Technical Report 1427

Validation Research for the FA48 Officer Selection Process

Kristin M. Repchick
Tiffani R. Chen
Stephen J. Zaccaro
Lucy Q. Wang
Desmond Consulting Group

Melissa J. Glorioso
U.S. Army Research Institute

January 2023

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

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

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE L. ZBYLUT, PhD.
Director

DISPOSITION

REPORT DOCUMENTATION PAGE

1. REPORT DATE (Month Year)	2. REPORT TYPE	3. DATES COVERED (Month Year)	
January 2023	Final	**START DATE** September 2018	**END DATE** March 2023

4. TITLE AND SUBTITLE
Validation Studies for the FA48 Officer Selection Process

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. COOPERATIVE AGREEMENT NUMBER
W911-NF-18-C-0076		

5d. PROGRAM ELEMENT NUMBER	5e. PROJECT NUMBER	5f. TASK NUMBER	5g. WORK UNIT NUMBER
6330007	A792	1030	

6. AUTHOR(S)
Repchick, Kristin R., Chen, Tiffani R., Zaccaro, Stephen J., Wang, Lucy Q., Glorioso, Melissa, J.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Desmond Consulting Group 14608 Snapdragon Circle North Potomac, MD 20878	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop: 5610) Fort Belvoir, Virginia 22060-5610	ARI	Technical Report 1427

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

13. SUPPLEMENTARY NOTES
ARI Research POC: Melissa Glorioso, Selection and Assignment Research Unit

14. ABSTRACT
The FA48 role requires a unique and highly specialized set of skills and attributes for success, as the job environment can be challenging, ill-defined, and ambiguous. FA48 officers are frequently assigned to positions outside of the United States and are required to interact with foreign officials and representatives from various countries, as well as personnel from agencies outside of the DoD. As FA48s are representing the U.S. Army in a foreign environment with minimal supervision or guidance, it is imperative that these officers possess the knowledge, skills, and abilities (KSAs) needed to be successful in this role. However, there is currently no standardized selection or assessment tool to identify officers for these roles. This report describes the predictive ability of a series of cognitive and noncognitive measures of FA48 KSAs as part of an ongoing effort by the U. S. Army Research Institute for the Behavioral and Social Sciences to develop and validate a selection tool for FA48 applicants. Results of a concurrent validation effort revealed that several individual difference measures showed promise for predicting FA48 performance. Additional data collection is needed to determine the best combination of measures for inclusion in a selection tool for FA48s.

15. SUBJECT TERMS
Officer branching, Classification, Functional area, Noncognitive assessment

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**C. THIS PAGE** Unclassified	Unclassified Unlimited	111

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER (Include area code)
Dr. Tonia Heffner	703-545-4408

Technical Report 1427

Validation Studies for the FA48
Officer Selection Process

Kristin M. Repchick
Tiffani R. Chen
Stephen J. Zaccaro
Lucy Q. Wang
Desmond Consulting Group

Melissa J. Glorioso
U.S. Army Research Institute

Selection and Assignment Research Unit
Tonia S. Heffner, Chief

January 2023

VALIDATION STUDIES FOR THE FA48 OFFICER SELECTION PROCESS

EXECUTIVE SUMMARY

Research Requirement

Foreign Area Officers (FA48s) are Soldiers who provide leadership and regional expertise in political-military operations. As such, FA48s in the Army require a unique and specialized skillset for success. Presently, the use of a valid and standardized selection assessment for officer assignments is not widespread across the Army; thus, current Army assessments likely do not include the unique KSAs needed for the FA48 job. An assessment tool measuring the KSAs most critical for performance in the FA48 role could significantly improve the application process and ensure the selection of candidates who are most likely to perform well, both in training and in the FA48 role. For these reasons, the overarching goal of this research effort is to identify the most critical attributes for FA48 job performance and subsequently develop and validate a selection tool to measure these attributes as a part of the application process.

Procedure

To gain deeper insight into what job performance entails for an FA48, as well as the needed KSAs to enable high levels of job performance, this research effort began with a literature review, followed by interviews with FA48 subject matter experts and the administration of a KSA survey. In particular, findings from these efforts were utilized to (1) determine the attributes that are most important for FA48 performance, (2) determine the attribute measures that should be included in an assessment battery, and (3) develop or select FA48 criterion measures to validate the assessment battery.

Following these procedures, the finalized assessment battery was assembled and subsequently administered to current FA48s for concurrent validation to evaluate the psychometric properties of the measures. Supervisor and peer ratings of performance were collected for FA48s who completed the survey to provide preliminary validation of the assessment. That is, to evaluate the validity of the developed assessment battery for predicting FA48 officer performance, a concurrent validation was conducted. Assessment measures administered to active duty FA48 officers were linked to current job performance criteria provided by the FA48 officer, their senior rater, and their peers to inform the best combination of assessment scores.

Findings

Multiple scales showed value in predicting self-rated performance criteria. Specifically, the biodata scales showed the most promise in predicting self-evaluations of FA48 performance as indicated by an eight-dimension rating scale and overall performance compared to peers. Written communication, oral communication and stress tolerance accounted for the largest portions of unique variance in these two criteria. Cognitive and ethical measures also showed indication of predictability, demonstrating significant or near-significant relationships with ratings on past officer evaluation reports (OERs). The figural series measure uniquely accounted

for 8% of the variance in ratings on past OERs, suggesting that cognitive abilities such as critical thinking show promise for predicting FA48 performance. While we were unable to examine the validity of predictor scales for other-rated performance measures, the results of the analyses conducted provide insight into the measures that should be included in a standardized selection tool for FA48s.

Utilization and Dissemination of Findings

The findings in this report offer several contributions toward development of a selection tool to measure the most critical attributes for FA48 job performance. Consideration of results across efforts (i.e., literature review, SME interviews, KSA survey, and concurrent validation) can be utilized to inform a longitudinal validation effort with FA48 applicants. Though the concurrent validation faced several limitations, the results provided important insights into the best set of measures for predicting FA48 performance. The majority of predictor measures included in the assessment battery demonstrated adequate psychometric properties, particularly the biodata measures. Criterion-related validity was established for several of the biodata scales. Analyses of ethical predictor measures showed value for their predictability of FA48 performance and also uncovered an interesting pattern of results that warrant further analysis.

VALIDATION STUDIES FOR THE FA48 OFFICER SELECTION PROCESS

CONTENTS

CONTENTS (continued)

APPENDICES

TABLES

CONTENTS (continued)

VALIDATION STUDIES FOR THE FA48 OFFICER SELECTION PROCESS

CHAPTER 1: OVERVIEW AND BACKGROUND

Statement of the Problem

Talent management for rising Army officers entails the appropriate selection and classification of personnel into job assignments that afford developmental opportunities for the skills required to succeed at higher levels of Army leadership. One such job assignment is Functional Area 48 or Foreign Area Officer (i.e., FA48). The FA48 role requires a unique and highly specialized set of skills and attributes for success in the challenging, ill-defined, and ambiguous job environment. FA48 officers often work outside of the United States and interact with foreign officials and stakeholders as representatives of the U. S. Army with minimal guidance. As such, high levels of leadership and technical abilities are required for successful execution of their job duties. To accomplish this objective, Army talent management must use a valid and standardized selection process. The current FA48 selection process does not involve the use of a standardized assessment, which may lead to high variability in performance among selected officers, a particularly undesirable outcome given the high-profile nature of the role.

