to create a new non-cognitive assessment that can be used for officer selection and assessment.

Procedure:

For this research, the first task involved conducting a meta-analysis to identify the personality traits that provide the best prediction of leadership potential and performance. The focus of this meta-analysis was on identifying those characteristics that are relevant to officers in the Army. To supplement the results of the meta-analysis, we also reviewed existing Army doctrine and program of instruction (POI) documentation that outlines the fundamental characteristics of officers. The goals of reviewing these materials were to 1) identify the characteristics of officers that are viewed as most important by Army leaders, and 2) ensure that we identified characteristics that are unique to Army officers but that may not have been examined in the previous research that was available for the meta-analytic review.

Building on the results of the meta-analysis and the review of Army doctrine, we also conducted focus groups with subject matter experts (SMEs) in the U.S. Army. The goal of these focus groups was to further supplement the literature reviews and identify characteristics that existing officers feel are important in their roles. Furthermore, discussing the characteristics of successful officers with SMEs was important to avoid relying too much on published research, especially given that the majority of this research was conducted in civilian samples.

Finally, in addition to examining the characteristics of successful officers, we also reviewed the literature on relevant item response theory (IRT) models and administration formats that may enhance scale validity. A number of models and methods that can help to improve the assessment of non-cognitive characteristics have been developed recently in the psychometrics literature. These advances were reviewed with a goal of identifying specific models or formats that 1) are useful for non-cognitive assessments, 2) are fake resistant, and 3) can be administered effectively in high-stakes settings.

Findings:

Based on the literature reviews and focus groups, we identified a set of 16 personality dimensions that can be administered in a new officer assessment. These 16 dimensions will help to address the competing goals of a) assessing a comprehensive set of personality characteristics and b) ensuring an efficient measure that can be administered in the limited amount of time available for officer assessments. In addition, our review of existing methodologies suggested that using a forced choice format is the most promising approach for reducing faking in high-stakes testing. Although other approaches have been proposed, the forced choice format has been the most widely researched and the available evidence suggests that this format is effective.

There are numerous variations on the forced choice format that can be used for personality assessment. However, our review suggests that the two alternative forced choice (2AFC) format appears to be the most promising *at this time*. Existing research has identified a number of potential alternative approaches that may have advantages over the 2AFC format. However, these alternative approaches have not yet been adequately tested and several questions remain about their use in high-stakes settings. In contrast, the 2AFC format has been widely supported and successfully used in high-stakes settings. Therefore, this approach can be implemented immediately while research continues on other promising alternatives.

Finally, we also recommend incorporating empirical bayes (EB) scoring, which has the potential to improve score precision and increase the reliability of the personality facets assessed by the proposed measure. Importantly, EB scores can also be combined with other models to obtain the benefits of both. For example, EB scoring could be used to augment IRT trait estimates from a 2AFC forced choice measure to obtain scores that are both resistant to faking and more reliable than traditional 2AFC measures.

Utilization and Dissemination of Findings:

This research provides a clearer picture of the characteristics of non-cognitive assessments that can be successfully used for officer selection and assignment requirements. Therefore, this report concludes with a proposed research plan to develop an officer assessment that incorporates the findings from this research. The proposed research plan would take approximately 36-48 months (depending on the specific approach used) and involve developing new statement pools, pretesting the new statements to estimate IRT parameters and remove low quality statements, and conducting a validation study of the new assessment. The approach described here follows current professional standards for scale development (American

Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014; SIOP, 2018) and will provide the initial evidence necessary to support the use of the new measure for officer selection and assessment requirements. Given the potential influence of personality on work performance (e.g., Barrick & Mount, 1991; Campbell & Knapp, 2001), this new assessment is likely to help identify individuals with high-potential for success as officers in the U.S. Army.

DEVELOPMENT OF A COMPUTER ADAPTIVE NON-COGNITIVE TEST FOR OFFICER ASSESSMENT APPLICATIONS: RESEARCH PLAN

CONTENTS

	Page
INTRODUCTION.....	1
Background	1
Purpose of the Current Research	1
META-ANALYSIS OF THE RELATIONSHIP BETWEEN PERSONALITY CHARACTERISTICS AND LEADERSHIP OUTCOMES.....	3
Methods	3
Literature Search	3
Inclusion Criteria	3
Study Coding	4
Meta-Analytic Procedure	6
Results	8
REVIEW OF EXISTING ARMY DOCTRINE AND FOCUS GROUPS WITH SUBJECT MATTER EXPERTS.....	22
Background	22
Review of Existing Army Doctrine.....	22
Focus Groups with SMEs.....	26
Focus Group Procedures	27
Focus Group Results	27
Summary and Recommendations.....	29
REVIEW OF POTENTIAL METHODS FOR NON-COGNITIVE ASSESSMENTS.....	36
Literature Review	36
Rapid Response Measurement (RRM)	36
Two-Alternative Forced Choice (2AFC) Format	37
Three-Alternative Forced Choice (3AFC) Format	38
Zinnes-Griggs (ZG) Model.....	40
Forced Choice Graded Preference (FCGP) Format	41

CONTENTS (Continued)

	Page
Empirical Bayes (EB) Scoring.....	41
Summary and Recommendations.....	43
RESEARCH PLAN TO DEVELOP A NON-COGNITIVE ASSESSMENT FOR OFFICER APPLICATIONS	47
Proposed Research Plan	47
Selecting an Underlying IRT Model.....	47
Statement Development and Pretesting	49
CAT Software and EB Scoring.....	50
Validation of the New Assessment.....	51
Anticipated Timeline	51
CONCLUSIONS	54
REFERENCES.....	55

LIST OF TABLES

Table 1. Big Five Traits and Corresponding Facets Included in the Review	6
Table 2. Average Reliability Estimates used for Corrections.....	7
Table 3. Relationships between Personality Dimensions and Leader Performance Ratings.....	10
Table 4. Relationships between Personality Dimensions and Leadership Emergence.....	15
Table 5. Relationships between Personality Dimensions and Transformational Leadership	16
Table 6. Relationships between Personality Dimensions and Transactional Leadership	18
Table 7. Relationships between Personality Dimensions and Physical Fitness.....	20
Table 8. Relationships between Personality Dimensions and Intentions to Remain in the Military	21
Table 9. Summary of the Talent Priorities (Behaviors) Desired by the Basic Branches in the Army.....	25
Table 10. Results of Previous Qualitative Reviews of Officer Characteristics	26
Table 11. Participants' Rank Order of Personality Characteristics for Officers.....	29
Table 12. Proposed Dimensions for the Officer Assessment	31

CONTENTS (Continued)

	Page
Table 13. Correlations between the Raw and Augmented (EB) Scores	44
Table 14. Reliabilities of the Raw and Augmented (EB) Scores.....	45
Table 15. Construct and Criterion Correlations for the Raw and Augmented (EB) Scores	46

LIST OF FIGURES

Figure 1. Army Leadership Requirements Model	23
Figure 2. Desired Characteristics for Officers in the Infantry Branch.....	24
Figure 3. Possible 36-Month Timeline for the Proposed Research Plan	52
Figure 4. Possible 48-Month Timeline for the Proposed Research Plan	53

DEVELOPMENT OF A COMPUTER ADAPTIVE NON-COGNITIVE TEST FOR OFFICER ASSESSMENT APPLICATIONS: RESEARCH PLAN

Background

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 temperament constructs 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 (Schmidt & Hunter, 1998). Given these potential advantages, many organizations are now using personality measures to make decisions in high-stakes selection and classification settings.

The use of personality assessments for selection and classification inevitably calls attention to the quality of their measurement. In fact, due to the limitations of traditional personality measures, some research has called into question their use in high-stakes settings (Morgeson et al., 2007). Although we believe that these traditional personality measures are useful in research and counseling settings where respondents have little motivation to respond dishonestly, they are unsuitable for making important personnel decisions for several reasons. First, in high-stakes testing situations, research shows that faking is an important concern in traditional personality assessments; i.e., test takers can discern the correct or socially desirable answers and, thus, increase or decrease their scores (White et al., 2008). This intentional distortion can severely undermine the utility of personality measures for personnel selection. Second, currently used scales were not constructed to measure well across all levels of the trait continuum. Specifically, because classical test theory methods were used to evaluate and choose items during scale development, only those items having moderately positive and moderately negative standing on the underlying trait continuum were retained; extreme and neutral items were discarded (Stark, Chernyshenko, & Drasgow, 2005). This affects the rank-order of high and low scoring individuals who are often of primary interest in selection contexts. Finally, traditional paper and pencil personality measures are inefficient and cumbersome to administer and maintain. They have rigid administration prescriptions in the sense that all items must be administered to every examinee in a prespecified order. This increases testing time and decreases test security through repeated item exposure. In addition, because organizations are often interested in only a subset of scales that vary depending on the occupation, it would be better to have a flexible way of choosing the constructs administered on particular occasions, an option not available in most current inventories.

Given the limitations of traditional personality measures, a new approach is needed to support officer selection and assessment requirements in the U.S. Army. Research in the psychometric literature has shown that IRT-based measures can mitigate faking (Drasgow et al., 2012; Trent et al., 2020) and improve the assessment of non-cognitive characteristics in selection settings (Stark et al., 2005). In addition, computer adaptive testing can also be used to generate more efficient measures that are up to 50% shorter than a static personality assessment (Stark et al., 2012) and more resistant to cheating (Drasgow et al., 2009). These advances could be incorporated into a new measure that has the potential to improve officer selection and assessment.

Purpose of the Current Research

To take advantage of recent psychometric advances, this report describes a rigorous research plan to create a new non-cognitive assessment that can support officer selection and assessment requirements. This plan was developed based on several tasks that helped to identify the key characteristics of successful officers and the most promising administration format for a high-stakes non-cognitive assessment. The first task involved conducting a meta-analysis to identify the personality traits that are likely to provide the best prediction of leadership potential and performance. This meta-analysis updated previous meta-analytic work conducted by the military to identify the characteristics of successful Soldiers (Dragow et al., 2012). However, the focus of the current review was on identifying those characteristics that are relevant to officers in the Army.

In addition to conducting the meta-analysis, we also reviewed existing doctrine and program of instruction (POI) documentation that outline the fundamental characteristics/attributes of Army officers. This included, for example, the Department of the Army Pamphlet (DA Pam) 600-3 on *Commissioned Officer Professional Development and Career Management* (Department of the Army, 2019), relevant field manuals (FM) such as FM 6-22 on *Leader Development* (Department of the Army, 2019), and branch-specific guidance on the characteristics of successful officers. The goals of reviewing these materials was to 1) identify the characteristics of officers that are viewed as most important by Army leaders, and 2) ensure that we identified characteristics that are unique to Army officers but that may not have been examined in the previous research that was available for the meta-analytic review.

Next, we also conducted focus groups with subject matter experts (SMEs) in the U.S. Army. The goal of these focus groups was to further supplement the literature reviews and identify characteristics that existing officers feel are important in their roles. Furthermore, discussing the characteristics of successful officers with SMEs was important to avoid relying too much on published research, especially given that the majority of this research was conducted in civilian samples.

Finally, in addition to examining the characteristics of successful officers, we also reviewed the literature on relevant item response theory (IRT) models and administration formats that may enhance scale validity. A number of models and methods have been developed recently in the psychometrics literature that may be able to improve the assessment of non-cognitive characteristics. These advances were reviewed with a goal of identifying specific models or formats that 1) are useful for non-cognitive assessments, 2) are fake resistant, and 3) can facilitate the assessment of non-cognitive characteristics in high-stakes settings.

META-ANALYSIS OF THE RELATIONSHIP BETWEEN PERSONALITY CHARACTERISTICS AND LEADERSHIP OUTCOMES

Methods

Literature Search

As described above, the first step we used to identify the characteristics of successful officers was to conduct a meta-analysis to examine the non-cognitive predictors of leadership outcomes. For this quantitative review, we conducted a broad search to identify relevant studies that reported relationships between personality and leadership outcomes and were published between 1990 and 2021. Given that consensus regarding the validity of the five-factor model started to emerge in the early 1990's (cf. Chernyskenko et al., 2011; Goldberg 1990; McCrae & Costa, 1987) we selected 1990 as the cutoff for our meta-analytic review. During our literature search, we focused on both military and civilian research to provide a more comprehensive evaluation of the personality literature and to identify a comprehensive set of personality characteristics.

To identify articles for this meta-analysis, we searched Google Scholar, PsycINFO (including PsycARTICLES), EBSCOHost, and ProQuest using the following keywords: *Personality, Personality Traits, Personality Facets, Personality Predictors, TAPAS, Big Five, Grit, Leaders, Leadership, Leader Performance, Leader Effectiveness, Leader Emergence, Transformational Leadership, Transactional Leadership, Military Leader Effectiveness, Military Leader Performance, Military Leader Emergence, Military, Officers, Officer Leadership, NCOs, Supervisors, Firefighter Leaders, and Police Leaders*. To better capture the personality research performed in military settings, we also examined research reported in the International Military Testing Association (IMTA) Annual Conference archives, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Technical Library, Canadian Armed Forces research reports, and U.S. Air Force and Navy research reports using the same search parameters. To identify additional unpublished conference presentations, we searched the past conference programs for both IMTA and the Society for Industrial and Organizational Psychology. Finally, a set of coders examined the references of the studies identified in our search to find additional relevant studies and technical reports.

Inclusion Criteria

The studies identified in our literature search were reviewed to determine if they were relevant for inclusion in the meta-analysis. Studies were included if they a) reported a relevant effect size (thus excluding a handful of qualitative studies that only reported interview data), b) reported results for at least one dimension of personality, and c) reported relationships with a measure of leader emergence, training performance, transformational leadership, or transactional leadership. For each of these outcomes, we only included studies that used other-reports (e.g., supervisors, peers, grades) of leadership performance. Self-reports of leadership performance were *excluded* from these analyses for several reasons. First, focusing on other-reports reduces the potential for common method variance, which can artificially inflate effect sizes between personality and leadership performance (Podsakoff et al., 2003). Second, self-reports were

excluded due to the potential influence of self-presentation bias on performance outcomes (Conway & Huffcutt, 1997; Hoffman et al., 2011; Paulhus & Vazire, 2007). Studies that did not report a leadership-relevant outcome or that focused solely on follower outcomes were also excluded from analyses.

For the personality predictors, studies reporting only other-reports of personality (e.g., reports of personality from subordinates or friends) were excluded from the meta-analysis as well because they potentially assess different aspects of personality than self-reports (i.e., reputation vs. identity; Nye & Roberts, 2013). As shown in previous research, informant-reports tend to have low levels of agreement with self-reports of personality (Kandler et al., 2010). Studies that examined contextualized personality measures (e.g., personality at work versus at home) were removed for the same reason. Although our preference would have been to include these alternative measures of personality as moderators of the relationship between personality and performance, there were not a sufficient number of studies to include them in these analyses (less than 5 total). Finally, a small number of studies using the *True Colors Personality Test* or the *MBTI* were also excluded because they do not map onto the Big Five traits and have questionable validity (Boyle, 1995).

After excluding all studies that did not meet our inclusion criteria, we examined the final pool of studies to remove any redundant samples. The majority of such cases occurred when multiple published articles or technical reports presented results drawn from the same data. When this was the case, the coders reviewed the studies and used whichever article contained the most comprehensive set of variables and relevant statistics (e.g., standard deviations, correlations). Based on these inclusion criteria, the final meta-analytic database included a total of 59 military studies and 75 civilian studies ($k = 134$) for analyses.

Study Coding

After obtaining the final database of relevant studies, we coded each study for the variables of interest. Consistent with previous meta-analyses (e.g., Hoffman et al., 2011), we randomly selected 10% of the studies and calculated the interrater agreement between two coders. Average interrater agreement was 94% and discrepancies between coders were discussed by the authors and resolved. Demographic variables (age, gender composition of the sample, job type, country, education level, and ethnicity) were recorded for each study that reported this information. The characteristics of each study (i.e., the sources of the leadership ratings, field vs. lab study, predictive vs. concurrent study design) were recorded as well.

Across all studies, numerous leadership outcomes were examined. Therefore, we coded each criterion variable into six categories representing the following aspects of leadership:

1. **Leader Emergence.** Emergence occurs when an individual without a formal leadership role begins to exert influence on others and/or be perceived as leader-like. Studies reporting the emergence of leaders from groups of equal peers or that examined recommendations for leadership roles were included in this category.

2. **Leader Performance Ratings.** This aspect of leader performance included studies that examined performance/effectiveness ratings of leaders provided by a) supervisors, b) peers, or c) instructors (i.e., academic grades).
3. **Transformational Leadership.** Studies that assessed an individual's ability to inspire and obtain support from followers in pursuit of shared goals were included in this category (Bass, 1985).
4. **Transactional Leadership.** This type of leader behavior is task-focused and refers to a leader's ability to obtain cooperation from subordinates by providing them with rewards/resources in exchange for performance (Bono & Judge, 2004).
5. **Intentions to Stay.** This leadership criterion was only examined in military studies and focused on intentions to remain in a military leadership role.
6. **Physical Fitness.** Fitness scores were only examined in military settings as these are an important criterion for success in many military leadership roles.

A number of different personality characteristics were examined in the studies included in this meta-analysis as well. These characteristics included both broad Big Five traits and narrower personality facets. Therefore, we examined relationships with both broad and narrow personality dimensions. To do so, the facet-level personality scales reported in each study were matched to their corresponding Big Five trait either using the relationships reported in the technical manual for each scale, if available, or by trained subject matter experts (SMEs). In addition, trained SMEs also examined the measures used to assess each personality characteristic to ensure that these constructs were being measured similarly in each study.

Personality facets were also coded using the taxonomy proposed by the Organisation for Economic Cooperation and Development (OECD) (Chernyskenko et al., 2018). This taxonomy was developed using both rational and empirical approaches to identify a 26-facet structure of personality that was used in the current research as much as possible to categorize the personality facets reported in each study. However, not all of the facets in the OECD taxonomy could be examined in the current study. For example, too few studies reported results for the Agreeableness facets of Trust and Respectfulness or the Conscientiousness facet of Perseverance. In addition, due to the relatively low number of studies examining the facets of Openness, there were not enough studies to calculate relationships at the facet level for this trait. Therefore, we used the intermediate aspects of personality reported by DeYoung et al. (2007) and Woo et al. (2014) to code the narrower dimensions of Openness.

Finally, we also included a number of additional personality characteristics that were of particular interest in the leadership literature. For example, we included Narcissism and Self-Efficacy because these dimensions are often of interest when examining leader behaviors (Braun 2017; Machida & Schaubroeck, 2011). In addition, we chose to examine the Agreeableness facet of Humility separately given that comparable measures in our dataset were mainly from work using the Modesty-Humility scale from the HEXACO model (Lee & Ashton, 2004). Given that Humility is sometimes considered a 6th broad dimension of personality, we decided to examine the few studies focusing on this dimension separately from the broader trait of Agreeableness. We also examined the Energy facet of Extraversion separately given that most measures of this trait in our database treated it this way as well. The final taxonomy of facets used in this study is summarized in Table 1.

Table 1. Big Five Traits and Corresponding Facets Included in the Review

Extraversion					
Assertiveness		Sociability		Sensation Seeking	
Agreeableness					
Empathy		Generosity		Cooperation	
Emotional Stability					
Stress Resistance - Anxiety		Emotional Control - Volatility		Optimism	
Conscientiousness					
Achievement Motivation	Orderliness	Self-control	Responsibility	Discipline	Honesty-Virtue
Openness					
Intellect ^a			Culture ^a		
Intellectual Efficiency	Ingenuity	Curiosity	Tolerance	Aesthetics	Depth
Other					
Energy ^b	Humility ^b	Hardiness ^b	Narcissism		Self-Efficacy

Note: The naming conventions used for this meta-analysis were drawn from Chernyskenko et al. (2018) except where otherwise noted. ^aThis naming convention is based on the paper by Woo et al. (2014). ^bThese facets are included under a Big 5 dimension in Chernyskenko et al. (2018), but examined separately in this study.

Finally, studies were also coded to account for moderators of interest within the study. Research in the military context often finds relationships that differ from those observed in civilian settings (e.g., Capon et al., 2007; Pulakos et al., 1989). This is not surprising and is likely due to the unique characteristics of the military, such as strong top-down control (cf. Bradley, 2006; Darr, 2011; Horn, 2006). For these reasons, we coded whether correlations were obtained from a military or civilian sample for use in subsequent moderator analyses. We also coded correlations based on the types of leaders included in each sample. Different relationships often emerge at different levels of leadership (e.g., lower-level managers vs. executives) because of the differences in tasks performed at each level (Hoffman et al., 2011). Therefore, we coded the samples into three main leadership levels including a) line managers/noncommissioned officers (NCOs), b) university (e.g., MBA) students or military academy trainees, or c) executives, managers, EMBA students, and current officers.

Meta-Analytic Procedure

We based the meta-analytic process on recommendations made by Hunter and Schmidt (2004) and corrected effect size estimates for unreliability, range restriction, and sampling error. Given that we account for indirect range restriction in our samples, we first corrected for unreliability in the predictor and criterion (Hunter et al., 2006). For personality measures, 57% of studies reported a reliability estimate for the measure, while 52% reported reliability estimates for the leadership criterion measure. For studies that did not report reliability estimates, we used the average reliability estimate across all studies that reported reliabilities for the corresponding construct. The average reliabilities used for these corrections are reported in Table 2.

Table 2. Average Reliability Estimates used for Corrections

Criterion Measures	Reliability
Leader Emergence	.89
Leader Performance Ratings	.89
Transformational leadership	.86
Transactional leadership	.82
Physical Fitness ^a	.88
Intentions to Stay	.92
Personality Measures	
Extraversion	.80
Assertiveness	.76
Sociability	.71
Sensation Seeking	.71
Agreeableness	.74
Empathy	.64
Generosity	.67
Cooperation	.72
Emotional Stability	.83
Stress Resistance-Anxiety	.67
Emotional Control - Volatility	.68
Optimism	.66
Conscientiousness	.80
Achievement Motivation	.76
Orderliness	.78
Self-Control	.83
Responsibility	.71
Discipline	.63
Honesty-Virtue	.79
Openness	.77
Intellectual Efficiency	.71
Ingenuity	.78
Curiosity	.62
Tolerance	.76
Aesthetics	.71
Depth	.88
Other	
Energy	.73
Humility	.78
Hardiness	.74
Narcissism	.68
Self-Efficacy	.76

Note. The Big Five reliability estimates are based on global measures of the corresponding trait. ^aNone of the studies in the meta-analysis included reliability estimates for the physical fitness measure. Therefore, we used the average reliability estimate across all criterion measures as the estimate for this outcome.