The FA48 position requires officers to possess and demonstrate a wide range of expertise. Operating across numerous areas of concentration (AOCs), FA48 officers serve as regional experts. According to the Department of the Army Pamphlet 600-3 (DA-Pam 600-3, 2014), FA48 officers "provide leadership and expertise in diverse organizations in joint, interagency, intergovernmental and multinational environments; advise senior leaders as regional experts; and offer unique warfighting competencies—cross-cultural capabilities, interpersonal communications, and foreign-language skills—that are critical to mission readiness of the Army in today's dynamic strategic environment" (p. 276). FA48s need a deep understanding of the political-military situation in their specific region of work, including cultural awareness and understanding, as well as strong cross-cultural communication skills (DA-Pam 600-3, 2014). Further, FA48s must exhibit a highly specialized combination of language skills, comprehensive regional expertise, and professional military skills (Franklin, 2010). These officers must thoroughly understand and apply U.S. and foreign political-military relationships to political-military analysis, policy, and plans, advising senior Army and Civilian leaders on regional military capabilities to aid the execution of security policies (DA-Pam 600-3, 2014). FA48s must have strong situational awareness skills to be able to perceive and understand the elements in their environment to effectively communicate this information to senior Army leaders (Endsley et al., 2000). In addition to these requirements, FA48s must also operate effectively in joint, interagency, intergovernmental and multinational (JIIM) environments, influence partners who they do not directly manage, and leverage capabilities beyond the Army to achieve their objectives (DA-Pam 600-3, 2014). Encompassing all of these duties is the requirement to successfully operate as both a strategic and operational leader in a complex, multinational environment.

Perhaps the most critical role that the FA48 must perform is to provide the linkage between foreign and U.S. political-military institutions (Franklin, 2010). To perform this duty, FA48s must interact with foreign officials and representatives from various countries, building partnerships and enabling foreign military cooperation, training and exercises (DA-Pam 600-3,

2014). A core aspect of the FA48 role is to develop and maintain relationships characterized by mutual respect and trust across cultural boundaries, which can be challenging with language barriers and differing cultural norms across regions. When FA48s begin an assignment in a foreign region, they may face a largely different environment in terms of accepted leadership style, hierarchy, and business conduct (Rockstuhl et al., 2011). FA48 officers must be able to communicate successfully and employ effective negotiating, influencing, or trust-building tactics despite the challenges of the multicultural environment in which they must occur. In a foreign environment, successful leaders often face political and cultural practices different from those of their home country and adapt their behavior accordingly (Rockstuhl et al., 2011). Additionally, FA48s often serve as the sole representative of the U.S. Army with little supervision or guidance. As such, they must maintain an outstanding level of personal discipline and integrity when faced with complex ethical dilemmas and effectively navigate ambiguous situations to make appropriate decisions (Thiel et al., 2012).

To execute these duties and effectively navigate the challenging performance environment, FA48 officers must possess competencies and skills including "cross-cultural capabilities, interpersonal communications, language skills, interagency integration, and regional political-military expertise" (DA-Pam 600-3, 2014, pp. 277). The unique performance requirements of this role require an equally unique set of competencies for success. Thus, the primary aim of this research effort is to determine the knowledge, skills and abilities (KSAs) that are most important for FA48 officers to possess. Once these attributes are identified, the secondary aim is to develop and validate a selection tool to assess these competencies during the Voluntary Transfer Incentive Program (VTIP) application process.

Existing Selection Assessment for FA48

Presently, the use of a valid and standardized selection assessment for officer assignments is not widespread across the U.S. Army. Thus, current assessments likely do not include the unique KSAs needed for the FA48 role. The current VTIP process, which is one of two ways in which FA48s are assessed for selection, does not include a standardized selection tool. Instead, a review board assesses the officers' language skills, evidenced by a degree in a foreign language, high scores on the Defense Language Aptitude Battery (DLAB), Bachelor's degree, and/or GRE scores. The board reviews applicant materials to ensure that applicants meet the minimum requirements and then selects officers for FA48-specific training; a years-long, intensive process that includes the completion of an appropriate graduate degree. Though officers receive specific in-region training after selection, several KSAs that are important for job performance remain unmeasured as a part of the process. Without a standardized selection tool to assess a fuller range of competencies, the potential for costly, inconsistent performance increases. An assessment tool measuring the KSAs most critical for performance in the FA48 role could significantly improve the application process and ensure the selection of candidates who are most likely to perform well, both in training and in the FA48 role. For these reasons, the overarching goal of this research effort is to identify the most critical attributes for FA48 job performance and subsequently develop and validate a selection tool to measure these attributes as a part of the application process.

CHAPTER 2: LITERATURE REVIEW

Purpose

To gain deeper insight into what job performance entails for an FA48, as well as the needed KSAs to enable high levels of job performance, this research effort began with an extensive literature review. Source materials were gathered from peer-reviewed, academic journals, military journals, technical reports, book chapters, periodicals, and conference proceedings that targeted the job functions of FA48 officers and similar leadership and expatriate positions. These sources were then reviewed for relevant attributes and skills that are predictive of FA48 performance. Additionally, a concentrated effort was made to identify (1) predictor measures for the identified attributes and (2) criterion measures and behavioral indicators reflective of effective FA48 job performance. The identified measures are leveraged in the later assessment development phase.

Methodology

Using a keyword search list (see Appendix A), literature from military sources was gathered and reviewed to gain a deeper understanding of the FA48 job environment. Material from the Army, as well as material describing the adjacent roles of FA48s in other services (e.g., Air Force and Navy), were reviewed and included in the literature review database. This background material describing the specific job requirements of FA48s and similar roles informed the subsequent search for sources related to the determined KSAs. Additionally, once the performance criterion space was more clearly defined using these sources, we conducted a search for materials and technical reports describing competencies necessary for performing the job requirements of FA48's (e.g., operating in a multicultural/foreign environment, building and maintaining cooperative relationships with foreign counterparts, strategic leadership, maintaining personal discipline, etc.).

After exhausting the military literature, the search was expanded to academic literature. As KSAs were identified, we searched for studies that included assessments of those KSAs in relevant populations. Several types of samples were identified, including expatriates, managers and leaders in their home country and abroad, managers and leaders working with diverse and multinational teams, and graduate and undergraduate students. For each source material, the following relevant information was documented in a searchable database of pertinent literature, including APA citation; material source (e.g., academic or military); type of material (e.g., theoretical, empirical, review paper); primary and secondary constructs of focus; predictor or criterion measures included; category of the constructs included (e.g., culture competence and cultural intelligence).

In addition to the literature database, a second database was constructed solely from empirical sources to display the validity evidence for each KSA identified in the literature review, including the study source, KSA measured, relevant criterion assessed, effect sizes for the KSAs and relevant criterion (i.e., the strength of the predictor-criterion relationships), and countries in which the studies were performed or nationality of the participants.

<center>**Narrative Findings**</center>

Narrative findings from the literature review suggested five categories of attributes required for success as an FA48 officer, each described below: culture competence and cultural intelligence, ethical conduct, interpersonal and relationship-oriented skills, leading in challenging environments, and temperament.

Culture Competence and Cultural Intelligence

Common across all the job duties of an FA48 is operating in a multicultural environment. The FA48 serves as a regional expert and must advise senior Army and Civilian leaders on political and cultural matters, as well as on regional military capabilities to inform the execution of security policies. In addition to maintaining understanding of their operating area, FA48s must communicate their understanding and collaborate with foreign counterparts. The KSAs most frequently mentioned in the literature as vital to meeting these challenging intercultural performance requirements were cultural competence and cultural intelligence.