Next, we corrected for range restriction in each sample. To obtain estimates of the unrestricted standard deviations for each personality construct, we examined relevant technical manuals (e.g., Costa & McCrae, 2008; Hogan, 2002) and large community datasets (e.g., the Eugene-Springfield Community Sample; Goldberg, 2018), and used these estimates to correct for range restriction in the predictor variables. Using these unrestricted standard deviations (SDs) and the restricted SDs from the samples, we then used the procedures described by Hunter et al. (2006) to correct for range restriction. Although the participants in the studies we analyzed were not *directly* selected for leadership roles based on their personality (i.e., direct range restriction), they were likely to self-select into these roles based on their perceptions of their personality. As noted by Hunter et al. (2006), self-selection is a form of indirect range restriction and, therefore, we corrected for indirect range restriction in each sample. For studies that did not report the SD of a measure or when an estimate of the unrestricted SD could not be identified, we used the average ratio of the SDs (i.e., the unrestricted/the restricted SD) for that variable to perform these corrections.

Finally, we also corrected the variance of correlations for sampling error. After performing these corrections, we calculated an 80% credibility interval around each of the estimates. Credibility intervals partial out sampling error variance and provide a range within which the true population effect can be found. Thus, credibility intervals can indicate the extent to which population scores vary, with larger variation suggesting possible moderators. To examine the role of moderators, we then calculated separate estimates of the correlations between personality and leadership performance in each of the moderating conditions (e.g., levels of leadership, military/civilian samples). For these analyses, separate results are only reported when 3 or more studies were available for calculating a meta-analytic correlation.

Results

The results of the meta-analyses are shown in Tables 3-8. Each of these tables illustrates the relationships between a number of personality dimensions and various leadership outcomes. Table 3 provides the most substantial number of results and reports the correlations with leader performance ratings. This table shows that a number of personality dimensions had small to moderate correlations with leader performance ratings. Some of the strongest relationships were with the Assertiveness facet of Extraversion, which ranged from .15 to .24 across military and civilian samples and various leadership levels. In addition, achievement motivation ($\rho_{(b)} = .20$), responsibility ($\rho_{(b)} = .22$), optimism ($\rho_{(b)} = .28$), and energy ($\rho_{(b)} = .25$) also had moderate meta-analytic relationships with leader performance ratings. In other words, a number of personality dimensions were related to leader performance ratings.

In contrast to the leader performance ratings, fewer personality characteristics have been examined as potential predictors of the other leadership outcomes examined here. Nevertheless, several small to moderate correlations were observed for other outcomes like leader emergence, transformational leadership, and transactional leadership. The meta-analytic correlations with these outcomes are shown in Tables 4-6. Across these outcomes, several consistent findings emerged. For example, the assertiveness and responsibility facets predicted several of these outcomes including leader emergence (responsibility $\rho_{(b)} = .28$), transformational leadership (assertiveness $\rho_{(b)} = .18$), and transactional leadership (assertiveness $\rho_{(b)} = .19$; responsibility $\rho_{(b)} =$

.19). Optimism was also correlated with both transformational ($\rho_{(b)} = .16$) and transactional leadership ($\rho_{(b)} = .24$). These results suggest that many of these facets are related to several leadership outcomes.

Because they were frequently examined in the literature and highly relevant to the military, we also examined the personality correlates of leaders' physical fitness and intentions to stay on the job. Not surprisingly, the energy facet was the strongest correlate of physical fitness ($\rho_{(b)} = .47$). Numerous other facets were related to intentions to remain on the job including empathy ($\rho_{(b)} = .36$), achievement motivation ($\rho_{(b)} = .33$), discipline ($\rho_{(b)} = .27$), emotional control ($\rho_{(b)} = .30$), optimism ($\rho_{(b)} = .30$), and self-efficacy ($\rho_{(b)} = .30$). Given these results, it seems that personality is a particularly strong correlate of leaders' intentions to stay on the job.

Several of the global personality traits were also strongly correlated with leadership outcomes. For example, in the military, conscientiousness was strongly correlated with leader performance ratings ($\rho_{(b)} = .29$ for all military studies but .22 for officers), transformational leadership ($\rho_{(b)} = .31$), and transactional leadership ($\rho_{(b)} = .21$). Similarly, emotional stability ($\rho_{(b)} = .22$) and openness ($\rho_{(b)} = .20$) were also moderately correlated with transactional leadership in the military. These results suggest that composites of the facets for a particular trait might be correlated with leadership outcomes even if the individual facets are not. However, the global trait of conscientiousness had a meta-analytic correlation of .00 with intentions to remain on the job, despite relatively strong correlations between this outcome and several of the facets associated with this global trait (e.g. achievement motivation, discipline). These results suggest that simply aggregating scores on the facets using equal weights and assuming correlations in the same direction may miss out on important variation across facets. Therefore, it is important to examine facets individually, taking into consideration that optimally weighted composites of those facets may be even stronger predictors of leadership outcomes (see also Nye et al., 2020).

Finally, we also found a number of moderators of the relationships between personality and leadership outcomes. When a sufficient number of studies were available for these analyses, we found evidence of differences across military and civilian samples. For example, the relationship between responsibility and leader performance ratings was .22 in civilian samples but only .10 in military samples. In addition, the correlation between optimism and leader performance ratings was .28 in civilian samples but .16 in military samples. Similarly, the trait of agreeableness was correlated .23 with transformational leadership in civilian samples but only .06 in military samples. These results suggest that the important characteristics of leaders may vary across settings, but only for some facets. Nevertheless, even within the military samples, there was also variation across different levels of leadership. Although the relationship between optimism and leader performance ratings was .16 across all military samples, the correlation varied from .15 for academy trainees to .20 for officers. As another example, the correlation between the energy facet and leader performance ratings was .11 in samples of non-commissioned officers (NCOs) but .28 for trainees. Although there was limited data to examine differences across leadership levels for many facets, particularly in military settings, these results suggest that some personality predictors of officer performance may be different from the predictors of performance in other military roles.

Table 3. Relationships between Personality Dimensions and Leader Performance Ratings

											80 % CV	
Trait	Population	Facet	Leadership Level	<i>N</i>	<i>k_s</i>	<i>k_c</i>	<i>r</i>	<i>ρ_(c)</i>	<i>ρ_(b)</i>	<i>SD_{ρ(b)}</i>	lower	upper
Extraversion												
	All	Global	All	38821	46	91	.09	.10	.13	.20	-.12	.38
	Civilian	Global	All	10339	30	49	.09	.10	.12	.15	-.06	.29
			Line Supervisors	723	4	4	.05	.05	.06	.06	N/A	N/A
			University Students	2499	7	11	.12	.14	.16	.14	.00	.33
			Managers/Executives	7117	20	34	.08	.09	.10	.16	-.08	.29
	Military	Global	All	28482	16	42	.09	.09	.14	.21	-.13	.40
			Academy Trainees	18744	11	30	.01	.01	.02	.15	-.17	.20
			Officers	925	4	5	.08	.09	.12	.14	-.03	.27
	All	Assertiveness	All	49136	33	101	.14	.15	.20	.12	.06	.34
	Civilian	Assertiveness	All	6642	7	19	.16	.17	.24	.13	.08	.40
			Managers/Executives	3284	3	10	.17	.18	.23	.15	.04	.41
	Military	Assertiveness	All	42494	26	82	.14	.15	.19	.11	.06	.33
			NCOs	10622	6	23	.12	.13	.16	.12	.03	.30
			Academy Trainees	28951	15	43	.15	.16	.21	.11	.08	.33
			Officers	2921	5	16	.11	.12	.15	.14	.00	.29
	All	Sociability	All	18455	25	55	.07	.07	.10	.10	-.02	.21
	Civilian	Sociability	All	4978	11	19	.11	.12	.16	.12	.02	.30
			Managers/Executives	3486	5	11	.10	.10	.13	.12	.00	.27
	Military	Sociability	All	13477	14	36	.05	.05	.07	.09	-.02	.16
			NCOs	4164	5	9	.08	.09	.11	.09	.01	.21
			Academy Trainees	7748	6	13	.03	.04	.05	.07	-.02	.12
			Officers	1565	3	14	.04	.05	.06	.11	-.02	.14
	All	Sensation Seeking	All	5043	5	13	.03	.03	.04	.10	-.07	.14
	Military	Sensation Seeking	All	3268	3	7	.02	.02	.02	.11	-.10	.15
Agreeableness												
	All	Global	All	38817	44	88	.06	.06	.09	.13	-.06	.24
	Civilian	Global	All	9730	28	45	.03	.03	.03	.16	-.15	.22
			Line Managers	782	5	5	.12	.13	.16	.16	-.01	.33
			University Students	2279	7	9	.07	.07	.09	.09	.00	.17
			Managers/Executives	6669	17	31	.00	.00	.00	.17	-.20	.20
	Military	Global	All	28204	16	43	.07	.07	.11	.11	-.02	.24
			NCOs	10615	3	11	.10	.10	.16	.08	.07	.26
			Academy Trainees	17135	11	28	.05	.06	.08	.11	-.05	.20
	All	Empathy	All	10364	18	33	.08	.08	.11	.15	-.07	.29
	Civilian	Empathy	All	3977	5	11	.15	.16	.23	.09	.13	.32

Trait	Population	Facet	Leadership Level	N	k_s	k_c	r	$\rho_{(c)}$	$\rho_{(b)}$	$SD_{\rho(b)}$	80 % CV	
											lower	upper
	Military	Empathy	All	6387	13	22	.03	.03	.04	.14	-.12	.20
			NCOs	3523	5	8	.05	.05	.07	.09	-.02	.17
			Academy Trainees	2507	6	10	-.01	-.01	-.02	.18	-.24	.20
	All	Generosity	All	9156	12	17	.01	.01	.01	.12	-.13	.15
	Civilian	Generosity	All	893	3	4	.04	.04	.05	.12	-.08	.19
	Military	Generosity	All	8263	9	13	.00	.00	.00	.12	-.14	.14
			NCOs	2622	4	6	.01	.01	.02	.06	-.02	.06
			Academy Trainees	5641	5	7	.00	.00	.00	.13	-.17	.16
	All	Cooperation	All	14088	17	37	.02	.02	.03	.18	-.20	.25
	Civilian	Cooperation	All	4885	7	14	.14	.15	.20	.08	.12	.28
			Managers/Executives	1803	4	8	.14	.15	.20	.11	.09	.31
	Military	Cooperation	All	9203	10	23	-.04	-.04	-.07	.15	-.25	.12
			Academy Trainees	6245	5	15	-.09	-.09	-.14	.11	-.27	-.01
Openness												
	All	Global	All	36117	42	84	.10	.11	.13	.12	-.01	.28
	Civilian	Global	All	7924	26	43	.10	.10	.13	.12	.01	.25
			Line Managers	782	5	5	.11	.12	.13	.11	.03	.23
			University Students	2049	6	6	.10	.11	.14	.13	-.02	.30
			Managers/Executives	5093	16	32	.09	.10	.13	.11	.02	.23
	Military	Global	All	28193	16	41	.10	.11	.13	.12	-.02	.28
			Academy Trainees	18025	11	28	.07	.08	.10	.12	-.05	.25
			Officers	454	3	4	.01	.01	-.01	.16	-.18	.17
	All	Intellectual Efficiency	All	15963	19	41	.08	.08	.11	.16	-.09	.31
	Civilian	Intellectual Efficiency	All	4737	7	16	.05	.05	.07	.15	-.12	.25
	Military	Intellectual Efficiency	All	11226	12	25	.09	.10	.13	.16	-.08	.33
			Academy Trainees	7435	6	12	.12	.13	.16	.18	-.07	.39
			Officers	1345	3	8	.00	.00	.00	.09	-.07	.07
	All	Ingenuity	All	3218	8	9	.09	.10	.13	.09	.03	.22
	Civilian	Ingenuity	All	2023	4	5	.12	.13	.16	.08	.08	.25
	Military	Ingenuity	All	1195	4	4	.05	.05	.06	.07	.02	.10
	All	Curiosity	All	7066	7	9	-.01	-.01	-.01	.07	-.09	.07
	Civilian	Curiosity	All	2287	3	3	-.05	-.05	-.07	.04	-.09	-.05
	Military	Curiosity	All	4779	4	6	.01	.01	.02	.07	-.06	.10
	All	Intellect aspect*	All	26247	21	59	.06	.06	.08	.15	-.10	.26
	Civilian	Intellect aspect*	All	9047	8	24	.04	.04	.05	.14	-.12	.23
	Military	Intellect aspect*	All	17200	13	35	.07	.07	.09	.15	-.09	.27
			Academy Trainees	12868	7	20	.08	.08	.11	.16	-.09	.31

Trait	Population	Facet	Leadership Level	N	k_s	k_c	r	$\rho_{(c)}$	$\rho_{(b)}$	$SD_{\rho(b)}$	80 % CV	
											lower	upper
	All	Aesthetics	All	1619	4	6	-.04	-.04	-.05	.11	-.16	.06
	All	Tolerance	All	9175	15	36	.04	.04	.06	.17	-.15	.26
	Civilian	Tolerance	All	2574	6	11	.09	.09	.12	.14	-.04	.29
	Military	Tolerance	All	6601	9	25	.02	.02	.03	.17	-.18	.23
			Academy Trainees	4989	5	11	.01	.01	.01	.17	-.20	.22
	All	Culture aspect*	All	12621	17	46	.04	.04	.05	.16	-.14	.24
	Civilian	Culture aspect*	All	5045	7	17	.06	.06	.08	.16	-.10	.27
	Military	Culture aspect*	All	7576	10	29	.02	.03	.03	.16	-.16	.22
			Academy Trainees	5717	6	14	.01	.01	.02	.16	-.18	.21
			Officers	1565	3	14	.08	.09	.11	.14	-.03	.25
Conscientiousness												
	All	Global	All	38798	51	92	.19	.20	.25	.16	.06	.45
	Civilian	Global	All	9491	31	46	.11	.12	.14	.13	.00	.28
			Lower Level	723	4	4	.10	.11	.12	.06	N/A	N/A
			University Students	1362	7	7	.15	.17	.20	.10	.11	.28
			Managers/Executives	7406	21	35	.10	.11	.13	.14	-.02	.29
	Military	Global	All	29307	20	46	.21	.22	.29	.15	.11	.47
			Academy Trainees	19012	13	32	.25	.27	.35	.14	.18	.52
			Officers	1306	5	6	.16	.17	.22	.15	.05	.40
	All	Achievement	All	37846	31	79	.14	.15	.20	.11	.07	.33
		Motivation										
	Civilian	Achievement	All	6269	7	15	.16	.17	.23	.08	.16	.31
		Motivation	Managers/Executives	2620	4	9	.13	.13	.18	.07	.13	.23
	Military	Achievement	All	31577	24	64	.14	.15	.19	.12	.05	.33
		Motivation	NCOs	9193	6	18	.13	.14	.18	.08	.10	.27
			Academy Trainees	19656	13	33	.14	.15	.19	.12	.04	.34
			Officers	2728	5	13	.14	.15	.20	.17	.01	.40
	All	Orderliness	All	9700	15	24	.03	.04	.04	.11	-.09	.17
	Civilian	Orderliness	All	2770	4	7	.04	.04	.02	.13	-.14	.18
	Military	Orderliness	All	6930	11	17	.03	.04	.05	.10	-.07	.16
			NCOs	3988	5	8	.07	.07	.09	.07	.03	.15
	All	Self-Control	All	3805	12	18	.03	.03	.03	.12	-.09	.15
	Civilian	Self-Control	All	1131	3	5	-.04	-.04	-.05	.13	-.20	.09
	Military	Self-Control	All	2674	9	13	.05	.06	.07	.09	.00	.13
			Academy Trainees	1776	5	7	.05	.05	.07	.08	.01	.12
	All	Responsibility	All	13898	18	43	.08	.09	.12	.11	-.01	.25

Trait	Population	Facet	Leadership Level	N	k_s	k_c	r	$\rho_{(c)}$	$\rho_{(b)}$	$SD_{\rho(b)}$	80 % CV	
											lower	upper
	Civilian	Responsibility	All	2374	3	5	.16	.16	.22	.10	.10	.34
	Military	Responsibility	All	11524	15	38	.07	.07	.10	.11	-.01	.21
			NCOs	4889	6	10	.08	.08	.11	.08	.02	.20
			Academy Trainees	5564	7	16	.06	.06	.09	.11	-.03	.21
	Military	Honesty-Virtue	All	5423	10	16	.01	.01	.02	.07	-.05	.08
			NCOs	3053	3	6	.03	.03	.03	.05	.01	.06
			Academy Trainees	1632	4	5	-.02	-.02	-.02	.05	N/A	N/A
			Officers	738	3	5	.01	.01	.02	.13	-.12	.15
	All	Discipline	All	33285	23	64	.04	.04	.05	.14	-.11	.22
	Civilian	Discipline	All	1521	3	5	-.01	-.02	-.02	.19	-.25	.20
	Military	Discipline	All	31764	20	59	.04	.04	.06	.13	-.10	.22
			NCOs	10094	6	20	.06	.06	.08	.09	-.03	.19
			Academy Trainees	20428	12	34	.04	.04	.06	.14	-.12	.23
Emotional Stability												
	All	Global	All	44441	49	99	.08	.09	.11	.12	-.04	.25
	Civilian	Global	All	9845	30	47	.10	.11	.14	.13	-.01	.29
			Line Managers	723	4	4	.12	.12	.14	.09	.08	.21
			University Students	1289	6	6	.12	.14	.16	.08	.11	.21
			Managers/Executives	7833	20	37	.10	.10	.14	.15	-.03	.30
	Military	Global	All	34596	19	52	.07	.08	.10	.12	-.05	.24
			Academy Trainees	24858	14	40	.04	.05	.06	.11	-.08	.19
			Officers	925	4	5	.07	.07	.10	.07	N/A	N/A
	All	Optimism	All	9939	17	26	.11	.11	.18	.10	.07	.29
	Civilian	Optimism	All	1720	4	6	.18	.18	.28	.11	.15	.41
	Military	Optimism	All	8219	13	20	.09	.10	.16	.08	.07	.24
			NCOs	2446	3	5	.10	.10	.16	.06	.10	.21
			Academy Trainees	5169	7	10	.09	.09	.15	.08	.07	.23
			Officers	604	3	5	.12	.13	.20	.16	.03	.36
	All	Emotional Control - Volatility	All	18156	19	38	.04	.05	.07	.16	-.13	.26
	Civilian	Emotional Control - Volatility	All	4912	7	15	.07	.08	.11	.17	.02	.30
			Managers/Executives	2574	3	8	.04	.04	.06	.18	-.16	.28
	Military	Emotional Control - Volatility	All	13244	12	23	.03	.04	.05	.15	-.14	.23
			NCOs	5964	4	11	.05	.05	.07	.08	-.01	.15
			Academy Trainees	6405	6	9	.03	.03	.03	.18	-.19	.25

Trait	Population	Facet	Leadership Level	N	k_s	k_c	r	$\rho_{(c)}$	$\rho_{(b)}$	$SD_{\rho_{(b)}}$	80 % CV	
											lower	upper
	All	Stress Resistance - Anxiety	All	20831	20	45	.07	.07	.11	.11	-.02	.23
	Civilian	Stress Resistance - Anxiety	All	4317	8	18	.08	.08	.13	.14	-.04	.29
			Managers/Executives	2794	5	13	.03	.04	.06	.12	-.08	.19
	Military	Stress Resistance - Anxiety	All	16514	12	27	.06	.07	.10	.10	-.01	.22
			NCOs	1371	3	4	.06	.07	.11	.03	N/A	N/A
			Academy Trainees	14649	8	21	.06	.07	.11	.10	-.01	.22
Other Dimensions of Personality												
	Military	Energy	All	24507	18	40	.18	.20	.25	.14	.08	.43
			NCOs	4889	6	10	.08	.08	.11	.06	.05	.17
			Academy Trainees	18623	10	26	.20	.22	.28	.12	.13	.44
	Military	Narcissism	All	7613	7	10	.01	.01	.01	.07	-.06	.08
	All	Military Identification	All	7672	5	13	.04	.04	.05	.06	-.01	.11
	All	Self-Efficacy	All	18565	22	43	.12	.13	.17	.09	.07	.26
	Civilian	Self-Efficacy	All	2485	6	10	.14	.15	.18	.07	.14	.22
	Military	Self-Efficacy	All	16080	16	33	.12	.13	.16	.09	.06	.26
			NCOs	5846	4	11	.08	.09	.11	.06	.05	.17
			Academy Trainees	9486	11	19	.14	.15	.19	.09	.10	.29
	Military	Humility	All	2399	3	5	-.01	-.01	-.01	.08	-.10	.08
	Military	Hardiness	All	29920	8	32	.12	.12	.17	.11	.04	.30

Note: N = total sample size. k_s = number of studies. k_c = number of correlations/effect sizes. r = uncorrected correlation. $\rho_{(c)}$ = operational meta-analytic value (corrected for unreliability and range restriction in the criterion). $\rho_{(b)}$ = Theoretical meta-analytic value (corrected for unreliability and range restriction in both the criterion and predictor). $SD_{\rho_{(b)}}$ = variance for the fully corrected correlation. 80% CV = 80% credibility interval around the estimate.