Cultural intelligence can be defined as the ability to grasp, reason, and behave effectively in situations characterized by cultural diversity. Falling under the umbrella of cultural competence, cultural intelligence is traditionally thought to have four sub-components: metacognitive, cognitive, motivational, and behavioral (Ang & van Dyne, 2008). These four sub-components of cultural intelligence have been empirically related to several important performance outcomes. Of note, individuals with high cultural intelligence have been shown to demonstrate higher levels of job performance in multinational environments (Chen et al., 2010; Van Woerkom & de Reuver, 2009). Culturally intelligent individuals were also more easily able to adjust to life in a foreign country (Chen et al., 2010). Such individuals are more likely to demonstrate an interest in international affairs and interacting with people of different backgrounds (van der Zee & van Oudenhoven, 2000). Similarly, individuals who possessed high levels of cultural competence, were also more likely to demonstrate higher levels of job performance (Van Woerkom & de Reuver, 2009) and be deemed more effective leaders (Groves & Feyerherm, 2011) in multicultural situations.

The literature review uncovered several additional KSAs deemed important to or indicative of cultural competence, including cultural perspective-taking, cultural flexibility, and (lack of) ethnocentrism. Cultural flexibility is the ability to move across multiple cultural environments with comfort and ease and is empirically related to important intercultural performance requirements, including leadership success in a global environment (Caligiuri & Tarique, 2012), job performance in a foreign environment, and effectively adjusting to the interactional style of a foreign culture (Shaffer et al., 2006). Ethnocentrism is the belief that one's own cultural group is superior to others. High levels of ethnocentrism can be a detriment to performance and adjustment in a multicultural environment, leader success, job performance, and adjustment in multicultural settings (Shaffer et al., 2006). Cultural perspective-taking, defined as the ability to think contextually and flexibly about issues from a foreign counterpart's perspective, is also critical to leader success in a foreign environment, facilitating successful collaboration and interactions with individuals from different cultures, especially where trusting relationships are needed (Mor et al., 2013).

<center>4</center>

Interpersonal and Relationship Oriented Skills

FA48 officers continuously interact with many different individuals, often of foreign cultures. As such, FA48s must create a positive working environment and develop relationships with varied stakeholders (e.g., host nation personnel, superiors) characterized by mutual respect and trust. In conjunction with understanding the military situation in foreign areas, these officers must understand cultural norms to effectively negotiate, influence, and build trust. Hand-in-hand with cultural intelligence and competence is the ability to enact effective interpersonal interactions, which involves skills such as relationship-building, social adroitness, social orientation, and conflict resolution. These skills have all been linked to the successful integration of expatriates into foreign countries in a cross-cultural leadership capacity, with interpersonal skills also predicting effective adaptation to cultural norms and practices expected in interpersonal dealings (Rockstuhl et al., 2011).

Throughout the literature, several KSAs emerged as important for meeting the social and interpersonal job requirements of the FA48 role. For example, political skill is a competency vital to the leadership function of creating and maintaining trusting relationships (Brouer et al., 2012). Defined as the ability to understand social interactions at work and use this understanding to influence others, appear genuine and sincere, and build strong social networks, political skill has been empirically related to the development of relationships characterized by mutual respect and trust (Ferris et al., 2005). Leaders with higher levels of political skill are better able to foster a positive work environment and instill a greater sense of commitment in those with whom they work (Ewen et al., 2013; Treadway et al., 2004). Another KSA important to interacting effectively with others is empathy, which is defined as the ability to understand and see situations from another's point of view. As FA48 officers are required to work with different individuals across different functional and cultural areas, empathy may be critical to understanding and integrating the viewpoints of all involved parties. Empathy has been empirically related to leaders' ability to execute tasks and their ability to create and manage relationships (Kellett et al., 2006). A broader construct frequently mentioned in the literature was social skills or competence. Social skill has been conceptualized as having multiple subfactors, including social expressiveness, social sensitivity, social control, active listening, and emotion regulation (Riggio et al., 2003). Taken together, social skills have been empirically related to leader performance (Riggio et al., 2003). For example, one important aspect of social skills, and social expressiveness in particular, is being able to effectively convey information. As FA48s serve as liaisons and are often required to convey important information to multiple parties, including foreign representatives and officials, it seems likely that both written and oral communication skills are critical to effective social expression.

Ethical Conduct

FA48 officers are typically one of, if not the only, representative of the U.S. Army in their area of concentration (AOC). Conducting oneself ethically and professionally as this representative requires a high level of personal discipline and the ability to make decisions that lead to ethical outcomes. Engaging in unethical behavior is likely a common temptation and could result in disastrous consequences for both the individual and the Army as a whole. There are severe implications for the quality and nature of the relationship between the U.S. and a foreign region if the integrity of the liaison is compromised. Therefore, officers chosen to be

FA48s must navigate ambiguous situations and make appropriate ethical decisions (Thiel et al., 2012). Through review of the literature on ethicality and personal discipline, several KSAs emerged as crucial to this component of the FA48 job, including ethical decision making, ethical leadership, integrity, moral awareness or identity, and Machiavellianism, each of which are briefly described below.

Ethical decision making, or the ability to process and understand ethical challenges and form appropriate judgments for how to respond based on ethical principles, is especially relevant to the FA48 role given the challenging ethical situations they are likely to face (Craft, 2013). To be successful, an FA48 officer must employ a sound decision-making process that reflects the strong ethical ideals of the U.S. Army. Ethical leadership, defined as "the demonstration of normatively appropriate conduct through personal actions and interpersonal relationships, and the promotion of such conduct to followers through two-way communication, reinforcement, and decision making" (Brown et al., 2005, p. 120), has been shown to be important not only for performance, but for maintaining perceptions of one's trustworthiness (De Hoogh & Den Hartog, 2008). Related to perceptions of trustworthiness is the concept of integrity, defined as the tendency to align words with actions, possess honest and strong moral principles, and demonstrate selflessness by doing the right thing regardless of personal and professional consequences. Integrity has been shown to increase trust in and commitment to leaders (Erkutlu & Chafra, 2016; Friedman et al., 2018) and, as such, is of vital importance to the FA48 role, wherein officers must be viewed as a trustworthy and highly ethical person to effectively execute their job duties. That is, FA48 officers may better foster trusting relationships with their foreign counterparts by practicing ethical leadership and demonstrating integrity.

Directly related to ethical behavior is moral awareness or identity, which is indicated by taking responsibility for behaviors, concern about harm to others, and acting in a manner consistent with high ethical standards. Moral awareness has been identified as a (negative) predictor of unethical behavior, including making unethical decisions and abusing leadership positions (Joosten et al., 2014; Moore et al., 2012). Given this, and based on the nature of the FA48 role, individuals high in moral awareness will be more likely to demonstrate ethical behavior on a consistent basis and less likely to participate in unethical behavior.

Finally, the construct of Machiavellianism emerged as a predictor of unethical conduct. Individuals high in Machiavellianism manipulate others in the pursuit of self-interested goals and lack concern for conventional morality. Not surprisingly, this trait has been empirically linked to one's propensity to engage in unethical behavior (Tang et al., 2008). As such, FA48s low in Machiavellianism may be more likely to succeed in the role and refrain from unethical conduct, such as using their power to manipulate others.

Leading in Challenging Environments

Inherently, the FA48 role is a position that requires the demonstration of leadership in an ambiguous and ill-defined environment. FA48 officers must effectively execute leadership functions, such as advising senior Army and civilian leaders on regional and JIIM environment characteristics, including political-military strategy and policy, and develop and implement political-military policy and plans, including the enablement of joint security operations and cross-cultural military partnerships. The literature regarding complex leader performance

requirements suggests that several leadership-related KSAs may be predictive of FA48 performance, including problem-solving skills, strategic thinking, sensemaking, forecasting, cognitive flexibility, tolerance for ambiguity, and adaptability.