Table 4. Relationships between Personality Dimensions and Leader Emergence

Leadership											80 % CV	
Trait	Population	Facet	Level	N	k_s	k_c	r	$\rho_{(c)}$	$\rho_{(b)}$	$SD_{\rho_{(b)}}$	lower	upper
Extraversion												
	All	Global	All	14419	18	25	.11	.12	.15	.11	.03	.28
	Civilian	Global	All	8696	13	17	.14	.15	.19	.07	.12	.26
	Military	Global	All	5723	5	8	.08	.08	.10	.13	-.06	.25
	All	Assertiveness	All	3872	11	12	.14	.14	.17	.10	.06	.29
	Civilian	Assertiveness	All	775	5	6	.21	.21	.26	.15	.09	.42
	Military	Assertiveness	All	3097	6	6	.12	.12	.15	.07	.08	.23
Agreeableness												
	All	Global	All	15085	20	27	.00	.00	-.01	.10	-.12	.11
	Civilian	Global	All	8755	14	18	-.01	-.01	-.02	.10	-.12	.09
	Military	Global	All	6330	6	9	.01	.01	.01	.10	-.11	.13
Openness												
	All	Global	All	15303	19	26	.05	.05	.06	.13	-.10	.23
	Civilian	Global	All	8306	12	16	.05	.05	.06	.12	-.08	.21
	Military	Global	All	6997	7	10	.06	.06	.07	.15	-.11	.25
Conscientiousness												
	All	Global	All	15981	20	27	.07	.07	.09	.08	.00	.18
	Civilian	Global	All	8696	13	17	.07	.08	.10	.07	.03	.17
	Military	Global	All	7285	7	10	.05	.06	.07	.09	-.03	.18
	All	Achievement Motivation	All	2058	6	6	.11	.12	.16	.12	.02	.29
	All	Orderliness	All	426	3	3	.04	.04	.05	.20	-.18	.29
	All	Self-Control	All	1305	3	3	-.03	-.03	-.04	.08	-.11	.04
	All	Responsibility	All	1133	4	5	.20	.22	.28	.17	.09	.48
Emotional Stability												
	All	Global	All	15499	20	27	.06	.07	.10	.10	-.02	.21
	Civilian	Global	All	8696	13	17	.08	.08	.11	.09	.01	.21
	Military	Global	All	6383	7	10	.04	.04	.07	.11	-.06	.20

Note: N = total sample size. k_s = number of studies. k_c = number of correlations/effect sizes. r = uncorrected correlation. $\rho_{(c)}$ = operational meta-analytic value (corrected for unreliability and range restriction in the criterion). $\rho_{(b)}$ = Theoretical meta-analytic value (corrected for unreliability and range restriction in both the criterion and predictor). $SD_{\rho_{(b)}}$ = variance for the fully corrected correlation. 80% CV = 80% credibility interval around the estimate.

Table 5. Relationships between Personality Dimensions and Transformational Leadership

Trait	Population	Facet	Leadership Level	N	k_s	k_c	r	$\rho_{(c)}$	$\rho_{(b)}$	$SD_{\rho(b)}$	80 % CV	
											lower	upper
Extraversion												
	All	Global	All	5730	21	33	.15	.16	.20	.14	.05	.34
	Civilian	Global	All	3374	18	29	.17	.18	.23	.17	.04	.41
	Military	Global	All	2356	3	4	.13	.14	.16	.04	.13	.18
	All	Assertiveness	All	2777	8	19	.13	.14	.18	.15	.01	.34
	Civilian	Assertiveness	All	6642	7	19	.16	.17	.24	.13	.08	.40
	Military	Assertiveness	All	737	3	7	.03	.03	.04	.15	-.11	.18
	All	Sociability	All	1775	5	15	.06	.07	.10	.23	-.17	.36
	Civilian	Sociability	All	1070	4	9	.15	.16	.22	.19	.01	.43
Agreeableness												
	All	Global	All	6147	22	35	.12	.12	.16	.20	-.07	.40
	Civilian	Global	All	3791	19	31	.17	.18	.23	.23	-.04	.50
	Military	Global	All	2356	3	4	.04	.04	.06	.04	.05	.07
	All	Empathy	All	2321	4	13	.12	.12	.17	.13	.03	.32
	Civilian	Empathy	All	1872	2	9	.13	.14	.20	.13	.05	.35
	Military	Empathy	All	449	2	4	.05	.06	.08	.07	N/A	N/A
Openness												
	All	Global	All	5507	18	30	.08	.08	.10	.13	-.04	.24
	Civilian	Global	All	3049	15	25	.09	.10	.12	.17	-.07	.31
	Military	Global	All	2356	3	5	.04	.04	.06	.04	.05	.07
	All	Intellectual Efficiency	All	1122	3	5	.04	.05	.06	.15	-.11	.24
	All	Culture aspect*	All	1086	3	10	.13	.14	.18	.17	.00	.36
Conscientiousness												
	All	Global	All	6022	21	35	.15	.16	.23	.18	.02	.43
	Civilian	Global	All	3666	18	31	.12	.13	.17	.19	-.04	.39
	Military	Global	All	2356	3	4	.20	.22	.31	.11	.17	.45
	All	Achievement Motivation	All	968	4	6	.00	.00	.00	.14	-.15	.14
Emotional Stability												
	All	Global	All	6082	22	35	.08	.08	.12	.13	-.03	.26
	Civilian	Global	All	3673	18	30	.10	.10	.15	.15	-.01	.30
	Military	Global	All	2409	4	5	.05	.05	.07	.08	-.02	.16
	All	Optimism	All	3228	3	7	.10	.10	.16	.09	.06	.27

All	Emotional Control - Volatility	All	1256	3	6	.09	.09	.14	.14	-.02	.30
All	Stress Resistance - Anxiety	All	5622	4	13	.06	.07	.11	.10	-.01	.22

Note: Transformational leadership is focused the interpersonal aspects of leading a group. N = total sample size. k_s = number of studies. k_c = number of correlations/effect sizes. r = uncorrected correlation. $\rho_{(c)}$ = operational meta-analytic value (corrected for unreliability and range restriction in the criterion). $\rho_{(b)}$ = Theoretical meta-analytic value (corrected for unreliability and range restriction in both the criterion and predictor). $SD_{\rho_{(b)}}$ = variance for the fully corrected correlation. 80% CV = 80% credibility interval around the estimate. *Naming convention based on Woo et al., 2014. N/A = unable to calculate credibility interval due to negative variance of the corrected correlation.

Table 6. Relationships between Personality Dimensions and Transactional Leadership

Trait	Population	Facet	Leadership	N	k_s	k_c	r	$\rho_{(c)}$	$\rho_{(b)}$	$SD_{\rho(b)}$	80 % CV	
			Level								lower	upper
Extraversion												
	All	Global	All	4560	16	34	.10	.11	.14	.15	-.03	.30
	Civilian	Global	All	3862	13	29	.11	.12	.15	.16	-.03	.32
	Military	Global	All	698	3	5	.07	.08	.09	.10	.02	.17
	All	Assertiveness	All	9364	7	19	.14	.15	.19	.09	.09	.29
	Civilian	Assertiveness	All	1549	3	9	.18	.19	.24	.15	.07	.41
	Military	Assertiveness	All	7815	4	10	.13	.14	.18	.07	.11	.26
	All	Sociability	All	4387	4	7	.02	.02	.03	.05	-.01	.06
Agreeableness												
	All	Global	All	3954	15	32	.05	.05	.05	.14	-.08	.19
	Civilian	Global	All	2688	12	25	.05	.05	.05	.14	-.08	.19
	Military	Global	All	1266	3	7	.04	.04	.05	.14	-.09	.20
	All	Empathy	All	3490	3	14	.06	.06	.09	.08	.02	.16
	Military	Empathy	All	449	2	4	.05	.06	.08	.07	N/A	N/A
	All	Cooperation	All	4757	5	14	.03	.04	.04	.19	-.20	.28
Openness												
	All	Global	All	5131	16	35	.09	.10	.13	.08	-.04	.30
	Civilian	Global	All	3717	12	28	.08	.08	.10	.09	-.06	.26
	Military	Global	All	1414	4	7	.14	.16	.20	.07	.04	.35
	All	Intellectual Aspect*	All	5320	3	10	.06	.07	.09	.04	-.05	.24
	All	Culture aspect*	All	4083	3	12	.01	.01	.02	.05	-.11	.15
Conscientiousness												
	All	Global	All	5177	16	35	.13	.14	.18	.19	-.04	.40
	Civilian	Global	All	3717	12	28	.12	.13	.16	.21	-.08	.41
	Military	Global	All	1460	4	7	.16	.17	.21	.12	.08	.35
	All	Achievement	All	6661	7	16	.12	.13	.17	.08	.09	.25
	All	Motivation	All	1221	3	5	.10	.11	.14	.09	.07	.22
	All	Orderliness	All	1221	3	5	.10	.11	.14	.09	.07	.22
	All	Responsibility	All	3900	4	9	.13	.14	.19	.14	.02	.37
	All	Discipline	All	3933	3	6	-.01	-.01	-.02	.18	-.24	.20
Emotional Stability												
	All	Global	All	4415	15	33	.12	.13	.16	.18	-.04	.37
	Civilian	Global	All	3717	12	28	.11	.11	.15	.19	-.07	.37

Military	Global	All	698	3	5	.17	.19	.22	.09	.16	.28
All	Optimism	All	927	4	6	.14	.16	.24	.07	N/A	N/A
All	Emotional Control - Volatility	All	4091	5	10	.05	.06	.08	.11	-.05	.21
All	Stress Resistance - Anxiety	All	905	3	6	.03	.03	.05	.15	-.11	.21

Note: Transactional leadership is focused on completing the tasks and accomplishing the goals of the group. N = total sample size. k_s = number of studies. k_c = number of correlations/effect sizes. r = uncorrected correlation. $\rho_{(c)}$ = operational meta-analytic value (corrected for unreliability and range restriction in the criterion). $\rho_{(b)}$ = Theoretical meta-analytic value (corrected for unreliability and range restriction in both the criterion and predictor). $SD_{\rho(b)}$ = variance for the fully corrected correlation. 80% CV = 80% credibility interval around the estimate. *Naming convention based on Woo et al., 2014. N/A = unable to calculate credibility interval due to a negative variance of the corrected correlation.

Table 7. Relationships between Personality Dimensions and Physical Fitness

Trait	Population	Facet	Leadership Level	N	k _s	k _c	r	ρ _(c)	ρ _(b)	SD _{ρ(b)}	80 % CV	
											lower	upper
Extraversion	Military	Global	All	706	3	4	.11	.12	.16	.04	N/A	N/A
		Assertiveness	All	7532	6	17	.04	.05	.06	.09	-.03	.15
Agreeableness	Military	All	All	8492	6	28	-.02	-.02	-.03	.10	-.13	.08
		Empathy	All	2957	5	9	.00	.00	.00	.08	-.08	.08
		Generosity	All	2124	3	4	-.02	-.02	-.03	.07	-.10	.04
		Cooperation	All	2695	5	11	-.04	-.04	-.05	.12	-.18	.08
Openness	Military	Tolerance	All	1999	4	7	-.03	-.03	-.04	.12	-.18	.09
Conscientiousness	Military	Global	All	942	4	5	.06	.06	.08	.11	-.02	.17
		Achievement	All	6207	6	16	.09	.09	.12	.10	.02	.23
		Motivation	All	2241	5	5	.05	.05	.07	.04	N/A	N/A
		Orderliness	All	652	3	3	-.03	-.03	-.04	.12	-.17	.09
		Self-Control	All	2830	6	8	.02	.02	.03	.07	-.03	.09
		Responsibility	All	5849	6	14	.03	.03	.04	.09	-.05	.13
		Discipline	All	5849	6	14	.03	.03	.04	.09	-.05	.13
Emotional Stability	Military	Global	All	759	3	5	.06	.07	.10	.12	-.02	.22
		Stress Resistance - Anxiety	All	3346	4	11	.06	.06	.09	.21	-.17	.35
		Emotional Control - Volatility	All	4071	5	8	-.01	-.01	-.01	.12	-.15	.13
		Optimism	All	2556	5	7	.01	.01	.01	.09	-.08	.10
Other Dimensions of Personality												
	Military	Energy	All	3872	6	11	.33	.36	.47	.20	.22	.72
		Self-Efficacy	All	3255	4	7	.08	.08	.11	.07	.04	.18

Note: N= total sample size. k_s = number of studies. k_c = number of correlations/effect sizes. r = uncorrected correlation. $\rho_{(c)}$ = operational meta-analytic value (corrected for unreliability and range restriction in the criterion). $\rho_{(b)}$ = Theoretical meta-analytic value (corrected for unreliability and range restriction in both the criterion and predictor). $SD_{\rho(b)}$ = variance for the fully corrected correlation. 80% CV = 80% credibility interval around the estimate. N/A = unable to calculate credibility interval due to negative variance of the corrected correlation.

Table 8. Relationships between Personality Dimensions and Intentions to Remain in the Military

Trait	Population	Facet	Leadership Level	N	k _s	k _c	r	ρ _(c)	ρ _(b)	SD _{ρ(b)}	80 % CV	
											lower	upper
Extraversion												
	Military	Global	All	27639	12	23	.14	.15	.19	.16	-.02	.40
		Assertiveness	All	21025	5	16	.13	.14	.17	.15	-.02	.36
Agreeableness												
	Military	All	All	16099	7	23	.16	.17	.24	.22	.04	.21
		Empathy	All	7007	6	8	.24	.26	.36	.17	.03	.17
		Generosity	All	7026	5	6	.13	.14	.20	.18	.03	.18
		Cooperation	All	1607	4	6	-.01	-.01	-.01	.08	.06	.05
Openness												
	Military	Global	All	10245	6	19	0	.01	.01	.06	.04	.05
Conscientiousness												
	Military	Global	All	561	3	3	.00	.01	.00	.24	-.29	.30
		Achievement	All	19297	6	14	.23	.25	.33	.13	.16	.49
		Motivation	All	6859	6	6	.00	.00	.00	.07	-.08	.08
		Orderliness	All	7100	6	7	.08	.09	.12	.07	.05	.20
		Responsibility	All	19223	6	13	.17	.18	.27	.10	.14	.40
		Discipline	All	19223	6	13	.17	.18	.27	.10	.14	.40
Emotional Stability												
	Military	Stress Resistance - Anxiety	All	2670	4	9	.00	.00	.00	.15	-.18	.18
		Emotional Control - Volatility	All	13239	5	9	.18	.20	.30	.16	.09	.51
		Optimism	All	6998	5	7	.18	.19	.30	.14	.12	.48
Other Dimensions of Personality												
	Military	Energy	All	8089	6	9	.11	.12	.16	.09	.05	.27
		Narcissism	All	6315	3	3	-.06	-.06	-.09	.04	-.14	-.04
		Self-Efficacy	All	12299	4	7	.21	.23	.30	.11	.16	.44
		Humility	All	1763	3	3	-.09	-.10	-.13	.05	-.16	-.09

Note: N = total sample size. k_s = number of studies. k_c = number of correlations/effect sizes. r = uncorrected correlation. $\rho_{(c)}$ = operational meta-analytic value (corrected for unreliability and range restriction in the criterion). $\rho_{(b)}$ = Theoretical meta-analytic value (corrected for unreliability and range restriction in both the criterion and predictor). $SD_{\rho(b)}$ = variance for the fully corrected correlation. 80% CV = 80% credibility interval around the estimate.

REVIEW OF EXISTING ARMY DOCTRINE AND FOCUS GROUPS WITH SUBJECT MATTER EXPERTS

Background

Although the results of the meta-analysis provide useful information about the relationship between personality traits and leadership, these results may not adequately represent the characteristics of successful officers in the Army. First, although the analyses included both military and civilian samples, the majority of the studies identified in our literature review were from civilian samples. Consequently, when examining the effects in combined civilian and military samples, it is possible that the results in the civilian samples masked important effects in the military. In addition, there were not a sufficient number of studies to examine the military context as a moderator of the correlation between personality dimensions and leadership performance in all cases. As a result, it is possible that the meta-analysis missed some characteristics of leaders that are important for Army officers but not necessarily for civilian leaders.

To address this issue, we next conducted a review of existing Army doctrine on leadership. The goal of this step was to identify the characteristics that the Army views as important for officers, even if they weren't represented in the meta-analysis. Second, building on both the meta-analytic results and the review of Army doctrine, we conducted focus groups with subject matter experts (SMEs) to get a different perspective on the characteristics that make officers successful. Current officers with experience in their roles may be able to identify additional characteristics that significantly influence their performance in their day-to-day tasks. Therefore, we summarized the results of the meta-analysis and the review of Army doctrine and asked SMEs (i.e., current officers) to identify additional characteristics that are required to make officers successful.

Review of Existing Army Doctrine

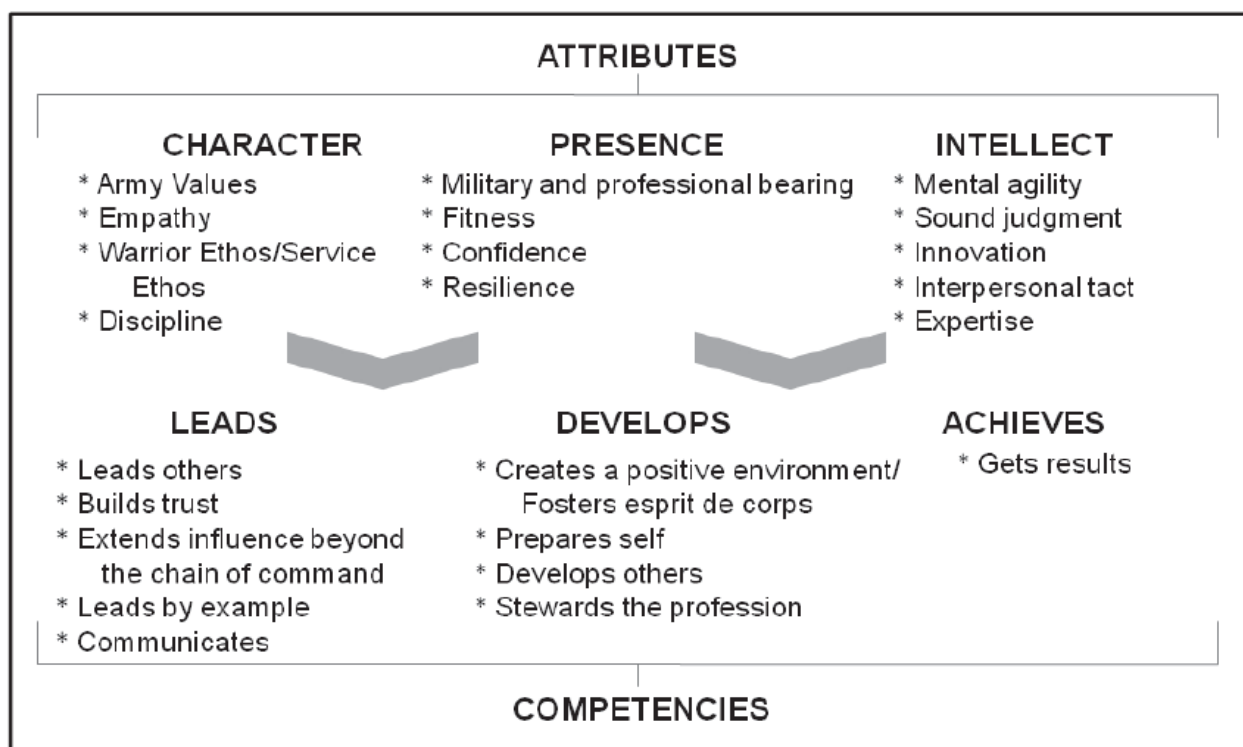
The first step was to conduct a review of existing Army doctrine. This involved reviewing publications on officer characteristics including Army Doctrine Publication (ADP) 6-22 on *Army Leadership* (Department of the Army, 2012), Field Manual (FM) 6-22 on *Leader Development* (Department of the Army, 2015), and descriptions of the required officer characteristics provided by 18 of the Basic Branches in the Army¹ (<https://vbo.army.mil/event?eid=833cab2b-0d7f-42b9-ae2-e147774f224f&roomId=9f56e343-8ee8-4c36-bcdb-9959364f9156>). In addition, we also reviewed a previously published job analysis of Army officer performance requirements (Paullin et al., 2011) and similar research conducted by the U.S. Marine Corps to identify the characteristics of individuals who are successful in Officer Candidate School (Harvey et al., 2018). From each of these sources, we searched for the personality characteristics that were identified as important for successful officers.

We first examined the Army leadership requirements model as described in ADP 6-22 and FM 6-22. This model describes “the expectations that the Army wants leaders to meet”

¹Here, only a subset of the Basic Branches published detailed descriptions of their branch talent priorities in 2022.

(Department of the Army, 2012, p. 5) and is shown in Figure 1. As shown in this figure, this model organizes the leader requirements into attributes and competencies. The attributes describe what a leader should be while the competencies describe what they should do (Department of the Army, 2012). For the purposes of this review, we focused on the leader attributes because these are most closely related to personality characteristics. The three broad attributes described in the model include Character, Presence, and Intellect. Within each of these broad attributes, the model also describes several narrower attributes such as Empathy and Discipline (under Character), Confidence and Resilience (under Presence) and Mental Agility and Interpersonal Tact (under Intellect). Together, these attributes describe the characteristics that the Army views as important for its leaders.

Figure 1. Army Leadership Requirements Model (Adapted from ADP 6-22)



Next, we examined the behaviors desired for officers in each of the 18 branches that published this information in 2022. The majority of the Basic Branches listed their desired characteristics for officers on the virtual branch outreach (VBO) site (<https://vbo.army.mil/event?eid=833cab2b-0d7f-42b9-aef2-e147774f224f&roomId=9f56e343-8ee8-4c36-bcdb-9959364f9156>). An example of the knowledge, skills, behaviors, and talent priorities provided by the Infantry Branch is provided in Figure 2.

Figure 2. Desired Characteristics for Officers in the Infantry Branch




Infantry

Year Group 2022

INTELLIGENCES: Bodily-Kinesthetic, Interpersonal, Spatial

SKILLS: Infantry officers must be able to operate in some of the most politically, economically, and environmentally adverse situations while at different threat levels. In order to operate in such environments, Infantry Officers must possess the highest levels of mental toughness, problem solving ability, and physical fitness. Using creativity and sound judgment, they must have the ability to devise and prioritize solution sets rapidly, motivate and employ Soldiers, and have an innate ability to adapt to fluid situations when facing any enemy across the entire threat spectrum. Additionally, Infantry Officers must have the ability to discriminate an action out of the norm and respond with the appropriate level of action. They must also demonstrate consistent command of these skills over extended periods of time.

KNOWLEDGE: The Infantry branch desires officers with academic backgrounds that span the entire spectrum of disciplines and majors offered at our nation's undergraduate institutions. Broad individual experiences contribute to the success of the branch.

> **RELEVANT TRAINING / EXPERIENCE:** Leadership role in athletics / student government; Cadet Troop Leading Time / Leader Development Time (CTLT / CLDT) with Infantry units. Overseas Academic Enrichment Program. Prior enlisted service in a Maneuver Branch. Contact Sports / Triathlons / Ultra Marathon Competitor; Coaching / Mentoring Experience (not all inclusive).

> **RELEVANT CERTIFICATIONS / ACCREDITATIONS:** EMT / First Responder Training; Cross Fit Instructor; PADI Cert; Free Fall; highest level of scouting (not all inclusive).

BEHAVIORS: (In addition to foundational)

> ADAPTABLE	> DISCIPLINED	> HONORABLE	> MENTALLY AGILE
> ASSERTIVE	> DUTIFUL	> INNOVATIVE	> PERSONAL COURAGE
> CONFIDENT	> ETHICAL / MORAL	> INTEGRITY	> RESPECTFUL
> DILIGENT	> FIT (PHYS / MENT)	> LOYAL	> SELFLESS SERVICE

TALENT PRIORITIES:

- MENTALLY TOUGH:** Stress tolerant and emotionally mature. Performs well even under extreme psychological duress.
- PROBLEM SOLVER:** Able to choose between best practices and unorthodox approaches to reach a solution. Accomplishes the task.
- PHYSICALLY FIT:** Physically tough, gritty, and tenacious. Performs well even under extreme physiological duress. Committed to a lifestyle of physical fitness.
- COMMUNICATOR:** Precise, efficient, and compelling in both written and spoken word.
- INNOVATIVE:** Creative, inquisitive, and insightful. Easily identifies new solutions and catalyzes change.
- MULTI-TASKER:** Rapidly processes and prioritizes multiple demands simultaneously. Takes appropriate action.

Certified by Branch Commandants, Approved by CAC Commander, Distributed by DCS-G1 DMPM, March 2021

12

To summarize the characteristics of successful officers across all 18 of these Basic Branches, we examined the frequency of each desired characteristic reported by the Basic Branches. For this review, we focused on the category of behaviors because these characteristics are most closely associated with personality traits rather than other cognitive characteristics (e.g., cognitive ability). The number of behaviors listed for each branch ranged from 5 (Armor) to 24 (Military Police). The frequencies with which each of the behaviors were mentioned by the Basic Branches are reported in Table 9. The behaviors described by the branches are listed in alphabetical order. In addition, to reduce the size of the table, only those behaviors that were mentioned by five or more of the Basic Branches are included in this table. As shown, several characteristics were mentioned by multiple Basic Branches. For example, Problem Solving was mentioned as a desired behavior by 15 of the 18 Basic Branches, Adaptable was mentioned by 11 of the Basic Branches, and Collaborative was mentioned by 10. The percentages of the Basic Branches that described each of these behaviors as desired for officers in their branch are also reported in Table 9.

Table 9. Summary of the Talent Priorities (Behaviors) Desired by the Basic Branches in the Army

Behaviors	AD	AG	AR	AV	CM	CY	EN	FA	FM	IN	MI	MP	MS	OD	EOD	QM	SC	TC	Total	Percentage
Adaptable	X			X	X		X		X	X		X	X			X	X	X	11	61%
Agile		X		X	X	X				X			X				X		7	39%
Assertive	X				X				X	X		X							5	28%
Collaborative	X	X			X	X					X		X	X	X	X	X		10	56%
Committed	X				X		X		X			X	X	X	X				8	44%
Confident					X	X		X		X							X		5	28%
Cooperation/Teamwork			X						X			X		X	X				5	28%
Critical Thinking	X	X		X	X	X		X			X		X	X	X		X	X	12	67%
Customer Focused		X							X				X	X	X	X	X		7	39%
Detail Focused		X			X	X	X	X			X	X		X	X				9	50%
Diligent					X	X			X	X	X		X						6	33%
Disciplined				X	X	X			X	X	X	X	X			X	X		10	56%
Ethical/Moral									X	X		X	X			X			5	28%
Expert					X	X	X		X			X	X	X	X	X			9	50%
Fit (Physical)	X	X			X		X			X	X	X							7	39%
Flexible		X		X	X			X						X	X	X			7	39%
Hard Working		X			X				X					X	X	X			6	33%
Initiative	X		X	X		X					X	X		X	X				8	44%
Innovative				X	X	X	X	X		X	X		X			X	X		10	56%
Perceptive				X		X	X				X		X	X	X				7	39%
Precise	X	X		X	X	X			X				X	X	X	X			10	56%
Problem Solving	X	X	X	X	X	X	X	X	X			X	X	X	X	X		X	15	83%
Proactive	X	X			X		X							X	X	X	X		8	44%
Resilient	X	X				X		X			X	X					X	X	8	44%
Responsible	X								X					X	X	X			5	28%
Stress Tolerant	X			X	X								X	X	X	X			7	39%

Note: AD = Air Defense Artillery; AG = Adjutant General; AR = Armor; AV = Aviation; CM = Chemical Corps; CY = Cyber Corps; EN = Engineers Corps; FA = Field Artillery; FM = Finance; IN = Infantry; MI = Military Intelligence Corps; MP = Military Police; MS = Medical Service Corps; OD = Ordnance; EOD = Explosive Ordnance Disposal; QM = Quartermaster; SC = Signal Corps; TC = Transportation Corps.

Finally, we also reviewed previous research focused on identifying the characteristics of successful officers. Here, the focus was on studies conducting job analyses of officers in the military, which provided a qualitative review of these characteristics and, therefore, could not be included in the meta-analysis described above. Paullin et al. (2011) conducted a job analysis to identify Army officer performance requirements while Harvey et al. (2018) focused on the characteristics of individuals who are successful in the U.S. Marine Corps Officer Candidate School. The personality traits identified as important for officers in these two qualitative reviews are shown in Table 10. Again, a number of personality dimensions were identified as important for officer performance.

Table 10. Results of Previous Qualitative Reviews of Officer Characteristics

Job Analysis of Army Officers (Paullin et al., 2011)	Characteristics of Marine Corps Officers (Harvey et al., 2018)
Achievement Motivation	Achievement
--	Adaptability
--	Competence
--	Decisiveness
Motivation to Lead	Dominance
Emotional Stability	Even Tempered
Initiative	Initiative
Openness	Intellectual Efficiency
--	Optimism
Health and Fitness Orientation	Physical Conditioning
--	Resilience
Conscientiousness	Responsibility
--	Sense of Purpose
--	Sociability
Team Orientation	Team Orientation
Learning Orientation	--
Self-Efficacy	--
Traditional Values	--

Note: Similar characteristics are shown next to each other in the table. "--" indicates that the characteristic was not identified in the corresponding review.

Focus Groups with SMEs

Given the wide range of potential characteristics identified in both the meta-analysis and the review of Army doctrine, we next conducted focus groups with SMEs to refine this list, verify our findings, and identify any additional characteristics that were not found in other sources. Before conducting the focus groups, we began by identifying an initial set of characteristics based on the previous reviews. To do so, we examined the most promising personality characteristics identified in the meta-analysis and the review of Army doctrine. We then used this initial list as a starting point for the focus groups with SMEs.

Focus Group Procedures

To conduct the focus groups, researchers coordinated with unit representatives from multiple Army installations to gain access to SMEs. For the purposes of this research, the target group of SMEs was junior officers in ranks O-1 to O-3. In addition, to get a more comprehensive view of officer characteristics, we requested officers from multiple MOS. In total, we conducted five focus groups with a total of 38 volunteer participants from multiple branches of the Army including the Infantry, Armor, Field Artillery, and Quartermaster branches, among others. All focus groups were conducted remotely at the request of the units.

As the focus groups progressed, the content was adapted to incorporate input from previous focus groups and to refine the list of potential personality dimensions. The initial focus groups began by asking participants to describe the role of an officer. Next, SMEs reviewed the initial list of personality dimensions that we identified based on the meta-analysis and the review of Army doctrine. This initial list also included definitions of each personality dimension to help facilitate the discussion. We then discussed the relevance of each dimension for Army Officers and asked the SMEs to identify any additional dimensions that were not found in our review. For the additional dimensions that were identified by the SMEs, we also asked them to define each dimension and describe the behaviors that are associated with it. After identifying a comprehensive list of personality characteristics, participants were asked to rank-order these characteristics from most to least important. Specifically, we asked the SMEs to identify the characteristics that are most important for officer performance. After the initial set of focus groups, the sessions began to focus less on describing the role of an officer and instead used the time to generate behavioral examples of the personality dimensions on our list. These behavioral examples helped to define the characteristics and provide ideas for writing statements in the subsequent scale development process.

Focus Group Results

Across all of the focus groups, several key themes emerged:

- Empathy was mentioned in several focus groups as an important characteristic of officers. Officers need to be able to empathize with their subordinates and understand the challenges they face to build unit cohesion and help to ensure Soldiers stay mentally and physically ready.
- It is essential to follow orders from senior leaders but officers also need to challenge the norms and stick up for their Soldiers when it is appropriate to do so. In this regard, officers should not be passive observers but active participants in the decision-making

process. Participants also emphasized that while following orders should be the priority, it may require officers to develop innovative solutions that both meet the intent of the orders and satisfy the needs of the Soldiers in their unit.

- Participants frequently talked about the need for officers to take responsibility, both for their own actions and for the actions of the Soldiers they are leading.
- It is important for officers to be assertive to communicate the needs of their units and persistent to ensure they have what they need.
- It is important for leaders to have “presence” but not to be domineering unless absolutely necessary.
- Officers also need to be confident both in themselves and their decisions.
- Not all characteristics will be equally important for officers in all branches. For example, physical conditioning was mentioned frequently by officers in the combat arms branches. In contrast, organization/order was mentioned more by officers in combat support branches.

As described above, participants were asked to rank order each dimension from most to least important for officer performance. Each group discussed these rankings and came to a consensus to identify a final order for the whole group. The rankings from the five focus groups that we conducted are summarized in Table 11. As shown in this table, there were a number of similarities across groups. For example, integrity and adaptability were consistently ranked among the top five most important characteristics for officers. Empathy was also ranked in the Top 10 in four of the five groups while resilience was rated in the top 10 in all five groups. There were also a number of differences across the sessions. For example, achievement was rated as important for some groups but not others. Responsibility was ranked 3rd for the first three groups but much lower for the last two. Similarly, team orientation was ranked as important for all of the groups but was considered much more important in some groups than in others. Nevertheless, the results of these rankings provide useful information about the most important characteristics of officers.

Table 11. Participants' Rank Order of Personality Characteristics for Officers

Rank Order	Focus Group #1	Focus Group #2	Focus Group #3	Focus Group #4	Focus Group #5
1	Integrity	Integrity	Self-Awareness	Integrity	Integrity
2	Adaptability	Team Orientation	Integrity	Empathy	Intellectual Efficiency
3	Responsibility (tied for 2 nd)	Responsibility	Responsibility	Communication	Achievement
4	Initiative	Dominance	Resilience	Adaptability	Adaptability
5	Resilience	Adaptability	Adaptability	Initiative	Resilience (tied for 4 th)
6	Order/ Organization	Resilience	Empathy	Persistence	Self-Awareness
7	Empathy	Selflessness	Achievement	Resilience	Initiative
8	Communication	Self-Awareness	Self-Efficacy	Even-Tempered	Dominance
9	Selflessness	Initiative	Dominance	Endurance	Empathy
10	Competence	Achievement	Tolerance	Dominance	Genuine
11	Caring	Assertiveness	Team Orientation	Physical Conditioning	Responsibility
12	Optimism	Self-Efficacy	Adjustment	Sociability	Persistent
13	Team Orientation	Intellectual Efficiency	Selflessness	Humility	Team Orientation
14	Commitment to Serve	Planning	Sociability	Team Orientation	Innovative
15	Physical Conditioning	Humility	Self-Efficacy	Disciplined Disobedience	Self-Efficacy
16	Self-Awareness	Tolerance	Even-Tempered	Intellectual Efficiency	Selflessness
17	Intellectual Efficiency	Curiosity	Persistence	Responsibility	Tolerance
18	Self-Efficacy	Even-Tempered	Optimism	Self-Efficacy	Organization/ Order
19	Sociability	Passionate	-- ^a	Selflessness	Even-Tempered
20	Curiosity	Approachable	-- ^a	Innovative	Sociability

^a This focus group suggested that the rest of the characteristics were much less important and stopped their ranking at 18.

Summary and Recommendations

To review the literature and identify the most important characteristics for an officer assessment, we conducted a meta-analysis of previous research, a review of Army doctrine, and focus groups with SMEs. Building on this research, we propose assessing the characteristics shown in Table 12. To select these characteristics, we tried to address the competing goals of a) assessing a comprehensive set of personality characteristics and b) ensuring an efficient measure that can be administered in the limited amount of time available for officer assessments. Therefore, we propose a maximum of 16 dimensions that can be administered in any one version

of the proposed assessment. This number of dimensions was selected based on previous research with the Tailored Adaptive Personality Assessment System (Drasgow et al., 2012), which is administered to enlisted Soldiers at the MEPS and typically includes 15-17 dimensions. Given the limited number of dimensions that can be assessed reliably in one administration, a number of characteristics that were identified as important in the current research (i.e., either in the meta-analysis, the review of Army doctrine, or the focus groups) had to be excluded. Table 12 provides the justification for the dimensions that were chosen as well as their definitions and example items.

There were various reasons why the other dimensions ranked in Table 11 were not included in the final list shown in Table 12. For example, some of the characteristics proposed by SMEs were either not defined consistently across focus groups (e.g., Disciplined Disobedience) or represented skills (e.g., Communication) or meta-cognition (e.g., Self-Awareness) that did not fit well in a broader personality framework. However, the main reason that many of the remaining dimensions were not included in Table 12 is that there was less evidence supporting their inclusion relative to the 16 dimensions that were chosen. For example, Order was another potential dimension mentioned by two of the focus groups. However, in one of those focus groups, this dimension was ranked low relative to many of the other dimensions. In addition, it was not mentioned consistently as a desired behavior in the basic branches and showed only weak correlations with many of the leadership performance dimensions in the meta-analysis. Similarly, Persistence seems like a promising dimension but was not mentioned frequently in the focus groups (or was ranked low when it was) or as a desired behavior by the basic branches. In addition, we did not find any validity evidence for this dimension in the meta-analysis. Due to the limited support for their inclusion, these and the other remaining dimensions were viewed as less important than the dimensions listed in Table 12.

Table 12. *Proposed Dimensions for the Officer Assessment*

Dimension	Description	Example Items	Justification
1. Integrity	High scoring individuals tend to behave in exemplary ways and try to do "what is right," even if doing so would be detrimental to their own interests.	• I have higher ethical standards than most people. • I try not to swear or behave badly in front of children. • I would never want to become rich by being dishonest.	• This characteristic was ranked #1 or #2 in all five focus groups.
2. Adaptability	High scoring individuals have the ability to adjust their thoughts and behaviors to meet the demands of the situation and tend to adjust quickly to changing circumstances.	• I quickly adapt to new situations. • I change my behavior to give me the greatest chance of success in any situation. • I believe I can adapt to any challenges that Army life offers.	• Consistently ranked as one of the top 5 most important characteristics for officers in the focus groups. • Mentioned as a desired trait in 61% of the basic branches.
3. Responsibility	High scoring individuals are dependable, reliable and make every effort to keep their promises.	• I am willing to make personal sacrifices to keep my promises. • I make an extra effort to show up to appointments on time. • When I make a mistake, I take responsibility for my actions.	• Frequently ranked as one of the top 5 most important characteristics for officers in the focus groups. • Correlated .10 with leader performance, .28 with leader emergence, .19 with transactional leadership, and .12 with intentions to stay in the meta-analysis.
4. Dominance	High scoring individuals tend to "take charge," have a strong leadership presence, and are often referred to by their peers as "natural leaders."	• After joining a group, I usually end up becoming the leader. • I enjoy supervising others. • When working on a team project, I like to take charge of things.	• Consistently ranked as one of the top 10 most important characteristics in the focus groups. • Conceptually, this facet is the most closely related to leadership. • Also a strong predictor in Army research with enlisted Soldiers.

5. Resilience	High scoring individuals bounce back from defeat or failures to focus on accomplishing their goals. They believe failures and bad situations are only temporary setbacks that can be quickly reversed.	• Even if I fail at something, I try to think about how I can do better next time. • I am able to take feedback and use it to improve myself. • If I don't succeed at first, I keep trying.	• Ranked in the top 10 most important characteristics in all five focus groups. • Mentioned as a desired trait in 44% of the basic branches.
6. Team Orientation	High scoring individuals prefer working in teams and helping people work together better.	• I enjoy working in teams much more than working by myself. • I feel that working in groups allows me to accomplish more. • I have a talent for making teams work well	• Ranked in the top 15 most important characteristics in all five focus groups. • Mentioned as a desired trait in 56% of the basic branches. • Also a strong predictor in Army research with enlisted Soldiers.
7. Empathy	High scoring individuals tend to care for others, treat them with respect, and consider their perspectives.	• I often consider the needs of others when making important decisions. • I listen to people's problems and try to help them if I can. • I try to treat everyone with kindness and respect.	• Consistently ranked as one of the top 10 most important characteristics in the focus groups.
8. Intellectual Efficiency	High scoring individuals are able to process information quickly and would be described by others as knowledgeable, astute, and intellectual.	• I am able to process information quickly. • I like to work on mentally challenging tasks. • It is easy for me to learn new things.	• Mentioned frequently as one of the most important characteristics in the focus groups. • Correlated .16 with leader performance in the meta-analysis. • Also a strong predictor in Army research with enlisted Soldiers.

9. Initiative	High scoring individuals tend to be self-motivated. They take action without the need to be prompted and complete tasks without the need for direct supervision.	• I tend to be highly self-motivated. • When I see a problem, I will address it without being told to. • I tend to finish my work tasks without much direct supervision.	• Mentioned frequently as one of the most important characteristics in the focus groups. • Mentioned as a desired trait in 44% of the basic branches.
10. Self-Efficacy	High scoring individuals are confident in their skills and abilities to accomplish any task that they take on.	• I am confident I can accomplish my goals no matter how difficult they may be. • I know I can excel in just about any situation. • I expect to be successful at whatever I do.	• Mentioned frequently as one of the most important characteristics in the focus groups. • Correlated .16 with leader performance and .30 with intentions to remain in the meta-analysis. • Identified as one of the most important characteristics in the job analysis conducted by Paullin et al. (2011).
11. Selflessness	High scoring individuals are generous with their time and resources.	• I contribute to charity regularly. • I am very generous with my time when I can help others. • It is important for me to work in a profession where I can help others.	• Mentioned frequently as one of the most important characteristics in the focus groups. • Correlated .20 with intent to stay in the meta-analysis.

12. Achievement	High scoring individuals are seen as hard working, ambitious, and resourceful.	• I make every effort to do more than what is expected of me. • I try to be the best at anything I do. • I tend to set high standards for myself.	• Correlated .20 with leader performance, .16 with leader emergence, .17 with transactional leadership, and .33 with intentions to stay in the meta-analysis. • Identified as one of the most important characteristics in the job analysis conducted by Paullin et al. (2011). • Mentioned as a desired trait in 33% of the basic branches (labeled “Diligent”). • Also one of the strongest predictors in Army research with enlisted Soldiers.
13. 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.	• I like to play high intensity sports. • I like to exercise. • I consider myself to be an athletic person.	• Mentioned as an important characteristics in the focus groups, particularly for combat arms officers. • Correlated .25 with leadership performance, .47 with physical fitness, and .16 with intentions to stay in the meta-analysis. • Mentioned as a desired trait in 39% of the Basic Branches. • Identified as one of the most important characteristics in the job analysis conducted by Paullin et al. (2011). • Also one of the strongest predictors in Army research with enlisted Soldiers.

14. Sociability	High scoring individuals tend to seek out and initiate social interactions.	• I enjoy talking with friends and neighbors. • I like to go out in big groups so I have more people to talk to. • I enjoy making friends and meeting new people.	• Correlated .11 with leader performance and .22 with transformational leadership in the meta-analysis. • Also a strong predictor in samples of non-commissioned officers (Nye et al., 2018).
15. Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.	• I am pretty happy with my life. • I have a positive outlook on life. • I tend to focus on the positive aspects of life even in the worst of situations.	• Correlated .20 with leader performance, .16 with transformational leadership, .24 with transactional leadership, and .30 with intentions to stay in the meta-analysis. • Also a strong predictor in samples of non-commissioned officers (Nye et al., 2018).
16. Even-Tempered	High scoring individuals tend to be calm and stable. They don't often exhibit anger, hostility, or aggression.	• I tend to handle difficult situations without getting emotional. • I can put up with more aggravation than most people. • It takes a lot to make me angry.	• Mentioned frequently as one of the most important characteristics in the focus groups. • Correlated .14 with transformational leadership and .30 with intentions to stay in the meta-analysis. • Mentioned as a desired trait in 39% of the Basic Branches.

REVIEW OF POTENTIAL METHODS FOR NON-COGNITIVE ASSESSMENTS

Literature Review

In this section, we review several potential methods for psychological measurement. Given that the goal of this effort was to develop a research plan to create a new personality assessment that can be administered in a computer adaptive testing (CAT) format, we focused our review on methods that could be implemented in such an assessment. This focus limited the scope of our review in that not all models fit these criteria. In addition, we also focused our review on potential methods that can mitigate the negative effects of faking in high-stakes settings. A key consideration for any personality measure that will be used in high-stakes testing is faking (Morgeson et al., 2007). Past research has shown that test takers can often discern the correct or socially desirable answers and, thus, increase or decrease their scores (White et al., 2008). This intentional distortion can severely undermine the utility of personality measures in personnel selection settings. Therefore, any measure that is intended for use in high-stakes settings will also need to address the important issue of faking.

Rapid Response Measurement (RRM)

RRM has recently been proposed as a method of reducing faking in high-stakes settings (Meade et al., 2020). To assess personality with this approach, test content is presented to test-takers in rapid succession. For example, Meade et al. (2020) used personality adjectives (e.g., “Talkative”) as the test stimuli and asked respondents to indicate whether each adjective was “like me” or “not like me.” Respondents are asked to make a choice as quickly as possible and are given a warning if they take longer than 2.5 seconds to respond. In addition, the response time is captured and used to weight individual responses such that faster responses are viewed as better indicators of an individual’s personality.

The key advantages of RRM are that it will be relatively quick to administer given the speeded nature of the test and can be resistant to faking. Regarding the latter point, the rapid administration of the RRM is presumably more difficult to fake because respondents have less time to consider the most socially desirable response option. Meade et al. (2020) provided evidence of this and found that faking was substantially lower on an RRM assessment than a more traditional Likert type format.