In unstructured and uncertain environments, a common function that is critical to effective leadership performance is problem-solving; the ability to analyze problems to make sound decisions and identify strategies to find solutions. Problem-solving skills are consistently deemed important for successful Army leadership, especially at higher levels (e.g., Zaccaro et al., 2012, 2015). Specific aspects of problem-solving, such as forecasting and sense-making, have also been shown to facilitate leading in challenging environments. Forecasting, the ability to make accurate predictions based on observations of the situation at hand, is important for effective problem-solving, as leaders must think through the potential implications of strategic decisions. This ability may be crucial for FA48 officers determining the implications of military policy and plans in their particular regions, as it has been empirically shown to lead to effective decision making and ethicality of decisions (Stenmark et al., 2010). Additionally, sensemaking, which involves developing an understanding of a dynamic situation by creating a mental representation of the important elements in the situation, has been empirically linked to the use of effective and ethical decision-making processes (Caughron et al., 2011). This may be particularly important for FA48 officers' assessment and understanding of their environment, which they use to guide the development and implementation of military strategy through obtaining and synthesizing relevant information.

Attributional complexity, or the "propensity of individuals to infer complex internal and external attributions in interpersonal situations" (Lakshman & Estay, 2015, p. 2), has been shown to be important for leaders, especially those working in cross-cultural contexts (Lakshman, 2013; Sun & Anderson, 2012). Cognitive flexibility is also important, particularly for problem-solving in challenging environments. Cognitive flexibility represents a curiosity or desire to think about the environment in novel ways and develop new approaches to problems. It can be defined as the tendency to entertain new approaches to solving problems, create new plans and ideas, and initiate and accept change and innovation (Putka, 2009; Wasko et al., 2019). Cognitive flexibility has an empirical relationship with international orientation, such that individuals with high levels of cognitive flexibility were more interested in interacting with foreign counterparts and experiencing new cultures (Leone et al., 2005). Both constructs may play important, yet different, roles in FA48s' ability to continuously respond to novel political, military, or other situations in their region.

Finally, adaptability emerged as an important predictor of leader effectiveness, especially in military contexts. Adaptability is the ability to adjust rapidly to new information, such as changing conditions and strategy, or unexpected obstacles, processes, and requirements (Baard et al., 2014). Such conditions are ubiquitous in the challenging multicultural environment within which FA48s must operate. These officers must change their behavior according to their dynamic and ambiguous environment, especially when faced with political and cultural practices different from those of their home country (Rockstuhl et al., 2011), and they must develop and maintain trusting relationships in these environments with important individuals such as foreign officials (Ang et al., 2007). This type of adjustment requires a high level of cultural knowledge and understanding of norms and practices, as well as the ability to execute appropriate and effective behaviors across situations.

Temperament

This final category of KSAs was created to capture the general personality traits that emerged in the literature review as predictive of both performance in a multicultural environment and job performance in general. Specifically, the five-factor model of personality (i.e., the Big 5, McCrae & Costa, 1997) has been extensively examined in various contexts as it relates to job performance, leadership, cultural outcomes, and several other outcomes relevant to the FA48 role. While it is possible that each of the five factors—agreeableness, conscientiousness, extraversion, neuroticism, and openness—has some relationship with performance outcomes that are important for FA48 officers, we specifically focus on conscientiousness, extraversion, and openness due to extensive support found in related research, as described below. Agreeableness and neuroticism may have some small effect, but these are not as well-supported by empirical evidence in the literature.

Conscientiousness is defined as a tendency to strive to do what is right, especially to do one's work or duty well and thoroughly, and has been extensively examined in relation to job performance. Particularly relevant to the FA48 job environment is the link between conscientiousness and effective leadership in a foreign environment, such as managers' task performance in intercultural assignments (Rockstuhl et al., 2015). Conscientiousness is also positively related to the cultural flexibility and overall success of leaders in a global environment (Caligiuri & Tarique, 2012). Additionally important, openness to experience, defined as the tendency to try new experiences and be imaginative and open-minded, is indicative of high interest in experiencing new cultures and easier adjustment to the working environment of an unfamiliar culture (Shaffer et al., 2006). Empirical research has also demonstrated a link between openness and the motivational and behavioral components of cultural intelligence (Ang et al., 2006) and the cultural flexibility of global leaders (Caligiuri & Tarique, 2012), which is another competency that shows promise for predicting FA48 job performance. Finally, extraversion is frequently mentioned in the literature related to multicultural performance and leadership. Extraversion is the tendency to engage in outgoing, talkative, energetic behavior, and to be assertive in social situations. Research has demonstrated that individuals high in extraversion tend to be more effective leaders in multicultural situations (Ng et al., 2008) and to exhibit higher levels of cultural intelligence (Ang et al., 2006) and cultural flexibility (Caligiuri & Tarique, 2012). Based on these findings, these Big 5 personality traits may add unique predictive value for FA48 performance.

Quantitative Findings

Frequencies

To assess the relative importance of the attributes identified in the literature review using a quantitative indicator, a frequency table was created to summarize the counts for appearance of each KSA by category (see Table 1). Overall, the literature review suggests that several KSAs in each of the five categories should be considered important for FA48 performance.

Notably, KSAs in the cultural competence category (e.g., cultural perspective-taking, cultural flexibility, ethnocentrism) and cultural intelligence (i.e., the four-factor model comprising metacognitive, cognitive, motivational, and behavioral; Ang & Van Dyne, 2008)

appeared most often across the literature, suggesting that these KSAs are extremely important for predicting performance in the FA48 role. This finding is unsurprising, given the multicultural job environment of FA48s. Ethical conduct KSAs, including ethical decision-making, self-regulation, integrity, authentic leadership, and Machiavellianism, were also frequently mentioned, suggesting these KSAs are also important for both ethical and interpersonal job performance requirements of FA48s. Interpersonal and relationship-oriented KSAs, including political skill and trust-building, also appeared frequently, as well as KSAs related to successful leadership performance (e.g., leader competencies, strategic thinking, leadership style, problem-solving, sensemaking). Finally, though not mentioned as frequently as others, temperament KSAs (e.g., Big 5 personality factors, self-efficacy), appeared frequently enough to also be considered necessary for high-performing FA48 officers.

Table 1. *Frequency of Cataloged KSAs*

Attribute	Frequency
Culture competence and intelligence	
Cultural competence*	35
Cultural intelligence*	17
Cultural flexibility*	3
Ethnocentrism*	3
Prior international experiences	3
Global identity	1
Ethical conduct	
Ethical decision-making, integrity*	13
Authentic behavior / leadership, ethical leadership*, transparency	9
Dark side personality traits (incl. Machiavellianism*), propensity to morally disengage, propensity to engage in unethical behavior	7
Moral awareness / identity*, value orientation, social responsibility	6
Self-control*	3
Interpersonal and relationship-oriented skills	
Interpersonal skills, social skills*, social intelligence, social competence*, social knowledge	17
Political skill*, negotiation	14
Trust-building behavior	11
Empathy*, perspective taking	7
Communication skills	4
Self-monitoring, social perceptiveness	4
Leading in challenging environments	
Leader competencies (incl. transformational leadership style)	31
Strategic thinking, systems thinking, systems perspective	12
Problem solving skill*	7

Table 1. *(Continued)*

Attribute	Frequency
Sensemaking*	6
Adaptability, behavioral flexibility	6
Sensegiving	2
Forecasting*	2
Cognitive complexity / flexibility*	1
Temperament	
Big 5 factors (incl. conscientiousness*, extraversion*, openness*)	19
Self-efficacy	4
Tolerance for ambiguity*	4
Creativity	3
Self-monitoring	3
Attributional complexity*	1

Note. *Attribute mentioned in empirical studies with effect size estimates listed in Table 2. Articles included in the count for Big 5 factors include both articles that mention all five attributes and those that only include a subset. Of these, conscientiousness, extraversion, and openness were the most frequently mentioned.