Despite these advantages, the RRM also has several apparent disadvantages in high-stakes settings. For example, as with all speeded tests, there are potential concerns with the construct validity of the assessment. Speeded tests include any test that requires individuals to respond rapidly to the test content such that many test-takers will not be able to respond to all of the items in the time allotted. Given the rapid response format, a speeded test will not only measure the construct of interest (e.g., personality), but will also assess how quickly individuals can process and respond to stimuli (Lord & Novick, 1968). Therefore, this format may unintentionally assess processing speed, which is a lower-order factor of cognitive ability (McGrew, 2009), in addition to the construct of interest. As evidence of this, Komar et al. (2010) found that a speeded test of personality resulted in less faking than a Likert-type measure, but only for individuals with lower cognitive ability. In other words, individuals with higher

cognitive ability were able to more quickly review the test content and select the most socially desirable response option, presumably due to their ability to process information more quickly. As a result, there are concerns that speeded tests will influence the psychometric properties of the assessment (Crocker & Algina, 1986) and several studies have indicated that speed and power (i.e., tests that provide sufficient time for examinees to respond to the items) tests are not equivalent (Rindler, 1979).

Another characteristic of the RRM that may limit the construct validity of this method is the response format, which may also be cognitively demanding for test-takers because they have to learn and remember specific responses. For example, respondents in the study conducted by Meade et al. (2020) pressed the “e” key to indicate that a personality adjective was “not like me” and the “i” key to indicate that the adjective was “like me.” Although respondents are given practice questions prior to the test, the ability to learn the required responses and recall them during the assessment, while also responding rapidly to each item, is strongly associated with cognitive ability (Nye et al., 2020). The cognitive load of the RRM is particularly important for the Army, which already uses the Armed Services Vocational Aptitude Battery (ASVAB), a measure of cognitive ability, for selection and classification. Therefore, any new measure that will be used for making selection or classification decisions would need to assess different individual characteristics and provide incremental validity over the ASVAB. Although the RRM has shown some promise, more research is needed to demonstrate that it would be useful in this context. At this time, this format does not appear ideal for a new measure of personality that will be used to make high-stakes personnel decisions. Instead, more research is needed to verify the utility and psychometric properties of this approach.

Two-Alternative Forced Choice (2AFC) Format

Another potential format for assessing non-cognitive characteristics is the two-alternative forced choice format. With this format, each item consists of two statements (e.g., “*I tend to work hard*” and “*I get along well with my coworkers*”) that are presented together in a pair. Test-takers are then asked to pick which of the two statements provides the best description of them. Conceptually, this format makes faking more difficult because respondents must choose the best option out of two statements rather than simply agreeing to an individual statement. Cao and Drasgow (2019) conducted a meta-analysis to examine the effectiveness of the forced choice format for reducing faking. Across 74 independent samples, these authors found that this format was useful for mitigating the effects of faking, with overall score inflation across low- and high-stakes conditions equal to $d = .06$. However, these authors also found several moderators of these effects. Forced choice measures that included multidimensional statement pairs (i.e., statements from different personality dimensions in each pair) that were matched on both social desirability and the extremity of the statements were more effective than unidimensional or unmatched pairs. This finding is consistent with other research conducted in operational settings, which has found little evidence of score inflation even when comparing scores obtained in high-stakes settings to scores collected for “research purposes only” (Drasgow et al., 2012; Trent et al., 2020).

Another important finding reported by Cao and Drasgow (2019) is that the method used to score the forced choice data is important. In the past, forcing respondents to choose between two response options resulted in ipsative scores, which are unsuitable for making personnel

selection and assignment decisions. However, recent advances in IRT have provided a way to overcome this limitation and obtain normative scores. Although several IRT models are available for scoring 2AFC items, the Multi-Unidimensional Pairwise Preference (MUPP) model (Stark, Chernyshenko, & Drasgow, 2005), which has been used successfully as part of the Tailored Adaptive Personality Assessment System (TAPAS; Drasgow et al., 2012), is the most widely used 2AFC IRT model in the military.

The 2AFC format has a number of advantages for use with officers in the U.S. Army. First, a substantial amount of research has supported the use of this format for personality assessment. In addition to the research cited above demonstrating that 2AFC measures can be highly resistant to faking (Cao & Drasgow, 2019; Drasgow et al., 2012; Trent et al., 2020), a substantial amount of research has demonstrated that 2AFC measures can predict performance outcomes in the military (see Drasgow et al., in progress, for a review). Based on this evidence, the TAPAS, which is a 2AFC personality measure, was recently designated as the principal tool for assessing personality in the Department of Defense (G. R. Cisneros, memorandum, March 29, 2022). Finally, due to the work on the TAPAS, a 2AFC measure can also be readily implemented in a computer adaptive testing format.

One disadvantage of the 2AFC format is that it requires large samples of individuals to pretest new statements before they can be used in this format. Pretesting is required to estimate the IRT parameters that are used to match statements based on their extremity. However, it is important to note that pretesting is required for CAT administration regardless of whether the forced choice methodology is used because the statement parameters are used to select the statements that are administered to test-takers. Therefore, despite the sample size requirements for pretesting, the 2AFC format seems promising for officer selection and assessment requirements. This is particularly true given the substantial amount of previous research on this format that has demonstrated that it can be resistant to faking while still showing validity for predicting important outcomes in the military.

Three Alternative Forced Choice (3AFC) Format

In the previous section, we described the potential advantages of the 2AFC format. However, over the last 10 years, there have been several developments in the psychometrics field that now enable forced choice formats involving more than two alternatives. For example, de la Torre et al. (2012) generalized Stark's (2002) model for pairwise preferences items to formats involving any number of response alternatives. Similarly, Drasgow et al. (in progress) demonstrated that the 3AFC responses can be decomposed into paired comparisons and scored using the MUPP model. Given these advances, it is now possible to use a forced choice format with three or more alternatives and to administer these items in an adaptive test. However, because increasing the number of alternative responses also potentially increases the cognitive load (i.e., and therefore the correlation with cognitive ability), we will focus on a 3AFC format in this review.

Because the 3AFC format still forces respondents to choose between multiple response alternatives, this format is likely to be resistant to faking in the same way as the 2AFC format. In addition to being fake resistant, 3AFC measures have several potential advantages over the

2AFC format. First, using a 3AFC format should increase the marginal and test-retest reliabilities of the assessment. Theoretically, increasing the number of response alternatives will also increase the amount of information that each item provides (see Brown & Maydeu-Olivares, 2011; de la Torre et al., 2012; Lee et al., 2019), which, in turn, translates into higher reliabilities. In fact, this was demonstrated in recent research supported by ARI, which showed that the marginal reliabilities of a 3AFC measure ranged from .75 to .85 (Drasgow et al., in progress). These reliabilities were higher than the reliability estimates for the same statements administered in a 2AFC format. Second, a 3AFC format would exponentially increase the number of potential items available for administration. Each item consists of three statements so the number of unique combinations is substantially larger than the number of potential pairs in the 2AFC format. Increasing the number of potential items can enhance test security, particularly when used in combination with adaptive testing (Drasgow et al., 2009). Third, the 3AFC format might lead to more positive test-taker reactions to items involving negatively worded statements. In the 3AFC format, respondents are asked to not only identify the statement that is “most like you” but also the statement that is “least like you.” Therefore, the ability to indicate that a negative statement is “least like you” may allow examinees to feel a greater psychological distance from the alternatives they believe describe them poorly. Finally, because 3AFC items provide more information (i.e., statements reflecting different personality facets) than single statement or 2AFC items, they may also require fewer overall items to achieve the same level of measurement precision (Lee et al., 2019).

Like the other models described here, the 3AFC format also has several disadvantages. For example, there is some evidence that the 3AFC format may be slightly less resistant to faking than the 2AFC format. When comparing the fakeability of these two approaches, Cao and Drasgow (2019) found that the standardized differences (i.e., Cohen’s d) between faking and honest conditions ranged from .00 to .22 (overall $d = .05$) for 2AFC measures. This was compared with Cohen’s d values ranging from .08 to .51 (overall $d = .37$) for measures with three or more response options. Despite the larger overall effects for measures with three or more alternatives, there are several caveats. First, Cao and Drasgow (2019) noted that many of the measures in their sample used tetrads (i.e., four alternatives) rather than the 3AFC format. In addition, they also noted that many of these measures included two positive and two negative statements, which may make it easier to identify the most socially desirable responses. Due to these differences, it is unclear whether similar results would be found with multidimensional 3AFC measures that include statements from three different dimensions.

Another potential disadvantage of 3AFC measures is that this format may be more cognitively demanding than 2AFC measures. 3AFC items are more complex and cognitively taxing in that examinees are asked to compare and rank among three, rather than two, alternatives. This could lead to increased fatigue and response errors that potentially offset the aforementioned increases in information. So, the theoretical benefits of increased precision may not translate into empirical benefits. Nevertheless, more research is needed to demonstrate that this is the case.

As the other potential disadvantages of the 3AFC format described above suggest, perhaps the most significant disadvantage of this approach is that it has not been adequately tested and more research is needed to ensure this format is useful for operational testing. At this

point, it is unclear whether or not this format is as fake-resistant as 2AFC measures or whether the potential increase in cognitive load will negate the increased information provided by the triplet format. By comparison, a substantial amount of research has demonstrated the potential utility of the 2AFC format for high-stakes testing (Cao & Drasgow, 2019; Drasgow et al., 2012; Nye et al., 2012, 2020; Trent et al., 2020). Therefore, more research is needed before the 3AFC format could be implemented for officer selection and assessment requirements.

Zinnes-Griggs (ZG) Model

Zinnes and Griggs (1974) proposed an alternative IRT model for scoring forced choice items. This model postulates that when raters are presented with a pair of behavioral statements and asked which is more descriptive of them, they will choose the statement that they perceive to be closest to their ideal point on the trait being assessed (Stark & Drasgow, 2002). With the Zinnes-Griggs (1974) model, each statement in a forced choice item is characterized by just one parameter representing its location on the trait continuum. If s and t represent the first and second statements, respectively, in a forced choice item, Zinnes and Griggs (1974) showed that the probability of choosing or preferring statement s to statement t is given by:

$$P_{st}(\theta) = 1 - \Phi(a_{st}) - \Phi(b_{st}) + 2\Phi(a_{st})\Phi(b_{st}),$$

where

$$\begin{aligned} a_{st} &= (2\theta - \mu_s - \mu_t)/\sqrt{3}, \\ b_{st} &= \mu_s - \mu_t \end{aligned}$$

In these equations, θ represents the examinee's ideal point, μ_s and μ_t represent the locations of statements s and t on the trait continuum, and $\Phi(a_{st})$ and $\Phi(b_{st})$ are cumulative standard normal density functions evaluated at a_{st} and b_{st} , respectively (Stark et al, 2011).

Given that the ZG model only incorporates a single location parameter for each statement, the biggest advantage of using this model is that these parameters can be estimated using ratings of the extremity of the statements obtained from subject matter experts (SMEs). Scores based on SME ratings correlate highly with scores based on IRT marginal maximum likelihood estimation (Stark et al., 2011) and using SME ratings instead of pretesting each of the statements could substantially reduce the amount of time and the number of Soldiers required to develop new scales. However, SMEs might provide more accurate estimates of the statement locations in some domains than others, which could affect the accuracy of trait scores (Stark et al., 2011).

Despite the efficiency of using SME ratings instead of pretest data, a significant disadvantage of the ZG model is that it assumes that both statements in a pair assess the same underlying dimension. This is a disadvantage because previous research has indicated that unidimensional pairs are easier to fake than multidimensional pairs (Cao & Drasgow, 2019). When examinees are presented with two statements that only differ based on their extremity on the latent trait that they assess, they can easily identify the most socially desirable responses. Therefore, the increased efficiency of the ZG model may come at the cost of increased faking.

Forced Choice Graded Preference (FCGP) Format

In the FCGP format, statements are presented in pairs and individuals are asked to indicate the extent to which they prefer one statement over the other (or the extent to which one statement describes their personality better than the other one) using a graded scale. For example, respondents might prefer statement A “much more” or “slightly more” than statement B, be ambivalent about statements A and B, or prefer statement B “slightly more” or “much more” than statement A. With a block size of $n = 2$ (i.e., 2AFC), the two statements from each block are simply presented as one pair. If the block size is $n \geq 3$, all possible pairs of the statements within the block (i.e., $\tilde{n} = n(n - 1)/2$ pairs) are presented to respondents one at a time. As with the 3AFC format, including block sizes of 3 or more increases the amount of information obtained from each item when compared with the two alternative response format.

Because the FCGP format still forces respondents to choose between two or more statements that are potentially assessing different dimensions, this approach is likely to have many of the same advantages as other forced choice formats. For example, it is likely to be resistant to faking. However, the advantage of the FCGP approach over the 2AFC or 3AFC formats described above is that it allows test-takers to indicate the extent of their preference, which can increase test-takers engagement. Dalal et al. (2021) found that test-takers react more positively to FCGP items than the more traditional 2AFC format. This is likely to be particularly salient for pairs of negative statements, which have traditionally engendered negative reactions from Soldiers in the Army. In addition, because respondents not only choose between two (or more) response alternatives but also indicate the extent of their preference, the FCGP format provides more information about their standing on the latent trait. Therefore, this format may also have higher reliability than other forced choice formats (Brown & Maydeu-Olivares, 2018).

One possible disadvantage is that the FCGP format may be susceptible to response biases because individuals can consistently choose extreme or neutral options for all graded responses. However, it is unclear that these response styles are a significant problem (Brown & Maydeu-Olivares, 2018), particularly when compared with the traditional Likert format. Perhaps the most significant disadvantage of this format is that it is still relatively new and, therefore, has not been thoroughly researched. Consequently, more research may be needed to demonstrate the utility and psychometric properties of this format in high-stakes settings before it can be ready for operational use.

Empirical Bayes (EB) Scoring

EB scoring is a potential method for improving the accuracy of test scores. EB scoring uses auxiliary information (i.e., information beyond the examinee’s responses to the test items) to increase the precision of trait estimates. The sources of auxiliary information for this method can vary and include things like background information that correlates with test scores (e.g., years of experience, biodata), pretest data from a previous administration of the measure, response times, or correlations among the measured constructs. Although each of these sources of information may be useful, we focus here on the correlations among the measured constructs for several reasons. First, using background information can raise potential fairness concerns if important selection decisions are made based on information not obtained from the test (Wainer

et al., 2001). In addition, using background data requires collecting this information from all applicants, which can require significant effort. Similarly, in an applicant setting, scores on previous test administrations may not be available. Finally, response times raise potential concerns about construct validity, similar to the issues raised for speeded tests (Crocker & Algina, 1986; Rindler, 1979). Therefore, we view the use of correlations among the measured constructs as the most viable source of auxiliary information for officer selection and assessment requirements because it is the most relevant and most likely to be available in a selection setting.

EB scoring has a long history in psychometrics and dates back to early work conducted by Kelly (1927) to provide improved estimates of an individual's true score on a measured characteristic. Kelly (1927) provided an equation to compute an examinee's predicted true score ($\hat{T}$) based on their observed test score (x), the group mean ($x.$), and the test's reliability (r).

$$\hat{T} = r(x - x.) + x.$$

When the observed score is reliable, the predicted true score will be similar to the observed score. In contrast, an unreliable observed score will pull the predicted true score closer to the group mean.

EB scoring can be generalized and applied to IRT scale scores. In other words, IRT scores are estimated for each of the traits assessed by a measure and then augmented post hoc using the EB approach. Essentially, empirical Bayes estimates of the true subscale scores ($\hat{\tau}_i$) for examinee j is a multivariate generalization of Kelly's equation such that

$$\hat{\tau}_i = \mathbf{x}. + \mathbf{S}^{true}(\mathbf{S}^{obs})^{-1}(\mathbf{x}_j - \mathbf{x}.)$$

where $\mathbf{x}_j$ is the vector of subscale scores for examinee j , $\mathbf{x}.$ is the mean vector of subscale scores for the group, $\mathbf{S}^{obs}$ is the observed covariance matrix among the subscales, and $\mathbf{S}^{true}$ is the true covariance matrix among the subscales (estimated using the formula provided by Wainer et al., 2001).

EB scores have several advantages over the original IRT trait estimates. First, previous research has shown that EB scoring can improve the psychometric properties of a measure. For example, Wainer et al. (2001) found that the marginal reliabilities of the augmented scores were as much as 10% higher than the original trait estimates. Conceptually, this should also improve the criterion-related validity of a measure, not only due to the improved reliability but also because the trait estimates are less biased and more precise. A second advantage of EB scoring is that, because augmented EB scores are computed post hoc, this approach could be easily combined with one of the other models described above to obtain the benefits of both. For example, EB scoring could be used to augment IRT trait estimates from a 2AFC forced choice measure to obtain scores that are both resistant to faking and more reliable than traditional 2AFC measures.

In contrast to many of the previous methods discussed in this review, EB scoring does not appear to have many substantial disadvantages for measuring officer characteristics. The only potential disadvantage is practical in that any existing software or testing technology that is used

by ARI would need to be updated to incorporate auxiliary information and calculate the augmented scores. Although this is possible, it may also necessitate the use of third-party software (e.g., Microsoft R) to incorporate these corrections, which would require detailed discussions with the owners of the current testing platforms (e.g., the Army Analytics Group). Again, this is feasible but would require time to incorporate these changes.

Summary and Recommendations

Given the review presented above, we recommend the following approaches to the selection and assessment of officers. First, we recommend using the forced choice format for the items. This format has been widely researched and the available evidence suggests that it may be most useful for high-stakes assessment settings (Cao & Drasgow, 2019; Drasgow et al., 2012; Trent et al., 2020). In addition, this format has also been used successfully by the U.S. Army in operational settings for several years. In contrast to the forced-choice format, many of the alternative approaches reviewed here include significant limitations or have not been adequately tested in high-stakes settings and require more research. For example, with the RRM, there are potential concerns about the construct validity of a response time measure and the potential for cognitive load. Therefore, in our opinion, the forced choice format remains the most promising way of addressing potential concerns about faking in high-stakes settings.

Even within the broader category of forced choice measures, numerous variations on this format are available. For example, there is ZG, 2AFC, 3AFC, or even the graded response format. Among these formats, ZG is not ideal for high-stakes settings because it only administers unidimensional pairs, which are substantially more susceptible to faking (Cao & Drasgow, 2019). The graded response format is promising but requires more research to support its use in operational testing. In contrast, both the 2AFC and 3AFC formats have empirical research to support their use. Conceptually, the 3AFC format has a number of advantages over 2AFC measures, including that it may be more reliable and require fewer items to achieve the same level of precision as the 2AFC items. However, the 3AFC format may have some disadvantages as well, including an increased cognitive load and a slightly higher susceptibility to faking. Therefore, despite the advantages of the 3AFC format, more research is needed to demonstrate the utility of this format in high-stakes settings and, at this point, the 2AFC format has been more widely supported and successfully used. For these reasons, we recommend using the 2AFC format for the proposed officer assessment.

In addition to using the 2AFC format to construct items, we also recommend using EB scoring. As described above, EB scoring can be used with many of the other approaches described here, including 2AFC measures. In addition, some evidence suggests that it can improve score precision and increase the reliability of the measure (Wainer et al., 2001). Consequently, EB scoring could potentially provide additional advantages above and beyond the 2AFC items. However, incorporating this approach would require some additional time to program the 2AFC CAT software to augment scores with information from the construct correlations.

Because EB scoring can be used with other formats and applied post hoc, it can also be implemented using archival data. Therefore, as a proof of concept, we compared the raw scores

from archival 2AFC data to the EB scores to determine the potential impact of this method on the reliability, construct validity, and criterion-related validity of the scores. The data for these analyses came from previous research conduct on the TAPAS (Zhang et al., 2020). This study collected TAPAS data on MTurk at two time points approximately 10 days apart. The sample size for the Time 1 administration was 781 and the sample size at Time 2 was 562. A 10-dimensional version of the TAPAS with 83 2AFC items was administered to all participants. Additional measures of the Big Five personality traits, subjective well-being, and core self-evaluations were also administered to examine the construct and criterion-related validity of the EB scores. Additional details about the measures used for this study are provided by Zhang et al. (2020).

The results of this study are provided in Tables 13-15 for both the raw TAPAS scores and the EB scores. Table 13 shows the correlations between the raw scores and EB scores at both time points. The lowest correlation between the corresponding scores was .96 and the average across all 10 facets was .98. These results suggest that the two scores are very highly correlated. Table 14 shows the reliabilities of both the raw and EB scores. Here, we report both the IRT marginal reliabilities at T1 and T2 and the test-retest reliability across time points. Overall, the results indicate that the EB scores resulted in higher reliabilities than the raw scores. However, the improvements in reliability varied across facets. For example, at T1, the marginal reliabilities were .11 and .07 *higher* with the EB scores for selflessness and achievement, respectively, than for the raw scores. In contrast, the reliability was only .01 higher for physical conditioning. Nevertheless, the average marginal reliabilities were .05 larger at T1, .04 larger at T2, and the average test-retest reliability was .03 larger. Although not a huge improvement, these differences aggregate over a number of much larger improvements at the facet-level. In addition, these increases at the facet-level would be even larger improvements when aggregated in a composite, as is often done to make selection decisions in the Army. Therefore, the reliability differences between the raw and EB scores may be meaningful.

Table 13. Correlations between the Raw and Augmented (EB) Scores

Facet	T1	T2
Achievement	.97	.98
Dominance	.97	.97
Even-Tempered	.98	.98
Intellectual	.99	.99
Efficiency	.99	.99
Optimism	.99	.99
Order	.98	.99
Physical	.99	.99
Conditioning	.99	.99
Selflessness	.96	.97
Sociability	.99	.99
Tolerance	.98	.98
Mean	.98	.98

Table 14. Reliabilities of the Raw and Augmented (EB) Scores

Facet	Marginal Reliability (T2)		Marginal Reliability (T2)		Test-Retest Reliability	
	Raw	EB	Raw	EB	Raw	EB
Achievement	.70	.77	.68	.75	.67	.70
Dominance	.78	.81	.78	.81	.76	.79
Even-Tempered	.69	.75	.71	.77	.73	.76
Intellectual Efficiency	.64	.71	.67	.71	.69	.71
Optimism	.78	.82	.79	.83	.74	.76
Order	.81	.83	.83	.84	.79	.80
Physical Conditioning	.85	.86	.86	.87	.79	.80
Selflessness	.64	.75	.68	.77	.67	.72
Sociability	.75	.79	.76	.80	.79	.80
Tolerance	.75	.78	.74	.78	.72	.74
Mean	.74	.79	.75	.79	.73	.76

Table 15 shows the correlations with the Big Five measure, subjective well-being, and core self-evaluations. Again, the improvements observed for the EB scores varied by facet. Some of the largest differences were observed for the even-tempered facet. For example, the correlation between the EB even-tempered scores and CSE was .10 higher than for the corresponding raw scores while the correlation with SWB was .08 higher. Similarly, the correlation between the order facet and CSE was .08 larger with EB scoring than with the raw scores. Although these were some of the largest increases in the correlations, the correlations for many other facets were also modestly higher. Again, these effects would be even larger when combined together in composites, like those reported for the TAPAS. Therefore, these results suggest that the EB scores could potentially improve the reliabilities, construct validities, and criterion-related validities of a 2AFC measure, though the magnitudes of these improvements will vary across facets.

Table 15. Construct and Criterion Correlations for the Raw and Augmented (EB) Scores

Facets	Construct Correlations										Criterion Correlations			
	BFI-A		BFI-C		BFI-E		BFI-N		BFI-O		SWB		CSE	
	Raw	EB	Raw	EB	Raw	EB	Raw	EB	Raw	EB	Raw	EB	Raw	EB
Achievement	.21	.26	.50	.52	.04	.06	-.15	-.20	.20	.27	.07	.09	.25	.28
Dominance	-.06	-.04	.10	.14	.47	.55	-.12	-.16	.13	.17	.06	.08	.11	.15
Even-Tempered	.42	.47	.22	.29	-.03	.03	-.50	-.56	.08	.13	.18	.26	.31	.41
Intellectual Efficiency	.03	.06	.15	.20	.02	.03	-.12	-.14	.32	.34	-.05	-.04	.07	.10
Optimism	.29	.33	.36	.40	.33	.37	-.61	-.65	.00	.02	.55	.55	.69	.70
Order	.09	.11	.53	.59	.07	.09	-.16	-.22	-.03	-.03	.13	.18	.24	.32
Physical Conditioning	.02	.04	.24	.26	.18	.22	-.19	-.23	.03	.01	.18	.22	.20	.24
Selflessness	.39	.42	.11	.17	-.03	-.02	.02	-.01	.27	.33	.04	.04	.03	.05
Sociability	.23	.25	.14	.17	.71	.73	-.32	-.36	.10	.11	.19	.23	.24	.30
Tolerance	.18	.23	.00	.02	.02	.02	.08	.07	.36	.40	-.06	-.06	-.10	-.09
Mean	.19	.22	.23	.28	.19	.21	.23	.26	.15	.18	.15	.17	.22	.26

Note. BFI-A=Big Five Inventory Agreeableness Scale; BFI-C=Big Five Inventory Conscientiousness Scale; BFI-E=Big Five Inventory Extraversion Scale; BFI-N=Big Five Inventory Neuroticism Scale; BFI-O=Big Five Inventory Openness Scale; SWB=Subjective Well-Being; CSE=Core Self-Evaluation.

RESEARCH PLAN TO DEVELOP A NON-COGNITIVE ASSESSMENT FOR OFFICER APPLICATIONS

In sum, we conducted a meta-analysis to identify the personality characteristics that have the strongest relationships with leadership performance. Although this provided useful initial information for identifying important officer characteristics, it also included a number of studies that examined civilian samples. Therefore, to create a more comprehensive list of non-cognitive characteristics and ensure that Army-specific characteristics are included, we also examined existing Army leadership doctrine and conducted focus groups with officers. Finally, we reviewed the psychometric literature to identify the most promising models for assessing personality in high-stakes settings.

This research provides a clearer picture of the characteristics of non-cognitive assessments that can be successfully used for officer selection and assignment requirements. Therefore, in this section, we build on this work to propose a research plan for developing a new non-cognitive assessment for officer applications. This research plan was developed with several criteria in mind. First, the assessment should focus on assessing a comprehensive set of personality characteristics that are relevant for Army officers while also balancing concerns about test-taker fatigue and motivation that may be associated with test length. Second, any assessment that is developed for officer selection requirements should be able to be administered in high-stakes settings. Finally, the assessment should be administered as a computer adaptive test (CAT) to achieve increased accuracy, efficiency, and test security (Drasgow et al., 2009; Stark et al., 2012).

Proposed Research Plan

Selecting an Underlying IRT Model

As described above, we propose assessing the 16 personality dimensions shown in Table 12. The first step in developing the proposed assessment is to write statements that assess each of these dimensions. Before writing these statements, it is important to first identify the item response theory (IRT) model that will be used because this has implications for the statements that are generated for the assessment. Adaptive testing requires that statements are calibrated and responses are scored using a formal item response theory (IRT) model. An important distinction in IRT is between dominance and ideal point IRT models. Dominance models assume that individuals will endorse a statement as long as their current standing on the latent trait is higher than the level of the trait being assessed by the statement. In contrast, ideal point models assume that individuals will only endorse a statement if the extremity of that statement matches their standing on the trait being assessed.

The distinction between dominance and ideal point models is not a trivial concern. First, the model that is chosen will impact the types of statements that are developed and included in the assessment. For example, ideal point models require statements that assess high, intermediate, and low levels of the construct (Chernyshenko et al., 2007). In contrast, dominance models often exclude statements that assess intermediate levels of the trait because they tend to have low item-total correlations (Cao, Drasgow, & Cho, 2015). In addition, choosing the correct IRT model is important because using a dominance model to score ideal point data can affect the accuracy of the test scores and the decisions that are made from them. Because intermediate

items are excluded, assessments developed using a dominance model will provide limited information at the high and low ends of the trait (Chernyshenko et al., 2007). This is particularly important for selection and assignment settings because the low information at the ends of the distribution will affect the rank-order of individuals and influence select-in (i.e., selecting the highest scorers) or select-out (i.e., screening out the lowest scorers) decisions. Finally, the scores that result from using an inaccurate IRT model can also influence the criterion-related validity of the measure (Carter et al., 2014; Chernyshenko et al., 2007). Therefore, it is important to select an appropriate IRT model.

The choice to use a dominance or ideal point model should be based on the fit of the model to the data (Nye et al., 2020). For personality statements, past research has demonstrated that ideal point models tend to fit the response data the best (Chernyshenko et al., 2007; Stark et al., 2006). Of particular importance for the proposed officer assessment, ideal point models have also been shown to fit response data from personality assessments that are used in the U.S. Army (Drasgow et al., in progress). Therefore, we propose developing the statements for the officer assessment using an ideal point framework that assesses the full range of the latent trait including high, intermediate, and low levels of the trait. Here, writing statements to reflect the intermediate levels of the trait can be particularly difficult. Therefore, we recommend following the framework provided by Cao, Drasgow, and Cho (2015) to develop these intermediate statements. However, once all of the statements are written, it will be important to test model-data fit to confirm that the ideal point model fits the data. Nevertheless, given past research in both civilian and military samples (Chernyshenko et al., 2007; Drasgow et al., in progress; Stark et al., 2006), we expect that the ideal point response process will fit well.

It is important to note that either dominance or ideal point models can be used with the 2AFC format proposed in the previous section of this report. One model that can be used to score the 2AFC format is the Multi-Unidimensional Pairwise Preference (MUPP) model (Stark, 2002; Stark, Chernyshenko, & Drasgow, 2005). This model assumes that when person j encounters statements s and t the person considers whether to endorse s and, independently, considers whether to endorse t . The probability of preferring a particular statement in a pair thus depends on the individual's standing on the traits assessed by statements s and t , as well as the model chosen to characterize the response process for the individual statements. This model can either be a dominance or an ideal point model. Given the superior fit of ideal point models to personality data, Stark (2002) proposed using the dichotomous case of the Generalized Graded Unfolding Model (GGUM; Roberts, Donoghue, & Laughlin, 2000), which is an ideal point model that has been shown to fit personality data reasonably well (Chernyshenko, Stark, Drasgow, & Roberts, 2007).

In subsequent sections of this report, we differentiate between the statements that are used to assess a particular personality construct and the items that are administered to test-takers. In traditional personality measures, which were primarily administered using single-statement items, there is no difference between statements and items. However, with 2AFC measures, each item would consist of two statements assessing different dimensions. In other words, the statement is the phrase that is designed to assess a particular construct but the item consists of the two statements that are paired together to mitigate the effects of faking. Therefore, although the total statement pool for all 16 dimensions shown in Table 12 may consist of 800+ individual statements (see below), a 2AFC assessment may only include 120 items. Nevertheless, the total number of items in an assessment will depend on a number of factors, including the number of

statements administered for each dimension, which can affect the reliability of the overall measure.

Statement Development and Pretesting

Next, the definitions and example items provided in Table 12 should be used to write statements for each of the proposed personality facets. If desired, statement pools from the TAPAS are readily available and can be used to assess several of these facets. However, many of the TAPAS statement pools are already used with other Army samples. Therefore, if a dedicated statement pool that is only used for officers is desired, separate statement pools will need to be developed for the proposed officer assessment. Regardless of whether these existing statement pools are used, additional statements will need to be written to assess the dimensions that are not currently available for the TAPAS. Currently, only the initiative and empathy facets do not have corresponding TAPAS statement pools. Nevertheless, even if existing TAPAS statement pools are used for the rest of the facets in the proposed assessment, the statements in these pools should still be reviewed and revised to ensure the terminology is appropriate for officers (i.e., no terminology that is specific to enlisted Soldiers is used) and that the content is consistent with the feedback received in the focus groups. For example, the TAPAS Dominance facet has traditionally been described as occurring when individuals are “domineering, “take charge,” and are “referred to by their peers as natural leaders.” However, the focus groups consistently noted that being “domineering” is a negative trait for officers *in most cases*. Therefore, the statement pools for this dimension will need to be reviewed and revised to ensure the content is consistent with this new definition and new statements may need to be added to ensure there are a sufficient number of statements for the assessment.

Large statement pools are required for an assessment to be administered as a computer adaptive test. Therefore, we propose developing approximately 60-80 initial statements for each dimension that match the definitions provided in Table 12. To be consistent with the ideal point framework, it is important that each statement pool assesses the full range of the latent trait and includes statements assessing high, intermediate, and low levels of the construct (Chernyshenko et al., 2007). To ensure sufficient content validity and coverage of each construct, these initial statements should be written by subject matter experts (SMEs) who either have a Ph.D. in Industrial and Organizational Psychology or extensive experience with personality assessment.

After developing the initial statement pools, each statement should be carefully reviewed to establish content validity. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014) suggest that content validity evidence can be provided by SME’s judgements about the relationships between the test content and the constructs that are being assessed. Therefore, a small group of SMEs with Ph.D.s in Industrial and Organizational Psychology or extensive experience with personality assessment should review the statement pools closely to ensure the content validity and readability (i.e., grammar, punctuation, length, clarity) of all statements. Statements that are confusing or that do not match the definition of the construct should be removed from the pool or revised to improve content validity.

Once the SMEs are satisfied with the content validity of the statement pools, all remaining statements will need to be pretested in samples of Soldiers to further refine the content

and examine the psychometric properties of the statements. To collect sufficient pretest data, each statement will need to be administered to approximately 500 individuals to ensure sufficient sample sizes for IRT estimation. In addition, subsamples of these participants will also need to rate the social desirability of each statement. These desirability ratings can then be used in combination with the IRT parameters to match statements and create the 2AFC items that are administered to test-takers. Due to the sample independence of IRT parameter estimates, the majority of the individuals required for pretesting could be enlisted Soldiers, which would be easier to obtain in large numbers than officers. However, officers would also be needed to provide the social desirability ratings. Although we do not expect IRT parameters to vary across enlisted Soldiers and officers, the social desirability of each statement could vary due to the different requirements in these roles. Nevertheless, each statement only needs to be administered to approximately 20-30 officers to obtain accurate social desirability ratings.

By combining statements into multiple forms, the full set of 800+ statements for all 16 dimensions can be pretested using samples of approximately 2,600 Soldiers. This number includes 2,400 enlisted Soldiers and approximately 200 officers to provide the social desirability ratings. However, if existing TAPAS statement pools are used for at least some of the dimensions, then the number of Soldiers required for pretesting could be substantially reduced. For example, if only the statement pools for initiative and empathy are pretested, then only approximately 800 Soldiers would be needed. This includes 600 enlisted Soldiers for estimating the IRT parameters. Note that 200 officers would still be needed to provide the social desirability ratings for all 16 dimensions, given that previous TAPAS research has primarily used enlisted Soldiers to provide the social desirability ratings in the existing statement pools. In both of these scenarios, the response data from these pretest samples can then be used to estimate the IRT parameters necessary for CAT administration and ensure that a sufficient number of statements assessing high, intermediate, and low levels of the construct are included in the statement pools. Using the IRT parameters, low quality statements, such as those with low discrimination parameters (i.e., $\alpha < .50$) or extreme item location parameters (i.e., $\delta > +$ or $- 9.0$), can be identified and removed from the pools.

CAT Software and EB Scoring

As described above, the proposed assessment will be administered as an adaptive test. If using the forced choice format as proposed in the current research, this can be done using the existing software that the Army uses to administer the TAPAS. Therefore, no new software would be needed for this assessment. This would be a significant advantage for the new assessment because the process of running security checks to get new software downloaded on an Army IT platform is extensive and can take a significant amount of time. Therefore, using existing software could save substantial time.

Despite the advantages of using existing software that has already been screened for security flaws, incorporating EB scoring would require an additional update to the software. Calculating the EB scores can be done on the back-end of the CAT administration and, therefore, would not require substantial changes to the software but could be implemented using the output from it. For example, a Microsoft R Script could be created that takes the scores for each individual and transforms them using the EB equations. There are multiple ways to do this but all would require coordinating with the programmers for whatever platform the new officer assessment would be administered on (e.g., the Army Analytics Group server) to figure out the

most efficient way of implementing these changes with the highest likelihood of success given the Army's current systems.

Validation of the New Assessment

Once the statement pools have been pretested and the final set of statements are identified, the new officer assessment will need to be validated to demonstrate its utility for predicting officer outcomes. Although a longitudinal study is preferable whenever possible, it is often not feasible to collect data at two time points with the same Soldiers. Therefore, we propose a concurrent validation study to collect validity evidence for this assessment. The results of this study can provide initial validity evidence while longitudinal data is collected for a future validation study.

For the concurrent validation study, we propose collecting data from approximately 800 officers in at least four different basic branches in the Army. Each of these officers would be administered the new officer assessment as well as a set of relevant criterion measures that assess important officer outcomes. This design would provide sufficient data to examine the predictive validity of the new assessment in the full sample and within each of the basic branches. In addition, it would also allow for comparisons across the branches to determine whether a non-cognitive assessment can identify individuals who are likely to be more successful in some branches than in others. This sort of differential validity is necessary to be useful for making officer assignment decisions. Once the data have been collected, regression analyses can be used to identify composites of personality traits that are relevant both for Army-wide officer selection and for assignment into the basic branches.

Anticipated Timeline

The proposed timeline for this project is shown in Figures 3 and 4. Here, we propose two different timelines depending on the approach desired by ARI. The key distinction between these two timelines is the number of new statements that need to be developed and pretested for the proposed assessment. If existing TAPAS statement pools are used for some of the proposed dimensions, the number of statements that need to be pretested, and the number of Soldiers required to pretest the statements, is dramatically reduced. In this case, the anticipated timeline for the proposed research is approximately 36 months. As shown in Figure 3, this would involve writing the initial statement pools for any new dimensions during the first six months and then pretesting them in the following six months. After pretesting is completed, the data for the validation study can be collected and analyzed over the next 24 months.

In contrast, if new statement pools are needed for the officer assessment, the proposed timeline would be extended to 48 months². This extended timeline is entirely due to the amount of time required to collect sufficient pretest data on all 16 dimensions. As noted above, pretesting new statement pools for all 16 dimensions would require approximately 2,600 Soldiers (including 200 officers). Due to the limited number of Soldiers available for data collections at any one time point, it would not be possible to collect this much data in six months as shown in

² Note, however, that even this timeline could be reduced if a static (i.e., non-adaptive) form is sufficient for the initial validation study. In this case, a subset of approximately 20 statements for each of the 16 facets could be developed, pretested, and used for the initial validation study while the much larger statement pools for the CAT administration are developed and pretested.

Figure 3. Instead, as shown in Figure 4, this would take at least 18 months, if not longer. Therefore, if we assume the same 24 month time period for the validation data collection and analyses, the anticipated timeline for creating and validating 16 new statement pools would be at least 48 months. This timeline is consistent with previous scale development work supported by ARI (e.g., Nye et al., 2019).

Figure 3. Possible 36-Month Timeline for the Proposed Research Plan

Activity	Months											
	1	2	3	4	5	6	7	8	9	10	11	12
Task 1: Develop New Statement Pools												
Project kickoff and execution plan												
Write and edit new statements												
Submit IRB and RMD for Pretest Data Collections												
Collect pretest data												
Task 2: Validation Study of the Officer Assessment												
Submit IRB and RMD for the validation study												
Update CAT software with new statement pools												
Collect officer validation data												
Analyze officer validation data												
Submit draft report												
Revise and resubmit the final report												
Activity	Months											
	13	14	15	16	17	18	19	20	21	22	23	24
Task 1: Develop New Statement Pools												
Project kickoff and execution plan												
Write and edit new statements												
Submit IRB and RMD for Pretest Data Collections												
Collect pretest data												
Task 2: Validation Study of the Officer Assessment												
Submit IRB and RMD for the validation study												
Update CAT software with new statement pools												
Collect officer validation data												
Analyze officer validation data												
Submit draft report												
Revise and resubmit the final report												
Activity	Months											
	25	26	27	28	29	30	31	32	33	34	35	36
Task 1: Develop New Statement Pools												
Project kickoff and execution plan												
Write and edit new statements												
Submit IRB and RMD for Pretest Data Collections												
Collect pretest data												
Task 2: Validation Study of the Officer Assessment												
Submit IRB and RMD for the validation study												
Update CAT software with new statement pools												
Collect officer validation data												
Analyze officer validation data												
Submit draft report												
Revise and resubmit the final report												

Figure 4. Possible 48-Month Timeline for the Proposed Research Plan

Activity	Months											
	1	2	3	4	5	6	7	8	9	10	11	12
Task 1: Develop New Statement Pools												
Project kickoff and execution plan												
Write and edit new statements												
Submit IRB and RMD for Pretest Data Collections												
Collect pretest data												
Task 2: Validation Study of the Officer Assessment												
Submit IRB and RMD for the validation study												
Update CAT software with new statement pools												
Collect officer validation data												
Analyze officer validation data												
Submit draft report												
Revise and resubmit the final report												
Activity	Months											
	13	14	15	16	17	18	19	20	21	22	23	24
Task 1: Develop New Statement Pools												
Project kickoff and execution plan												
Write and edit new statements												
Submit IRB and RMD for Pretest Data Collections												
Collect pretest data												
Task 2: Validation Study of the Officer Assessment												
Submit IRB and RMD for the validation study												
Update CAT software with new statement pools												
Collect officer validation data												
Analyze officer validation data												
Submit draft report												
Revise and resubmit the final report												
Activity	Months											
	25	26	27	28	29	30	31	32	33	34	35	36
Task 1: Develop New Statement Pools												
Project kickoff and execution plan												
Write and edit new statements												
Submit IRB and RMD for Pretest Data Collections												
Collect pretest data												
Task 2: Validation Study of the Officer Assessment												
Submit IRB and RMD for the validation study												
Update CAT software with new statement pools												
Collect officer validation data												
Analyze officer validation data												
Submit draft report												
Revise and resubmit the final report												
Activity	Months											
	37	38	39	40	41	42	43	44	45	46	47	48
Task 1: Develop New Statement Pools												
Project kickoff and execution plan												
Write and edit new statements												
Submit IRB and RMD for Pretest Data Collections												
Collect pretest data												
Task 2: Validation Study of the Officer Assessment												
Submit IRB and RMD for the validation study												
Update CAT software with new statement pools												
Collect officer validation data												
Analyze officer validation data												
Submit draft report												
Revise and resubmit the final report												

CONCLUSIONS

The purpose of the current research was to do a preliminary evaluation to identify the potential characteristics of a new assessment that can be used for officer selection and assessment requirements. The first steps in this process involved conducting a meta-analysis, a qualitative review of Army doctrine, and focus groups to identify the most important characteristics of successful officers. Based on this work, we proposed assessing 16 different personality facets to identify individuals with a high potential for success in an officer role. Although several other promising facets were also identified, we limited the number of facets in the proposed assessment due to potential concerns about test-taker fatigue and motivation. Nevertheless, a 16-dimension assessment will still measure numerous useful characteristics and is consistent with other non-cognitive assessments used by the Army (e.g., Drasgow et al., in progress).

In addition to the review of potential characteristics that can be assessed, we also reviewed several potential methods that can be used to assess these characteristics. Here, again, we identified a number of potentially useful methods and formats. However, the bulk of the existing research suggests that a forced choice format may be most useful for high-stakes settings where faking is a potential concern (e.g., Cao & Drasgow, 2019). Based on this evidence, the 2AFC or 3AFC formats are both promising options. Although the feasibility of the 3AFC format has been demonstrated in Army samples (Drasgow et al., in progress), the sample sizes were relatively small in that research and several research questions remain for this format. Therefore, given the more substantial body of research for the 2AFC format, we believe that this format can be adopted more quickly and implemented immediately for an officer assessment. In contrast, although the 3AFC format also shows promise, additional research would be needed before we could confidently recommend this format for operational testing.

Finally, this report concludes with a proposed research plan to develop an officer assessment that incorporates the findings from this research. The proposed research plan would take approximately 36-48 months and involve developing new statement pools, pretesting the new statements to estimate IRT parameters and remove low quality statements, and conducting a validation study of the new assessment. The approach described here follows current professional standards for scale development (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014; SIOP, 2018). In addition, although the validation process is never complete, the results of the proposed research plan will provide the initial evidence necessary to support the use of the new measure for officer selection and assessment requirements. Given the potential influence of personality on work performance (e.g., Barrick & Mount, 1991; Campbell & Knapp, 2001), this new assessment is likely to help identify individuals with high potential for success as officers in the U.S. Army.

REFERENCES

*Indicates studies that were included in the meta-analysis.