Effect Sizes

While frequency provides a good initial assessment of KSA importance, more information is needed to provide a deeper understanding of the role these KSAs play in the FA48 job environment. As such, to assess the importance of each KSA more rigorously, the empirical relationships between the KSAs and relevant fo criteria were evaluated via an examination of effect sizes. Specifically, the strength of relevant KSA-criterion relationships and the context in which these relationships were demonstrated were evaluated. KSAs that demonstrated significant and strong empirical relationships with relevant criteria (e.g., performance in a foreign country, leadership ability, adjustment to a foreign culture) were deemed especially important for FA48 job performance. In addition to the above, information was captured regarding the nationality of the participants and the countries in which the studies were conducted. KSAs that demonstrated strong predictive ability for participants operating in a foreign country were deemed especially critical, as this is a definitional element of the FA48 position.

Table 2 displays the significant effect sizes found for each KSA as a part of the literature review; note that nonsignificant effect sizes were not included in the table. In general, these findings support the conclusions drawn from the frequency assessment and provide additional evidence of the importance of many of the KSAs preliminarily deemed to be important for FA48 job performance. Many of the attributes that were mentioned most often in the literature also showed the strongest empirical relationships with relevant criteria. Culture competence and intelligence clearly emerged as the most critical category of KSAs, as evidenced by the high frequency counts, effect sizes, and ability to predict performance across cultural settings. Almost all performance requirements are qualified by the fact that they must be executed in a foreign

setting. In addition to the cultural KSAs, several variables across the other four categories emerged as having strong empirical evidence supporting their importance (e.g., moral identity / awareness, cognitive flexibility).

The strength of the empirical relationships between the attributes and criteria, as well as the ability to predict effectiveness across countries, can be used to make judgements about which KSAs in each category are most likely predictive of FA48 job performance. As such, based on this empirical evidence, KSAs in the temperament category may be more important than the frequency counts suggest. Several studies in various cultures showed significant relationships between temperament variables and criteria of interest (e.g., tolerance for ambiguity, Zaccaro et al., 2012, 2015). The findings from the effect size table were used in conjunction with the frequency counts and narrative findings to narrow down the list of KSAs for inclusion in an FA48 selection assessment battery.

Summary

The purpose of the literature review was to gain deeper insight into the job performance requirements of the FA48 role and the KSAs necessary to meet those requirements. Both the narrative and quantitative findings resulting from the development of the literature review databases and effect size table provided useful information by which to evaluate the importance of the KSAs identified. The quantitative findings provided criteria by which to evaluate the narrative findings more rigorously. The findings of the literature review informed the subsequent development of the SME interview protocol and the KSA survey, detailed in the following chapters.

Table 2. *Effect Sizes of Cataloged KSAs*

Attribute	Effect Size	Country	Performance Criteria	Source
Culture competence and intelligence				
Cultural competence	$b = 0.27; 0.12$	Expat. sample	Adjustment, job performance	Chen et al., 2010
	$b = 0.12 - 0.34$	Netherlands, Italy	International orientation	Leone et al., 2005
	$b = 0.17 - 0.23$	Expat. sample	Adjustment	Peltokorpi & Froese, 2012
	$R^2 = 0.14 - 0.28$	Netherlands	International orientation, multicultural activity	van der Zee & van Oudenhoven, 2000
	$B = 0.68$	Multicultural	Job performance	Van Woerkom & de Reuver, 2009
	$b = 0.10 - 0.11$	Multicultural	Leader performance	Groves & Feyerherm, 2011
Cultural intelligence	$b = 0.15 - 0.17$, $b = 0.10 - 0.13$ $b = 0.30 - 0.40$	USA, Singapore Multicultural	Adjustment, decision making	Ang et al., 2007
	$b = 0.47$	Multicultural	Team and leader performance	Groves & Feyerherm, 2011
	$b = 0.13 - 0.41$	USA, East Asian countries	Cooperative negotiation, relationship mgmt.	Imai & Gelfand, 2010
	$b = 0.63$	Expat. sample	Adjustment	Lee & Kartika, 2014
	$b = 0.20$	USA	Intercultural cooperation, cultural perspective taking	Mor et al., 2013
	$b = 0.11 - 0.24$	Switzerland	Leader effectiveness	Rockstuhl et al., 2011
Cultural flexibility	$b = 0.17$	Multicultural	Global leadership success	Caligiuri & Tarique, 2012
	$b = 0.11 - 0.23*$, $b = -0.02 - 0.32$	Multicultural	Cultural, interaction, & work adjustment, contextual & task performance	Shaffer et al., 2006
	$b = 0.12$	Multicultural	Global leadership success	Caligiuri & Tarique, 2012

Table 2. *(Continued)*

Attribute	Effect Size	Country	Performance Criteria	Source
Ethnocentrism	$b = 0.08 - 0.13*$, $b = 0.26 - 0.18*$	Multicultural	Cultural, interaction, & work adjustment, contextual & task performance	Shaffer et al., 2006
Ethical conduct				
Integrity	$b = 0.16 - 0.35$	Turkey	Follower org. commitment	Erkutlu & Chafra, 2016
	$b = 0.26$, $b = 0.49$, $b = 0.24$	USA-Korea, USA-India, USA-Taiwan	Behavioral integrity perceptions	Friedman et al., 2018
Self-control	$b = 0.29 - 0.30$	USA	Moral awareness, cheating	Gino et al., 2011
	$b = 0.25 - 0.33$	Netherlands	Deviant leader behavior	Joosten et al., 2014
Moral awareness / identity	$b = -0.49$	USA	Cheating	Gino et al., 2011
	$b = -0.24$	Netherlands	Deviant leader behavior	Joosten et al., 2014
	$b = 0.27 - 0.49$	USA	Unethical behavior	Moore et al., 2012
Ethical leadership	$b = 0.33 - 0.41$	Netherlands	Top mgmt. team effectiveness, follower optimism	De Hoogh & Den Hartog, 2008
	$B = 0.12 - 0.27$, $B = 0.21 - 0.27$	Pakistan, Australia	Exposure to workplace bullying, interactional justice	Ahmad, 2018
Machiavellianism	$b = 0.18 - 0.27$	USA	Behavioral intentions	Jones & Kavanagh, 1996
	$b = 0.09$	USA	Propensity to engage in unethical behavior	Tang et al., 2008
Interpersonal and relationship-oriented skills				
Political skill	$b = 0.06 - 0.36$	USA	Relationship quality, leader & contextual perf.	Brouer et al., 2012
	$b = 0.10 - 0.21$	Germany	Leader effectiveness	Ewen et al., 2013
	$b = 0.58$, $r = 0.33 - 0.41$	USA	Employee perceptions of org. support, trust, org. commitment, cynicism	Treadway et al., 2004