*Allen, M. T., Bynum, B. H., Erk, R. T., Babin, N. E., & Young, M. C. (2014). *Selecting soldiers and civilians into the US army officer candidate school: Developing empirical selection composites* (Technical Report 1343). U.S. Army Research Institute for the Behavioral and Social Sciences.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

*Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360 feedback. *Journal of Vocational Behavior*, 66(3), 532-548. <https://doi.org/10.1016/j.jvb.2004.05.003>

*Atwater, L. E., & Yammarino, F. J. (1993). Personal attributes as predictors of superiors' and subordinates' perceptions of military academy leadership. *Human Relations*, 46(5), 645-668. <https://doi.org/10.1177/001872679304600504>

*Atwater, L. E., Dionne, S. D., Avolio, B., Camobreco, J. F., & Lau, A. W. (1999). A longitudinal study of the leadership development process: Individual differences predicting leader effectiveness. *Human Relations*, 52(12), 1543-1562. <https://doi.org/10.1023/A:1016985019728>

*Barbuto, J. E., Phipps, K. A., & Xu, Y. (2010). Testing relationships between personality, conflict styles and effectiveness. *International Journal of Conflict Management*, 21(4), 434-447. <https://doi.org/10.1108/10444061011079967>

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1-26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>

*Barrick, M. R., Parks, L., & Mount, M. K. (2005). Self-monitoring as a moderator of the relationships between personality traits and performance. *Personnel Psychology*, 58(3), 745-767.

*Bartone, P. T., Eid, J., Johnsen, B. H., Laberg, J. C., & Snook, S. A. (2009). Big five personality factors, hardiness, and social judgment as predictors of leader performance. *Leadership & Organization Development Journal*, 30(6), 498-521. <https://doi.org/10.1108/01437730910981908>

*Bartone, P. T., Kelly, D. R., & Matthews, M. D. (2013). Psychological hardiness predicts adaptability in military leaders: A prospective study. *International Journal of Selection and Assessment*, 21(2), 200-210. <https://doi.org/10.1111/ijsa.12029>

- *Bartone, P. T., Snook, S. A., & Tremble T. R., (2002). Cognitive and personality predictors of leader performance in West Point cadets. *Military Psychology, 14*(4), 321-338. https://doi.org/10.1207/S15327876MP1404_6
- Bass, B. M. (1985) *Leadership and Performance Beyond Expectation*. The Free Press.
- *Bauer, T. N., Erdogan, B., Liden, R. C., & Wayne, S. J. (2006). A longitudinal study of the moderating role of extraversion: Leader-member exchange, performance, and turnover during new executive development. *Journal of Applied Psychology, 91*(2), 298-310. <https://doi.org/10.1037/0021-9010.91.2.298>
- *Bergner, S. (2020). Being smart is not enough: personality traits and vocational interests incrementally predict intention, status and success of leaders and entrepreneurs beyond cognitive ability. *Frontiers in Psychology, 11*, Article 204. <https://doi.org/10.3389/fpsyg.2020.00204>
- *Bergner, S., Neubauer, A. C., & Kreuzthaler, A. (2010). Broad and narrow personality traits for predicting managerial success. *European Journal of Work and Organizational Psychology, 19*(2), 177-199. <https://doi.org/10.1080/13594320902819728>
- *Berry, C. M., Page, R. C., & Sackett, P. R. (2007). Effects of self-deceptive enhancement on personality–job performance relationships. *International Journal of Selection and Assessment, 15*(1), 94-109. <https://doi.org/10.1111/j.1468-2389.2007.00374.x>
- *Bilgic, R., & Sümer, H. C. (2009). Predicting military performance from specific personality measures: A validity study. *International Journal of Selection and Assessment, 17*(2), 231-238. <https://doi.org/10.1111/j.1468-2389.2009.00465.x>
- *Black, J. (2000). Personality testing and police selection: Utility of the Big Five. *New Zealand Journal of Psychology, 29*(1), 2-9. <https://www.psychology.org.nz/journal-archive/NZJP-Vol291-2000-1-Black.pdf>
- Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology, 89*(5), 901-910. <https://doi.org/10.1037/0021-9010.89.5.901>
- *Boswell, W. R., Roehling, M. V., & Boudreau, J. W. (2006). The role of personality, situational, and demographic variables in predicting job search among European managers. *Personality and Individual Differences, 40*(4), 783-794. <https://doi.org/10.1016/j.paid.2005.09.008>
- *Boudreau, J. W., Boswell, W. R., & Judge, T. A. (2001). Effects of Personality on Executive Career Success in the United States and Europe. *Journal of Vocational Behavior, 58*(1), 53–81. <https://doi.org/10.1006/jvbe.2000.1755>

- Boyle, G. J. (1995). Myers-Briggs type indicator (MBTI): Some psychometric limitations. *Australian Psychologist*, 30(1), 71-74. <https://doi.org/10.1111/j.1742-9544.1995.tb01750.x>
- Bradley, P. (2006). Obedience to military authority: A psychological perspective. In C. L. Mantle (Ed.), *The unwilling and the reluctant: Theoretical perspectives on disobedience in the military* (pp. 13–41). Canadian Defence Academy Press.
- *Bradley, J. P., Nicol, A. A. M., Charbonneau, D., & Meyer, J. P. (2002). Personality correlates of leadership development in Canadian forces officer candidates. *Canadian Journal of Behavioural Sciences*, 34(2), 92-103. <https://doi.org/10.1037/h0087159>
- Braun, S. (2017). Leader narcissism and outcomes in organizations: a review at multiple levels of analysis and implications for future research. *Frontiers in Psychology*, 8, Article 773. doi.org/10.3389/fpsyg.2017.00773
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460-502. doi.org/10.1177/0013164410375112
- Brown, A. & Maydeu-Olivares, A. (2018). Ordinal Factor Analysis of Graded-Preference Questionnaire Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(1), 516-529. <https://doi.org/10.1080/10705511.2017.1392247>
- *Brown, M. E., & Treviño, L. K. (2009). Leader–follower values congruence: Are socialized charismatic leaders better able to achieve it? *Journal of Applied Psychology*, 94(2), 478-490. <https://doi.org/10.1037/a0014069>
- *Brunell, A. B., Gentry, W. A., Campbell, W. K., Hoffman, B. J., Kuhnert, K. W., & DeMarree, K. G. (2008). Leader Emergence: The Case of the Narcissistic Leader. *Personality and Social Psychology Bulletin*, 34(12), 1663-1676. <https://doi.org/10.1177/0146167208324101>
- *Bynum, B. H., & Young, M. C. (2020). *Validation of the Cadet Background and Experience Form to Support Army ROTC Personnel Assessment (2015-2018)* (Technical Report 1383). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Byrne, J. C., Dominick, P. G., Smither, J. W., & Reilly, R. R. (2007). Examination of the Discriminant, Convergent, and Criterion-Related Validity of Self-Ratings on the Emotional Competence Inventory. *International Journal of Selection and Assessment*, 15(3), 341-353. <https://doi.org/10.1111/j.1468-2389.2007.00393.x>
- *Caligiuri, P., & Tarique, I. (2012). Dynamic cross-cultural competencies and global leadership effectiveness. *Journal of World Business*, 47(4), 612-622. <https://doi.org/10.1016/j.jwb.2012.01.014>

- *Calleja, J. A., Hoggan, B. L., & Temby, P. (2020). Individual predictors of tactical planning performance in junior military officers. *Military Psychology*, 32(2), 149-163. <https://doi.org/10.1080/08995605.2019.1691405>
- Campbell, J. P., & Knapp, D. J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. Lawrence Erlbaum Associates, Inc.
- *Campbell, L., Simpson, J. A., Stewart, M., & Manning, J. G. (2002). The formation of status hierarchies in leaderless groups: The role of male waist-to-hip ratio. *Human Nature*, 13(3), 345–362. <https://doi.org/10.1007/s12110-002-1019-0>
- Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, 104(11), 1347-1368. <https://doi.org/10.1037/apl0000414>
- Cao, M., Drasgow, F., & Cho, S. (2015). Developing ideal intermediate personality items for the ideal point model. *Organizational Research Methods*, 18(2), 252-275. <https://doi.org/10.1177/1094428114555993>
- *Carter, J. C. (2009). Transformational leadership and pastoral leader effectiveness. *Pastoral Psychology*, 58(3), 261-271. <https://doi.org/10.1007/s11089-008-0182-6>
- Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M. C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology*, 99(4), 564-586. <https://doi.org/10.1037/a0034688>
- Capon, J., Chernyshenko, O. S., & Stark, S. (2007). Applicability of civilian retention theory in the New Zealand military. *New Zealand Journal of Psychology*, 36(1), 50–56. https://www.psychology.org.nz/journal-archive/36-1_Capon-Chernyshenko-Stark_pg50.pdf
- Chernyshenko, O. S., Kankaraš, M., & Drasgow, F. (2018). *Social and emotional skills for student success and well-being: Conceptual framework for the OECD study on social and emotional skills* (OECD Education Working Papers, No. 173). Organisation for Economic Cooperation and Development. <https://doi.org/10.1787/db1d8e59-en>
- Chernyshenko, O. S., Stark, S., & Drasgow, F. (2011). Individual differences: Their measurement and validity. In S. E. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology: Vol 2. Selecting and Developing Members for the Organization*. (pp. 117-151). American Psychological Association.
- Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumption of an ideal point response process: Toward increasing the

- flexibility of personality measures. *Psychological Assessment*, 19(1), 88-106.
<https://doi.org/10.1037/1040-3590.19.1.88>
- *Christiansen, N. D., & Robie, C. (2011). Further consideration of the use of narrow trait scales. *Canadian Journal of Behavioural Science*, 43(3), 183-194.
<https://doi.org/10.1037/a0023069>
- *Cogliser, C. C., Gardner, W. L., Gavin, M. B., & Broberg, J. C. (2012). Big five personality factors and leader emergence in virtual teams: Relationships with team trustworthiness, member performance contributions, and team performance. *Group & Organization Management*, 37(6), 752-784. <https://doi.org/10.1177/105960111246426>
- *Colbert, A. E., Judge, T. A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly*, 23(4), 670-685.
<https://doi.org/10.1016/j.leaqua.2012.03.004>
- *Connelly, M. S., Gilbert, J. A., Zaccaro, S. J., Threlfall, K. V., Marks, M. A., & Mumford, M. D. (2000). Exploring the relationship of leadership skills and knowledge to leader performance. *The Leadership Quarterly*, 11(1), 65-86. [https://doi.org/10.1016/S1048-9843\(99\)00043-0](https://doi.org/10.1016/S1048-9843(99)00043-0)
- Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10(4), 331-360. https://doi.org/10.1207/s15327043hup1004_2.
- *Cortina, J.M., Doherty, M.L., Schmitt, N., Kaufman, G., & Smith, R.G. (1992). The "big five" personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology*, 45(1), 119-140. <https://doi.org/10.1111/j.1744-6570.1992.tb00847.x>
- Costa, P. T., Jr., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Vol. 2. Personality measurement and testing* (pp. 179–198). Sage Publications, Inc. <https://doi.org/10.4135/9781849200479.n9>
- *Courtright, S. H., Colbert, A. E., & Choi, D. (2014). Fired up or burned out? How developmental challenge differentially impacts leader behavior. *Journal of Applied Psychology*, 99(4), 681-696. <https://doi.org/10.1037/a0035790>
- *Crant, J. M., & Bateman, T. S. (2000). Charismatic leadership viewed from above: the impact of proactive personality. *Journal of Organizational Behavior*, 21(1), 63-75.
[https://doi.org/10.1002/\(SICI\)1099-1379\(200002\)21:1<63::AID-JOB8>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1099-1379(200002)21:1<63::AID-JOB8>3.0.CO;2-J)
- *Crespo, M. R. (2007). *Intelligence and Personality as Antecedents to Leadership Effectiveness and Extrinsic Career Success: The Predictive Power of Individual Differences* [Doctoral dissertation, Columbia University].

- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston.
- Dalal, D. K., Zhu, X. S., Rangel, B., Boyce, A. S., & Lobene, E. (2021). Improving applicant reactions to forced-choice personality measurement: Interventions to reduce threats to test takers' self-concepts. *Journal of Business and Psychology*, 36(1), 55-70. <https://doi.org/10.1007/s10869-019-09655-6>
- *Dalton, M., & Wilson, M. (2000). The relationship of the five factor model of personality to job performance for a group of Middle Eastern expatriate managers. *Journal of Cross Cultural Psychology*, 31(2), 250–258. <https://doi.org/10.1177/0022022100031002007>
- Darr, W. (2011). Military personality research: A meta-analysis of the Self-Description Inventory. *Military Psychology*, 23(3), 272- 296. <https://doi.org/10.1080/08995605.2011.570583>
- *De Hoogh, A. H., Den Hartog, D. N., & Koopman, P. L. (2005). Linking the Big Five-Factors of personality to charismatic and transactional leadership: Perceived dynamic work environment as a moderator. *Journal of Organizational Behavior*, 26(7), 839-865. <https://doi.org/10.1002/job.344>
- de la Torre, J., Ponsoda, V., Leenen, I., & Hontangas, P. (2012, April). *Examining the viability of recent models for forced-choice data* [Paper presentation]. Meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.
- *Deluga, R. J. (1998). Leader–member exchange quality and effectiveness ratings: The role of subordinate-supervisor conscientiousness similarity. *Group and Organization Management*, 23(2), 189–216. <https://doi.org/10.1177/1059601198232006>
- Department of the Army (2012). *Army leadership* (Army Doctrine Publication 6-22). Author.
- Department of the Army. (2015). *Leader development* (Field Manual 6-22). Author
- Department of the Army. (2019). *Army leadership and the profession* (Army Doctrine Publication 6-22). Author.
- Department of the Army. (2019). *Officer professional development and career management* (Pamphlet 600-3). Author.
- *Detrick P., Chibnall, J. T., & Luebbert (2005). Relationship between personality and academy performance. *Applied H.R.M. Research*, 10(2), 99-102. http://applyhrm.asp.radford.edu/Volume%2010/MS%2010_2_%20Detrick.pdf
- *De Vries, R. E. (2012). Personality predictors of leadership styles and the self–other agreement problem. *The Leadership Quarterly*, 23(5), 809-821.

<https://doi.org/10.1016/j.leaqua.2012.03.002>

- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880-896. <https://doi.org/10.1037/0022-3514.93.5.880>.
- Drasgow, F., Nye, C. D., Guo, J., & Tay, L. (2009). Cheating on proctored tests: The other side of the unproctored debate. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2(1), 46-48. Doi: 10.1111/j.1754-9434.2008.01106.x
- Drasgow, F., Nye, C. D., O'Brien, E., Chernyshenko, O. S., Stark, S. (in preparation). *Improving the Tailored Adaptive Personality Assessment System (TAPAS)*. Technical Report for the U.S. Army Research Institute for the Behavioral and Social Sciences.
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Driskell J. E., Hogan J., Salas E., & Hoskin, B. (1994). Cognitive and personality predictors of training performance. *Military Psychology*, 6(1), 31-46. https://doi.org/10.1207/s15327876mp0601_2
- *Eid, J., Johnsen, B. H., Bartone, P. T., & Nissestad, O. A. (2008). Growing transformational leaders: Exploring the role of personality hardiness. *Leadership & Organization Development Journal*, 29(1), 4-23. <https://doi.org/10.1108/01437730810845270>
- *Fearing, B. K. (2015). *Moderated mediation of leader's traits and effectiveness: The role of stress*. [Doctoral dissertation, Illinois Institute of Technology].
- *Flemming, W. D. (April 2004). *Predicting leadership effectiveness: Contributions of critical thinking, personality and derailers* [Paper presentation]. The 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL, United States.
- *Føllesdal, H., & Hagtvet, K. (2013). Does emotional intelligence as ability predict transformational leadership? A multilevel approach. *The Leadership Quarterly*, 24(5), 747-762. <https://doi.org/10.1016/j.leaqua.2013.07.004>
- *Fosse, T. H., Buch, R., Säfvenbom, R., & Martinussen, M. (2015). The impact of personality and self-efficacy on academic and military performance: The mediating role of self-efficacy. *Journal of Military Studies*, 6(1), 47-69. <https://doi.org/10.1515/jms-2016-0197>
- *Foti, R. J., & Hauenstein, N. M. A., (2007). Pattern and Variable Approaches in Leadership Emergence and Effectiveness. *Journal of Applied Psychology*, 92(2), 347-355. <https://doi.org/10.1037/0021-9010.92.2.347>

- *Gellatly, I. R., Paunonen, S. V., Meyer, J. P., Jackson, D. N., & Goffin, R. D. (1991). Personality, vocational interest, and cognitive predictors of managerial job performance and satisfaction. *Personality and Individual Differences*, 12(3), 221-231. [https://doi.org/10.1016/0191-8869\(91\)90108-N](https://doi.org/10.1016/0191-8869(91)90108-N)
- *Goffin, R. D., & Anderson, D. W. (2007). The self-rater's personality and self-other disagreement in multi-source performance ratings: Is disagreement healthy? *Journal of Managerial Psychology*, 22(3), 271-289. <https://doi.org/10.1108/02683940710733098>
- Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216-1229. doi.org/10.1037/0022-3514.59.6.1216
- Goldberg, L. R. (2018). *Documentation and Sample Demographics*. Harvard Dataverse.
- *Grimstad, Ø., & Haugen, E. (2017). *Personality and performance ratings in a military training context: a study of personality predictions of interview ratings of leader potential, and military performance in a longitudinal perspective* [Master's thesis, BI Norwegian Business School].
- *Grunes, P. (2011). *An examination of the relationship between emotional intelligence, leadership style and perceived leadership outcomes in Australian educational institutions* [Doctoral dissertation, Queensland University of Technology].
- *Hakstian, A. R., Scratchley, L. S., MacLeod, A. A., Tweed, R. G. & Siddarth, S. (1997). Selection of telemarketing employees by standardized assessment procedures. *Psychology & Marketing*, 14(7), 703-726. [https://doi.org/10.1002/\(SICI\)1520-6793\(199710\)14:7<703::AID-MAR5>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1520-6793(199710)14:7<703::AID-MAR5>3.0.CO;2-K)
- *Hall, B. F. (1994). *Leader traits and behavior as determinants of leadership effectiveness in a military academy*. Air Force Institute of Technology.
- *Harms, P. D., Roberts, B. W., & Wood, D. (2007). Who shall lead? An integrative personality approach to the study of the antecedents of status in informal social organizations. *Journal of Research in Personality*, 41(3), 689-699. <https://doi.org/10.1016/j.jrp.2006.08.001>
- Harvey, J., Hyland, J., Nye, C. D., Patterson, J., & Morath, R. (2018, October). *Marine Corps non-cognitive testing: Final briefing/report*. Presentation given to the Marine Corps Recruiting Command.
- *Hendricks, J. W., & Payne, S. C. (2007). Beyond the big five: Leader goal orientation as a predictor of leadership effectiveness. *Human Performance*, 20(4), 317-343. <https://doi.org/10.1080/08959280701521983>

- *Hetland, H. & Sandal, G. M. (2003). Transformational leadership in Norway: Outcomes and personality correlates. *European Journal of Work and Organizational Psychology*, 12(2), 147-170. <https://doi.org/10.1080/13594320344000057>
- *Hoffman, B. J., & Frost, B. C. (2006). Multiple intelligences of transformational leaders: An empirical examination. *International Journal of Manpower*, 27(1), 37-51. <https://doi.org/10.1108/01437720610652826>
- Hoffman, B. J., Woehr, D. J., Maldagen-Youngjohn, R., & Lyons, B. D. (2011). Great man or great myth? A quantitative review of the relationship between individual differences and leader effectiveness. *Journal of Occupational and Organizational Psychology*, 84(2), 347-381. <https://doi.org/10.1348/096317909X485207>
- Hogan, R. (2002). *The Hogan personality inventory manual*. Hogan Assessment Systems.
- *Hogan, J., Hogan, R. & Murtha, T. (1992). Validation of a personality measure of managerial performance. *Journal of Business & Psychology*, 7(2), 225-237. <https://doi.org/10.1007/BF01013931>
- *Hong, Y., Catano, V. M., & Liao, H. (2011). Leader emergence: The role of emotional intelligence and motivation to lead. *Leadership & Organization Development Journal*, 32(4), 320-343. <https://doi.org/10.1108/01437731111134625>
- *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). U.S. Army Research Institute for the Behavioral and Social Sciences.
- Horn, B. (2006). "But . . . It's not my fault!"—Disobedience as a function of fear. In C. L. Mantle (Ed.), *The unwilling and the reluctant: Theoretical perspectives on disobedience in the military* (pp. 169–192). Canadian Defence Academy Press.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications, Inc.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91(3), 594-612. <https://doi.org/10.1037/0021-9010.91.3.594>
- *Hwang, G.S. (1988). *Validity of the California Psychological Inventory for Police Selection* [Unpublished master's thesis. North Texas State University].
- *Judge, T. A., & Bono, J. E., (2000). Five factor model of personality and transformational leadership. *Journal of Applied Psychology*, 85(5), 751-765. <https://doi.org/10.1037/0021-9010.85.5.751>

- *Judge, T. A., LePine, J. A., & Rich, B. L. (2006). Loving yourself abundantly: relationship of the narcissistic personality to self-and other perceptions of workplace deviance, leadership, and task and contextual performance. *Journal of Applied Psychology*, 91(4), 762-776. <https://doi.org/10.1037/0021-9010.91.4.762>
- *Judge, T.A., Thoresen, C.J., Pucik, V., & Welbourne, T.M. (1999). Managerial coping with organizational change: A dispositional perspective. *Journal of Applied Psychology*, 84(1), 107-122. <https://dx.doi.org/10.1037/0021-9010.84.1.107>
- *Kalshoven, K., Den Hartog, D. N., & De Hoogh, A. H. (2011). Ethical leader behavior and big five factors of personality. *Journal of Business Ethics*, 100(2), 349-366. <https://doi.org/10.1007/s10551-010-0685-9>
- Kandler, C., Bleidorn, W., Riemann, R., Spinath, F. M., Thiel, W., & Angleitner, A. (2010). Sources of cumulative continuity in personality: A longitudinal multiple-rater twin study. *Journal of Personality and Social Psychology*, 98(6), 995-1008. <https://doi.org/10.1037/a0019558>
- Kelley, T. L. (1927). *The interpretation of educational measurements*. World Book Company.
- *Kelly, D. R., Matthews, M. D., & Bartone, P. T. (2014). Grit and hardiness as predictors of performance among West Point cadets. *Military Psychology*, 26(4), 327-342. <https://doi.org/10.1037/mil0000050>
- *Knapp, D. J., McCloy, R. A., Heffner, T. S. (2004). *Validation of Measures Designed to Maximize 21st-Century Army NCO Performance* (Technical Report 1145). U.S. Army Research Institute for the Behavioral and Social Sciences.
- Komar, S., Komar, J. A., Robie, C., & Taggar, S. (2010). Speeding personality measures to reduce faking: A self-regulatory model. *Journal of Personnel Psychology*, 9(3), 126-137. <https://doi.org/10.1027/1866-5888/a000016>
- *Kubisiak, U. C., Horgen, K. E., Connell, P. W., Lentz, E., Zu, X., Borman, W. C., ... & Young, M. C. (2005). *Concurrent validation of the NLSI for US Army Drill Sergeants* (Study Note 2006-01). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Lall, R., Holmes, E. K., Brinkmeyer, K. R., Johnson, W. B., & Yatko, B. R. (1999). Personality characteristics of future military leaders. *Military Medicine*, 164(12), 906-910. <https://doi.org/10.1093/milmed/164.12.906>
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329-358. https://doi.org/10.1207/s15327906mbr3902_8

- Lee, P., Joo, S. H., Stark, S., & Chernyshenko, O. S. (2019). GGUM-RANK statement and person parameter estimation with multidimensional forced choice triplets. *Applied Psychological Measurement*, 43(3), 226-240. <https://doi.org/10.1177/0146621618768294>
- *Legree, P. J., Kilcullen, R. N., Putka, D. J., & Wasko, L. E. (2014). Identifying the leaders of tomorrow: Validating predictors of leader performance. *Military Psychology*, 26(4), 292-309. <https://doi.org/10.1037/mil0000054>
- *Legree, P. J., Purl, J., Kilcullen, R. N., & Young, M. C. (2019). *Cadet training and personality metrics longitudinally predict officer in-unit performance: R = .37* (Technical Report 1377). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Leslie, J. B., Dalton, M., Ernst, C., & Deal, J. (2002). *Managerial effectiveness in a global context* (CCL No. 184). Center of Creative Leadership. <https://cclinnovation.org/wp-content/uploads/2020/03/managerial-effectiveness-in-a-global-context.pdf>
- *Lim, B. C., & Ployhart, R. E. (2004). Transformational leadership: Relations to the five-factor model and team performance in typical and maximum contexts. *Journal of Applied Psychology*, 89(4), 610–621. <https://doi.org/10.1037/0021-9010.89.4.610>
- *Lindberg, J. T. (2006). *The relative and incremental validity of the Big Five and maladaptive personality characteristics for predicting leadership effectiveness* [Doctoral dissertation, North Carolina State University].
- Lord, F., & Novick, M. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley.
- *Lyons, T. J., Bayless, A. & Park, R. K. (2001). Relationship of cognitive, biographical, and personality measures with the training and job performance of detention enforcement officers in a federal government agency. *Applied H.R.M. Research*, 6(1-2), 67-70.
- Machida, M., & Schaubroeck, J. (2011). The role of self-efficacy beliefs in leader development. *Journal of Leadership & Organizational Studies*, 18(4), 459-468. <https://doi.org/10.1177/1548051811404419>
- *MacLaren, N. G., Yammarino, F. J., Dionne, S. D., Sayama, H., Mumford, M. D., Connelly, S., et al. (2020). Testing the babble hypothesis: Speaking time predicts leader emergence in small groups. *The Leadership Quarterly*, 31(5), Article 101409. <https://doi.org/10.1016/j.leaqua.2020.101409>
- *Maddi, S. R., Matthews, M. D., Kelly, D. R., Villarreal, B., & White, M. (2012). The role of hardiness and grit in predicting performance and retention of USMA cadets. *Military Psychology*, 24(1), 19-28. <https://doi.org/10.1080/08995605.2012.639672>
- *Mael, F. A., & Hirsch, A. C. (1993). Rainforest empiricism and quasi-rationality: two approaches to objective biodata. *Personnel Psychology*, 46(4), 719-738.