Table 2. *(Continued)*

Attribute	Effect Size	Country	Performance Criteria	Source
Empathy	b = 0.42 - 0.54	Multicultural	Task leadership, relations leadership	Kellett et al., 2006
	b = 0.56	USA	Leader performance	Sadri et al., 2011
Social skill / competence	r = 0.34 - 0.46	USA	Initiating structure, group rated effectiveness, consideration	Riggio et al., 2003
	r = 0.17 - 0.20	USA	Peer ratings of performance	Zaccaro et al., 1995
	r = 0.25	USA	Adaptability	Zaccaro et al., 2009
Leading in challenging environments				
Problem solving	b = 0.16	USA	Continuance in the Army	Zaccaro et al., 2012
Sensemaking	b = 0.27 - 0.56	USA	Ethicality, recognizing circumstances, consideration	Caughron et al., 2011
	b = 0.35 - 0.60	USA	Ethical decisions, ethical culture	Zeni et al., 2016
Forecasting	b = 0.28 - 0.41,	USA	Decision ethicality	Stenmark et al., 2010
	b = 0.41 - 0.60	USA	Decision ethicality, forecast quality	Stenmark et al., 2011
Cog. complexity	b = 0.28	Netherlands, Italy	International orientation	Leone et al., 2005
Temperament				
Attributional complexity	b = 0.11 - 0.21	India, France	Subordinate satisfaction, performance	Lakshman & Estay, 2016
Conscientiousness	β = 0.26, β = 0.22	Singapore	Interactional adjustment, metacognitive CQ	Ang et al., 2006
	b = 0.24 - 0.30	Netherlands	Ethical leadership, role clarification	Kalshoven et al., 2011
	b = 0.24	Netherlands, Italy	International orientation	Leone et al., 2005
	r = 0.20	Singapore	Leader effectiveness	Ng et al., 2008
	b = 0.21	Multicultural	Task performance in intercultural assignments	Rockstuhl et al., 2015

Table 2. *(Continued)*

Attribute	Effect Size	Country	Performance Criteria	Source
	$b = 0.20 - 0.14$	Multicultural	Cultural flexibility, tolerance of ambiguity, ethnocentrism, global leadership success	Caligiuri & Tarique, 2012
	$b = 0.14 - 0.17*$	Multicultural	Cultural, interaction, & work adjustment, contextual & task performance	Shaffer et al., 2006
Extraversion	$\beta = 0.15 - 0.16$	Singapore	Motivational & behavioral CQ	Ang et al., 2006
	$b = 0.48$	USA	Strategic thinking competency	Dragoni et al., 2011
	$r = 0.19 - 0.35$	USA	Promotions, leader emergence, influencing	Foti & Hauenstein, 2007
	$r = 0.19$	Singapore	Leader effectiveness	Ng et al., 2008
	$b = 0.19$	Multicultural	Interpersonal OCB	Rockstuhl et al., 2015
	$b = 0.32* - 0.18$	Multicultural	Cultural flexibility, tolerance of ambiguity, ethnocentrism, global leadership success	Caliguiru & Tarique, 2012
	$b = 0.01 - 0.36$	Multicultural	Cultural, interaction and work adjustment, contextual & task performance	Shaffer et al., 2006
Openness	$\beta = 0.17 - 0.28$	Singapore	Motivational & behavioral CQ	Ang et al., 2006
	$b = 0.16$	Multicultural	Decision making	Ang et al., 2007
	$b = 0.08 - 0.38$	Multicultural	Cultural, interaction and work adjustment, contextual & task performance	Shaffer et al., 2006
	$b = -0.47 - 0.28$	Multicultural	Cultural flexibility, tolerance of ambiguity, ethnocentrism, global leadership success	Caligiuri & Tarique, 2012
Tolerance for ambiguity	$b = 0.38$	USA	Continuance in the Army	Zaccaro et al., 2012
	$b = 0.17$	Multicultural	Global leadership success	Caligiuri & Tarique, 2012

Note. *Coefficient reversed to facilitate comparison. All coefficients are $p < 0.05$ or $p < 0.01$. Coefficient ranges represent multi-faceted constructs and/or multiple performance criteria.

15

CHAPTER 3: SME INTERVIEWS

Purpose

The next phase of this research effort involved the interviewing of subject matter experts (SMEs) on the FA48 role with two purposes: first, to obtain information on the KSAs necessary for successful FA48 job performance and, second, to collect information on the dimensions of FA48 job performance. As FA48 SMEs have worked in the FA48 role and/or have witnessed incidents of both successful and unsuccessful FA48 performance, these interviewees were able to provide in-depth, expert knowledge through a guided interview process. The information and conclusions drawn from these interviews were used to corroborate and complement the literature review process.

Methodology

The U.S. Army Research Institute for the Behavioral and Social Sciences identified SMEs and contacted these potential interview candidates for participation. Candidates who were currently or previously Foreign Area Officers of COL rank or below were prioritized for recruitment and participation. The research team aimed to interview participants across multiple areas of concentration and positions within the FA48 branch to obtain information from a diverse and representative sample. The identified candidates were contacted via email, receiving a brief description of the project and how their participation could support it. They were advised that their participation was completely voluntary and that there would be no repercussions for declining to participate. Candidates who agreed to participate scheduled a 30-minute interview by phone or on the Microsoft Teams platform.

An ARI team member followed the structured interview protocol (see Appendix B), focusing exclusively on the KSAs required for successful FA48 job performance and the components of FA48 job performance. Additionally, SMEs were also asked to briefly describe their background and experience with the FA48 role. In all interviews, one or two members of the research team were present to take notes and capture the SME responses. These notes were reviewed, organized, and edited for organization and clarity, and provided the basis for the content coding as described below. No names or any other type of personal information were retained in the notes; any mentioned names or personal information were redacted in the review process. All interviewee responses have been kept confidential.

Sample Characteristics

A total of 17 interviews were conducted with officers who had between six and 19 years' experience in the FA48 role across a wide range of specific assignments. Of the 17 officers, eight were COLs and nine were LTCs. The officers represented a variety of AOCs, including three 48Js-Africa, south of the Sahara, 48Bs- Latin America, 48Gs – Middle East, and 48Es- Eurasia, two 48Is – Southeast Asia, and one 48H- Northeast Asia, 48F- China, and 48D- South Asia.

Content Coding

Using a confirmatory approach to validate each KSA that appeared across the interviews, the research team created codeable units in the form of KSA-related statements pulled from the interview notes. The resultant 43 KSA statements were then coded for importance (where possible) as indicated in the interview with a level of *Low*, *Medium*, or *High*. In most instances where KSA importance was captured, SME statements reflected high importance. In a handful of cases, SMEs mentioned low or medium importance regarding a particular KSA. An additional category of *Other* was applied to predictors that were not specific KSAs or that had not previously been identified from the literature review.

An exploratory approach was applied via a thematic analysis of the interview responses pertaining to FA48 job performance. Using the interview notes, we developed a set of codeable units consisting of statements that referred to a distinct performance topic. Multiple statements referring to the same performance topic were condensed into one codeable unit so that each unique performance theme could only be coded once per interview. These units were then coded for job type and AOC.

Findings

KSAs

For each KSA, several descriptive statistics were calculated, including frequency of appearance and frequency of each category of importance. Five KSAs were validated by 10 or more interviewees: cultural intelligence, political skill, written communication, oral communication, and language skill. Another five KSAs were validated by nine interviewees: self-regulation, behavioral flexibility, strategic thinking, adaptability, and work motivation. Several other KSAs reached five, six, or seven mentions across the interviews.

Many potential predictors identified in the *Other* category were mentioned frequently enough to warrant consideration in the assessment, including family support, prior international experience, and performance as an Army officer. Three additional predictors mentioned by participants--language skill, intelligence, and education--were not considered for inclusion in the assessment, as they are currently evaluated as part of the FA48 application process.

The results from the qualitative KSA analysis are presented in Table 3, including (1) SME frequency, which indicates the number of times the attribute was mentioned by an interviewee and (2) SME importance, which indicates the number of times an interviewee mentioned that the attribute was of high criticality for FA48 job performance. Frequency of mention and number of high ratings were evaluated in a holistic process to determine the most critical attributes for inclusion in the FA48 assessment battery. Very few instances of low or medium importance were coded; therefore, only frequency of high importance was considered and included in the table. Only three KSAs (cultural intelligence, political skill and written communication) received SME importance ratings of "high" and "medium." For each of these three KSAs, only one instance of "medium" importance was coded. All other codeable units regarding these KSAs were unable to be coded for importance. Frequency of mention for predictors/themes coded in the *Other* category are as follows: language skill (11), performance as

17

an Army officer (7), cultural distance (6), nature of Army assignments (6), prior international experience (5), humility (5), family compatibility / support (4), intelligence (3), and education (3).

Table 3. *Qualitative KSA Findings from SME Interviews*

Attribute	SME Frequency	SME Importance
Culture competence and intelligence		
Cultural perspective-taking	3	1
Cultural intelligence	12	4
Cultural open-mindedness	1	0
Cultural flexibility	5	2
Ethical conduct		
Integrity	6	4
Ethical decision-making	6	4
Self-regulation	9	5
Moral awareness	3	2
Interpersonal and relationship-oriented skills		
Political skill	12	7
Written communication	11	6
Oral communication	11	6
Conflict management	1	0
Negotiation	1	0
Behavioral flexibility	9	3
Social control	2	0
Active listening	6	4
Social expressiveness	2	1
Emotion regulation	2	2
Empathy	4	4
Social sensitivity	7	4
Leading in challenging environments		
Tolerance for ambiguity	7	2
Strategic thinking	9	4
Strategic communication	7	5
Adaptability	9	2
Cognitive flexibility	2	1
Problem solving	3	2
Cognitive complexity	6	4
Peer leadership	6	1

Table 3. *(Continued)*

Attribute	SME Frequency	SME Importance
Sense making	4	3
Temperament		
Stress tolerance	6	3
Self-awareness	7	3
Work motivation	9	6
Emotional stability	1	1
Conscientiousness	2	0
Patience / Long-term orientation	6	1
Curiosity	3	1
Learning goal orientation	5	1
Extraversion	4	0

Note. SME Frequency = frequency of mention in SME interviews ($n = 17$); SME Importance = number of high importance ratings coded in SME interviews.

Performance

The content-coding of performance-related statements revealed four high-level themes across 270 codeable units: performance indicators ($n = 141$), red flags ($n = 40$), performance evaluation ($n = 71$), and performance environment ($n = 18$). Performance indicators include statements related to observable, behavioral signifiers of performance; notably, these reflected both high and low indicators of a behavioral concept (i.e., inability to communicate well and ability to communicate effectively were coded as communication effectiveness). Performance evaluation statements include statements related to the assessment of FA48 job performance (e.g., how FA48s are evaluated, who evaluates them, etc.) Performance environment statements refer to descriptive indicators of the context within which FA48s work and how these affect performance requirements. Red flags are statements referring to behaviors or other such "troubling" indicators associated with unsuccessful FA48 performance (largely reflecting ethical or moral conduct). These statements consisted only of negative indicators, separate from positive performance indicators (i.e., the inverse of the negative indicator would not reflect high performance).

Each of the four high-level themes were coded with up to two additional levels of sub-themes. Within performance indicators, level one sub-themes were assignment experience, behavior, Officer Evaluation Report (OER), work products, effectiveness of programs, quality of advice, reputation, and other were coded. Notably, codeable units reflecting behavior fell into 16 additional level two sub-themes. Examples include communication effectiveness, interacting with others, ability to influence, and ability to think strategically.

Within performance evaluation, three level one sub-themes emerged: two regarding performance evaluators (reflecting positive sentiment and negative sentiment) and one consisting of general comments about the performance evaluation process. Findings regarding performance evaluation and which individuals are best suited to judge performance were mixed. Of note, eight

participants indicated that the senior rater is not in a good place to judge performance, while four participants suggested the opposite. Additionally, four participants indicated that peers are not well-suited to judge performance, while eight participants indicated the opposite. Three participants indicated that the secondary rater is well-suited to judge FAO performance, while four participants suggested the opposite. Further, several comments indicated that the formal performance evaluation process contains several sources of bias, including subjective evaluation criteria in various assignments/roles and lack of ability of senior raters and raters to actually observe FAO performance. For instance, co-location of secondary raters, as well as visibility/exposure to senior leadership, is often dependent on the region.

Within red flags, 32 codeable units were considered Behaviors (a level one subtheme) across three additional level two sub-themes (drug / alcohol use, lack of self-discipline, and ethical conduct issues). Of these, 21 codable units reflected examples of ethical conduct issues.

Finally, 18 codeable units reflected the nature of the performance environment. Of note, seven participants described the FAO role as an "interface between dissimilar cultures and organizations." As the units coded under this theme were largely uniform in content, no sub-themes emerged.

Table 4 includes the coded themes and subthemes with frequency counts considered in the development and/or selection of criterion measures. Due to sample size limitations, results were not analyzed by job type or AOC. As this information was captured in reference to coded statements, it is possible to reference particular coded material as necessary. This information was considered in developing criterion measures to ensure development of a performance rating form to encompass behavioral indicators applicable across FAO roles and regions.

Table 4. *Frequency of Coded Performance Themes and Sub-themes*

Level 1 Sub-theme	Level 2 Sub-theme	Frequency
Performance indicators		
Assignment experience		8
Behaviors	Ability to think strategically	3
	Achieving relevant objectives	6
	Adaptability	8
	Army professionalism	5
	Balances multiple authorities	4
	Communication effectiveness	11
	Decision-making	4
	Gives back to profession	4
	Host senior leader	4
	Influencing others	7
	Interpersonal interactions	12
	Language proficiency	4

Table 4. *(Continued)*

Level 1 Sub-theme	Level 2 Sub-theme	Frequency
	Remain calm under pressure	3
	Takes initiative	9
	Team leader	6
	Time management/deadlines	6
OER	Block check	4
	Consistent high evals	2
	Stratification	2
Other indicator	Invited to meetings	2
Program effectiveness		4
Quality of advice		4
Reputation		3
Work products	Oral presentations	4
	Written work	8
Red flags		
Behaviors	Drug/alcohol use	4
	Ethical conduct issue	21
	Lack of self-discipline	7
Family struggles		2
Lost confidence		4
Performance evaluation		
Evaluator (bad)	Ambassador	1
	Civilian	1
	Peers	4
	Rater	4
	Senior rater	8
	Subordinates	1
Evaluator (good)	360 evaluation	1
	Ambassador	1
	Chief of Mission	1
	Country Desk Officers	1
	Division Chiefs	2
	Intermediate rater in Embassy	1
	Peers	8
	Rater	3
	Senior DIA rep	1

Table 4. *(Continued)*

Level 1 Sub-theme	Level 2 Sub-theme	Frequency
	Senior rater	4
	Senior service reps	1
	Person from basic branch	1
	Subordinates	2
	Supervisor in country	1
General comments	Bias in formal rating system	10
	Criteria differ by job	8
Performance environment		
Higher military presence		1
Influence people you do not have authority over		1
Interface with dissimilar organizations and cultures		7
Strategic complexity		3
Work for multiple bosses		2
Work in small teams / groups		3
Work may not have much visibility		1

Note. Sub-themes with fewer than two mentions were only included for Performance Evaluation.

Summary

The purpose of the SME interviews was to provide further insight into the job performance requirements of the FA48 role and the KSAs necessary to meet those requirements. Both the KSA and performance requirement findings resulting from coding of the interviews provided useful information by which to evaluate the importance of the KSAs and performance requirements identified. The findings of the interviews were considered concurrently with findings from a KSA survey, detailed in the next chapter, to determine both the predictor and criterion measures for inclusion in the assessment battery.