<https://doi.org/10.1111/j.1744-6570.1993.tb01566.x>

- *Marcus, B., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (2007). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Human Performance*, 20(3), 275-285. <https://doi.org/10.1080/08959280701333362>
- *Matthew, C. T. (2009). Leader creativity as a predictor of leading change in organizations. *Journal of Applied Social Psychology*, 39(1), 1-41. <https://doi.org/10.1111/j.1559-1816.2008.00427.x>
- *Mayer, J. D., & Skimmyhorn, W. (2017). Personality attributes that predict cadet performance at West Point. *Journal of Research in Personality*, 66, 14-26. <https://doi.org/10.1016/j.jrp.2016.10.012>
- *McClean, E. J., Martin, S. R., Emich, K. J., & Woodruff, C. T. (2018). The social consequences of voice: An examination of voice type and gender on status and subsequent leader emergence. *Academy of Management Journal*, 61(5), 1869-1891. <https://doi.org/10.5465/amj.2016.0148>
- *McCormack, L., & Mellor, D. (2002). The role of personality in leadership: An application of the five-factor model in the Australian military. *Military Psychology*, 14(3), 179-197. https://doi.org/10.1207/S15327876MP1403_1
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81-90. <https://doi.org/10.1037/0022-3514.52.1.81>
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1-10. <https://doi.org/10.1016/j.intell.2008.08.004>
- Meade, A. W., Pappalardo, G., Braddy, P. W., & Fleenor, J. W. (2020). Rapid response measurement: Development of a faking-resistant assessment method for personality. *Organizational Research Methods*, 23(1), 181-207. <https://doi.org/10.1177/1094428118795295>
- *Milan, L. M. (2002). *Analog scales as temperament measures in the Baseline Officer Longitudinal Data Set (BOLDS)* (Technical Report 1126). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Miller, M. L., James, D. R., & Cobb, M. G. (2011). *The impact of accelerated promotion rates on Drill Sergeant performance* (Research Report 1935). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Mills, C. J., & Bohannon, W. E. (1980). Personality characteristics of effective state police officers. *Journal of Applied Psychology*, 65(6), 680-684.

<https://doi.org/10.1037/0021-9010.65.6.680>

- *Mohan, R., Chandran, R. and Arumugam, S. (2016, Nov.). *Personality predictors of selection of women officers' in Indian Armed Forces* [Paper presentation]. The 58th Annual Conference of the International Military Testing Association. New Delhi, India.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729. <https://doi.org/10.1111/j.1744-6570.2007.00089.x>
- *Nahrgang, J. D., Morgeson, F. P., & Ilies, R. (2009). The development of leader-member exchanges: Exploring how personality and performance influence leader and member relationships over time. *Organizational Behavior and Human Decision Processes*, 108(2), 256-266. <https://doi.org/10.1016/j.obhdp.2008.09.002>
- *Neubert, M.J., & Taggar, S. (2004). Pathways to informal leadership: The moderating role of gender on the relationship of individual differences and team member network centrality to informal leadership emergence. *The Leadership Quarterly*, 15(2), 175-194. <https://doi.org/10.1016/j.leaqua.2004.02.006>
- *Ng, K. Y., Ang, S., & Chan, K. Y. (2008). Personality and leader effectiveness: a moderated mediation model of leadership self-efficacy, job demands, and job autonomy. *Journal of Applied Psychology*, 93(4), 733-743. <https://doi.org/10.1037/0021-9010.93.4.733>
- *Nikolaou, I., & Robertson IV, I. T. (2001). The five-factor model of personality and work behaviour in Greece. *European Journal of Work and Organizational Psychology*, 10(2), 161-186. <https://doi.org/10.1080/13594320143000618>
- 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). U.S. Army Research Institute for the Behavioral and Social Sciences.
- Nye, C. D., Joo, S. H., Zhang, B., & Stark, S. (2020). Advancing and evaluating IRT model data fit indices in organizational research. *Organizational Research Methods*, 23(3), 457-486. <https://doi.org/10.1177/1094428119833158>
- *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). U.S. Army Research Institute for the Behavioral and Social Sciences.
- Nye, C. D., & Roberts, B. W. (2013). A developmental perspective on the importance of personality for understanding workplace behavior. In N. Christiansen & R. Tett (Eds.), *Handbook of Personality at Work* (pp. 796-818). Routledge.

- Nye, C. D., Rounds, J., Kirkendall, C., Drasgow, F., Chernyshenko, O. S., & Stark, S. (2019). *Adaptive Vocational Interest Diagnostic: Development and initial validation*. (Technical Report 1378). U.S. Army Research Institute for the Behavioral and Social Sciences. doi: 10.4324/9781315678924
- *Nye, C. D., White, L. A., Drasgow, F., Chernyshenko, O. S., Stark, S., Graves, C. R., & Muhammad, R. S. (2020). *Examining enhanced suitability screening for predicting drill sergeant training and job outcomes* (Technical Report 1388). U.S. Army Research Institute for the Behavioral and Social Sciences.
- Nye, C. D., White, L. A., Drasgow, F., Prasad, J., Chernyshenko, O. S., & Stark, S. (2020). Examining personality for the selection and classification of soldiers: Validity and differential validity across jobs. *Military Psychology*, 32(1), 60-70.
<https://doi.org/10.1080/08995605.2019.1652482>
- *O'Connor, P. J., & Jackson, C. J. (2010). Applying a psychobiological model of personality to the study of leadership. *Journal of Individual Differences*, 31(4), 185-197.
<https://doi.org/10.1027/1614-0001/a000027>
- *Oh, I. S., Le, H., Whitman, D. S., Kim, K., Yoo, T. Y., Hwang, J. O., & Kim, C. S. (2014). The incremental validity of honesty–humility over cognitive ability and the big five personality traits. *Human Performance*, 27(3), 206-224.
<https://doi.org/10.1080/08959285.2014.913594>
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.
- Paullin, C., Sinclair, A. L., Moriarty, K. O., Vasilopoulos, N. L., Campbell, R. C., & Russell, T. L. (2011). *Army officer job analyses: Identifying performance requirements to inform officer selection and assignment* (Technical Report 1295). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Paunonen, S. V., Lönnqvist, J., Verkasalo, M., Leikas, S., & Nissinen, V. (2006). Narcissism and emergent leadership in military cadets. *Leadership Quarterly*, 17(5), 475-486.
<https://doi.org/10.1016/j.leaqua.2006.06.003>
- *Perkins, A. M., & Corr, P. J. (2006). Cognitive ability as a buffer to neuroticism: Churchill's secret weapon? *Personality and Individual Differences*, 40(1), 39-51.
<https://doi.org/10.1016/j.paid.2005.05.012>
- *Peterson, S. J., Walumbwa, F. O., Byron, K., & Myrowitz, J. (2009). CEO positive psychological traits, transformational leadership, and firm performance in high-technology start-up and established firms. *Journal of Management*, 35(2), 348-368.
<https://doi.org/10.1177/0149206307312512>

- *Ployhart, R.E., Lim, B.C. & Chan K.Y. (2001). Exploring relations between typical and maximal performance ratings and the five-factor model of personality. *Personnel Psychology*, 54(4), 809-843. <https://doi.org/10.1111/j.1744-6570.2001.tb00233.x>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology*, 74(5), 770-780. <https://doi.org/10.1037/0021-9010.74.5.770>
- Pugh, G. (1985). The California Psychological Inventory and police selection. *Journal of Police Science and Administration*, 13(2), 172-177.
- Rindler, S. E. (1979). Pitfalls in assessing test speededness. *Journal of Educational Measurement*, 16(4), 261-270. <https://doi.org/10.1111/j.1745-3984.1979.tb00107.x>
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses using item response theory. *Applied Psychological Measurement*, 24(1), 2-32. <https://doi.org/10.1177/01466216000241001>
- *Rockstuhl, T., Seiler, S., Ang, S., Van Dyne, L., & Annen, H. (2011). Beyond general intelligence (IQ) and emotional intelligence (EQ): The role of cultural intelligence (CQ) on cross-border leadership effectiveness in a globalized world. *Journal of Social Issues*, 67(4), 825-840. <https://doi.org/10.1111/j.1540-4560.2011.01730.x>
- *Rose, M. R., Barron, L. G., Carretta, T. R., Arnold, R. D., & Howse, W. R. (2014). Early identification of unmanned aircraft pilots using measures of personality and aptitude. *The International Journal of Aviation Psychology*, 24(1), 36-52. <https://doi.org/10.1080/10508414.2014.860849>
- *Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal*, 48(5), 845-858. <https://doi.org/10.5465/AMJ.2005.18803926>
- *Rueb, J. D., Erskine, H. J., & Foti, R. J. (2008). Intelligence, dominance, masculinity, and self-monitoring: Predicting leadership emergence in a military setting. *Military Psychology*, 20(4), 237-252. <https://doi.org/10.1080/08995600802345139>
- *Russell, T. L., & Tremble, T. (2011). *Development and validation of measures for selecting Soldiers for the Officer Candidate School* (Study Note 2011-02). U.S. Army Research Institute for the Behavioral and Social Sciences.

- *Russell, T. L., Paullin, C. J., Legree, P. J., Kilcullen, R. N., & Young, M. C. (2017). *Identifying and validating selection tools for predicting officer performance and retention* (Research Note 2017-01). U.S. Army Research Institute for the Behavioral and Social Sciences.
- *Salgado, J. F. & Rumbos, A. (1997). Personality and job performance in financial service managers. *International Journal of Selection and Assessment*, 5(2), 91-100. <https://doi.org/10.1111/1468-2389.00049>
- *Saville, P., Sik, G., Nyfield, G., Hackston, J., & MacIver, R. (1996). A demonstration of the validity of the Occupational Personality Questionnaire (OPQ) in the measurement of job competencies across time and in separate organisations. *Applied Psychology*, 45(3), 243-262. <https://doi.org/10.1111/j.1464-0597.1996.tb00767.x>
- *Schaumburg, R. L., & Flynn, F. J. (2012). Uneasy lies the head that wears the crown: The link between guilt proneness and leadership. *Journal of Personality and Social Psychology*, 103(2), 327-342. <https://doi.org/10.1037/a0028127>
- *Schmit, M., Kihm, J. A., & Robie, C. (2000). Development of a global measure of personality. *Personnel Psychology*, 53(1), 153-193. <https://doi.org/10.1111/j.1744-6570.2000.tb00198.x>
- *Shao, L., & Webber, S. (2006). A cross-cultural test of the ‘five-factor model of personality and transformational leadership’. *Journal of Business Research*, 59(8), 936-944. <https://doi.org/10.1016/j.jbusres.2006.02.005>
- *Smith, J. A., & Foti, R. J. (1998). A pattern approach to the study of leader emergence. *Leadership Quarterly*, 9(2), 147-160. [https://doi.org/10.1016/S1048-9843\(98\)90002-9](https://doi.org/10.1016/S1048-9843(98)90002-9)
- Society for Industrial and Organizational Psychology (SIOP) (2018). *Principles for the Validation and Use of Personnel Selection Procedures* (5th ed.). Author.
- *Sosik, J. J., & Megerian, L. E. (1999). Understanding leader emotional intelligence and performance: the role of self-other agreement on transformational leadership perceptions. *Group and Organization Management*, 24(3), 367-390. <https://doi.org/10.1177/1059601199243006>
- *Spector, P. E., Schneider, J. R., Vance, C. A., & Hezlett, S. A. (2000). The relation of cognitive ability and personality traits to assessment center performance. *Journal of Applied Social Psychology*, 30(7), 1474-1491. <https://doi.org/10.1111/j.1559-1816.2000.tb02531.x>
- Stark, S. (2002). *A new IRT approach to test construction and scoring designed to reduce the effects of faking in personality assessment* [Doctoral Dissertation, University of Illinois at Urbana-Champaign].
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-

- unidimensional pairwise preference model. *Applied Psychological Measurement*, 29(3), 184-201. <https://doi.org/10.1177/0146621604273988>
- Stark, S., Chernyshenko, O. S., Drasgow, F., & White, L. A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods*, 15(3), 463-487. <https://doi.org/10.1177/1094428112444611>
- Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91(1), 25-39. <https://doi.org/10.1037/0021-9010.91.1.25>
- Stark, S., Chernyshenko, O. S., & Guenole, N. (2011). Can subject matter experts' ratings of statement extremity be used to streamline the development of unidimensional pairwise preference scales? *Organizational Research Methods*, 14(2), 256-278. <https://doi.org/10.1177/1094428109356712>
- Stark, S., & Drasgow, F. (2002). An EM approach to parameter estimation for the Zinnes and Griggs paired comparison IRT model. *Applied Psychological Measurement*, 26(2), 208-227. <https://doi.org/10.1177/01421602026002007>
- *Stewart, G. L., Barrick, M. R., & Parks, L. (2003). *A theoretical and empirical analysis of broad and narrow traits* [Paper presentation]. The Annual Conference for the Society of Industrial and Organizational Psychology, Orlando, FL, United States.
- *Strang, S. E., & Kuhnert, K. W. (2009). Personality and leadership developmental levels as predictors of leader performance. *The Leadership Quarterly*, 20(3), 421-433. <https://doi.org/10.1016/j.leaqua.2009.03.009>
- *Stutzman, N. S. (2017). *Examining the Relationship Between Personality and Performance: Does Personality Predict Performance for Female Leaders?* [Doctoral dissertation, Columbia University].
- *Taggar, S., Hackett, R. & Saha, S. (1999). Leadership emergence in autonomous work teams: Antecedents and outcomes. *Personnel Psychology*, 52(4), 899-926. <https://doi.org/10.1111/j.1744-6570.1999.tb00184.x>
- *Taylor, A. M. (2007). *The validity of personality trait interactions for the prediction of managerial job performance* [Master's Thesis, University of South Florida].
- *Thomas, J. L. (1999). *Personality and motivational predictors of military leadership assessment in the United States Army Reserve Officer Training Corps* [Doctoral dissertation, Wayne State University].

- Trent, J. D., Barron, L. G., Rose, M. R., & Carretta, T. R. (2020). Tailored Adaptive Personality Assessment System (TAPAS) as an indicator for counterproductive work behavior: Comparing validity in applicant, honest, and directed faking conditions. *Military Psychology*, 32(1), 51-59. <https://doi.org/10.1080/08995605.2019.1652481>
- *Van der Linden, D., Oostrom, J. K., Born, M. P., Van der Molen, H. T., & Serlie, A. W. (2014). Knowing what to do in social situations: The general factor of personality and performance on situational judgment tests. *Journal of Personnel Psychology*, 13(2), 107–115. <https://doi.org/10.1027/1866-5888/a000113>
- *Van Iddekinge, C. H., Ferris, G. R., & Heffner, T. S. (2009). Test of a multistage model of distal and proximal antecedents of leader performance. *Personnel Psychology*, 62(3), 463-495. <https://doi.org/10.1111/j.1744-6570.2009.01145.x>
- *Van Woerkom, M., & De Reuver, R. S. (2009). Predicting excellent management performance in an intercultural context: a study of the influence of multicultural personality on transformational leadership and performance. *The International Journal of Human Resource Management*, 20(10), 2013-2029. <https://doi.org/10.1080/09585190903175589>
- *Van Emmerik, I. H., & Euwema, M. C. (2007). Who is offering a helping hand? Associations between personality and OCBs, and the moderating role of team leader effectiveness. *Journal of Managerial Psychology*, 22(6), 530-548. <https://doi.org/10.1108/02683940710778422>
- *Von Wittich, D., & Antonakis, J. (2011). The KAI cognitive style inventory: Was it personality all along? *Personality and Individual Differences*, 50(7), 1044-1049. <https://doi.org/10.1016/j.paid.2011.01.022>
- Wainer, H., Vevea, J. L., Camacho, F., Reeve, B. B., Rosa, K., Nelson, L., Swygert, K. A., & Thissen, D. (2001). Augmented Scores — 'Borrowing Strength' to computer scores based on small numbers of items. In D. E. Thissen & H. E. Wainer (Eds.), *Test scoring* (pp. 329-370). Routledge.
- *Waldman, D. A., Atwater, L. E., Davidson, R. E. (2004). The role of individualism and the five factor model in the prediction of performance in a leaderless group discussion. *Journal of Personality*, 72(1), 1-28. <https://doi.org/10.1111/j.0022-3506.2004.00254.x>.
- *Walumbwa, F. O., & Schaubroeck, J. (2009). Leader personality traits and employee voice behavior: mediating roles of ethical leadership and work group psychological safety. *Journal of Applied Psychology*, 94(5), 1275-1286. <https://doi.org/10.1037/a0015848>
- *Washington, R. R., Sutton, C. D., & Feild, H. S. (2006). Individual differences in servant leadership: The roles of values and personality. *Leadership & Organization Development Journal*, 27(8), 700-716. <https://doi.org/10.1108/01437730610709309>

- 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. doi:10.1111/j.1754-9434.2008.00049.x
- Woo, S. E., Chernyshenko, O. S., Longley, A., Zhang, Z. X., Chiu, C. Y., & Stark, S. E. (2014). Openness to experience: Its lower level structure, measurement, and cross-cultural equivalence. *Journal of Personality Assessment*, 96(1), 29-45. doi.org/10.1080/00223891.2013.806328
- *Wright T. A., Cropanzano R., & Meyer, D. G. (2004). State and trait correlates of job performance: A tale of two perspectives. *Journal of Business and Psychology*, 18(3), 365-383. <http://dx.doi.org/10.1023/B:JOB.0000016708.22925.72>
- *Wright, B. S., Doerner, W. G., & Speir, J. C. (1990). Pre-employment psychological testing as a predictor of police performance during an FTO program. *American Journal of Police*, 9(4), 65-83.
- *Xu, X., Yu, F., & Shi, J. (2011). Ethical leadership and leaders' personalities. *Social Behavior and Personality: An International Journal*, 39(3), 361-368. <https://doi.org/10.2224/sbp.2011.39.3.361>
- *Yeh, S. C. J., Yuan, K. S., Chen, S. H. S., Lo, Y. Y., Chou, H. C., Huang, S., & Wan, T. T. (2016). The moderating effect of leadership on the relationship between personality and performance. *Journal of Nursing Management*, 24(7), 869-883. <https://doi.org/10.1111/jonm.12391>
- *Young, M., & Dulewicz, V. (2009). A study into leadership and management competencies predicting superior performance in the British Royal Navy. *Journal of Management Development*, 28(9), 794-820. <https://doi.org/10.1108/02621710910987665>
- *Zaccaro, S.J., Gilrane, V.L., Robbins, J.M., Bartholomew, L.N., Young, M.C., Kilcullen, R.N., Connelly, S., & Young, W.Y. (2012). *Officer individual differences: predicting long-term continuance and performance in the U. S Army* (Technical Report 1324). U. S. Army Research Institute for the Behavioral and Social Sciences.
- Zhang, B., Sun, T., Drasgow, F., Chernyshenko, O. S., Nye, C. D., Stark, S., & White, L. A. (2020). Though forced, still valid: Psychometric equivalence of forced-choice and single-statement measures. *Organizational Research Methods*, 23(3), 569-590. doi.org/10.1177/1094428119836486
- Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika*, 39(3), 327-350. <https://doi.org/10.1007/BF02291707>