CHAPTER 4: KSA SURVEY

Purpose

The next step in the research effort was the administration of a KSA survey to FA48 SMEs with the purpose of validating findings from the literature review and quantitatively determining the most critical KSAs for FA48 performance. Specifically, the survey served to determine which predictor attributes among those identified are the most critical for FA48 performance and, subsequently, should be included in an assessment battery. The 43 KSAs and their definitions (see Appendix C) were included in the survey were selected based upon several factors: (1) the frequency with which the attributes were mentioned across sources in the literature review, (2) the strength of attribute-criterion relationships gathered from empirical sources, (3) predictive validities in previous Army studies, especially those predicting officer performance, (4) predictive validities across cultures, and (5) strong conceptual evidence for linking a KSA to one or more performance requirements.

Methodology

The same pool of interview participants recruited for the structured SME interviews were also recruited for the KSA survey. The potential participants received an email with a link to the survey that directed them to an Army survey platform where they received brief instructions and were asked to make distinctions about the most important attributes for FA48 performance. Specifically, participants were asked to rate each KSA for importance ("Very Unimportant" to "Very Important") and ability to differentiate performance ("Not at all" to "Very Strongly") on a 1 to 5 scale. Participants were also asked to select and rank the top 10 KSAs most important for FAO performance; they were given an option to select "Cannot rate" in the case that they could not make an accurate determination.

Sample Characteristics

A total of 23 FAOs participated in the online KSA survey. Of those, 10 were LTCs and 13 were COLs from a variety of AOCs, including four 48Bs, 48Es and 48Js, three 48Fs and 48Gs, two 48Is and 48Ds, and one 48H. Previous assignment experience across the survey participants was rich and diverse, such that 22 FAOs indicated previous assignment overseas with a U.S. Country Team, 15 held Army Operational/Institutional assignments, and 16 held Political-Military assignments.

Descriptive Analysis

To determine the findings from the KSA survey, a cleaned, organized database consisting of ratings for each KSA was created from the participant responses. The "cannot rate" response option was re-coded to reflect missing data and was not used to calculate overall ratings. Average response across participants was calculated to create overall ratings of importance and ability to differentiate performance for each KSA. A sum of participant responses was calculated to indicate the number of times a KSA was indicated to be in the top 10 and, from those, an average ranking was created.

To determine the relative importance of each KSA for FA48 performance, a criticality score was created by summing the scores for importance and ability to differentiate performance. Scores for both factors were considered of conceptually equal importance; thus, a weighting method was not utilized to calculate criticality.

Findings

To determine the most important attributes for FA48 performance based on the survey data, the criticality score was examined in conjunction with scores for all four ratings included in the survey. The KSA data was ranked by criticality score. As both "importance" and "differentiates performance" were assessed on a 1 to 5 scale, the maximum criticality score for any KSA was 10. The data consistently showed that KSAs with the highest criticality scores also demonstrated high scores on the other rating factors. As such, a criticality score threshold of 8.5 was applied to determine the most critical KSAs for FA48 performance. A total of 14 KSAs exhibited criticality scores of 8.5 or above; see Table 5 for an overview.

Table 5. *KSAs with Criticality Scores of 8.5 or Higher*

KSA	Criticality Score	Importance Rating	Differentiates Performance	Top 10 Count	Avg. Top 10 Ranking
Political skill	9.16	4.48	4.68	13	7.77
Written communication	9.11	4.57	4.55	18	6.39
Tolerance for ambiguity	9.04	4.70	4.35	8	6.00
Oral communication	9.02	4.61	4.41	18	5.72
Cultural perspective-taking	9.00	4.74	4.26	5	5.40
Strategic thinking	8.91	4.61	4.30	15	4.73
Stress tolerance	8.87	4.52	4.35	7	5.71
Integrity	8.78	4.74	4.04	14	2.43
Ethical leadership	8.70	4.65	4.04	10	2.90
Strategic communication	8.70	4.43	4.26	7	5.29
Adaptability	8.57	4.43	4.14	10	5.30
Cultural intelligence	8.57	4.61	3.96	9	5.56
Cultural open-mindedness	8.52	4.57	3.96	3	7.33
Cognitive flexibility	8.52	4.30	4.22	6	7.33

Summary

Findings from the SME interviews and KSA survey, in addition to the literature review completed previously, were taken together to inform the attributes recommended for measurement in the assessment battery. In particular, findings from all three efforts were utilized to (1) finalize the attributes that are most important for FA48 performance, (2) finalize the attribute measures that should be included in an assessment battery, and (3) develop or select FA48 criterion measures to validate the assessment battery.

CHAPTER 5: MEASURE IDENTIFICATION AND DEVELOPMENT

Findings from the SME interviews and KSA survey were interpreted holistically and in conjunction to determine the KSAs and performance requirements, as well as the requisite predictor and criterion measures that warrant inclusion in an FA48 assessment battery for concurrent validation. Notably, because the criteria for inclusion of a particular KSA in the battery included the availability of a valid measure, the set of predictors to be included was not finalized until after acceptable measures were identified.

Predictor Measures

Attribute Selection

To determine the KSAs most critical for FA48 performance and finalize the list of attributes recommended for measurement, we evaluated quantitative and qualitative findings from the literature review, SME interviews, and the KSA survey holistically. To account for differences in relative importance across the data sources, a comprehensive, multi-step approach was utilized to determine the most important attributes for inclusion in the assessment battery. Specifically, three levels of inclusion criteria were created from both the qualitative and quantitative data to inform the selection of attributes for measurement.

Due to the objectivity and specificity of the assessment method, the KSA survey was considered the first criteria for selection of attributes to be measured. Using a criticality cutoff score of 8.5, a total of 14 KSAs were identified in step one: political skill, written communication, tolerance for ambiguity, oral communication, cultural perspective-taking, strategic thinking, stress tolerance, integrity, ethical leadership, strategic communication, adaptability, cultural intelligence, cultural open-mindedness, and cognitive flexibility.

A second set of criteria considered KSAs that fell just below the 8.5 criticality cutoff score but were frequently mentioned in the qualitative assessment. KSAs with criticality scores between 8 and 8.5 and that had a frequency of five or more mentions in the SME interviews and at least one high importance rating were identified. This set of criteria resulted in the identification of an additional six KSAs for consideration in step two: ethical decision-making, self-regulation, behavioral flexibility, cognitive complexity, self-awareness, and work motivation.

A third and final set of inclusion criteria considered KSAs that were mentioned frequently in the SME interviews (5+ times), despite a criticality score below 8.0, as well as those with criticality scores between 8 and 8.5, but lower frequency of mentions in the SME interviews (one to five mentions). Through this set of criteria, six additional KSAs were identified in step three: moral awareness, social sensitivity, active listening, peer leadership, patience / long-term orientation, and learning goal orientation.

After the application of these three criteria, a total of 26 KSAs were identified in the multi-step selection process. These KSAs were used to select attributes and accompanying measures for inclusion in the assessment battery. Specifically, the 26 attributes that emerged from the quantitative and qualitative data findings were evaluated in relation to the quality of

measures available to assess them, including validity evidence, format, fit for context, and availability. The process used to identify and evaluate these measures is described below.

Attribute Measure Identification and Selection

To identify the most appropriate measures for inclusion in an assessment battery to be validated for use in the FA48 application process, the initial literature review was expanded to determine the most valid and usable measures of attributes deemed critical to FA48 performance. Once the source materials from the original literature review were exhausted, the search was expanded to additional military, academic, and practitioner literature containing measures of the focal attributes.

To determine the usability of pre-existing measures, a database documenting the available measures for each critical attribute across the five categories was created to delineate various indicators of measure quality, including measure validity, reliability, intended use, and other characteristics. For each measure included in the database, a usability rating was provided on a scale of 1 to 3 (low to high). This rating reflected the suitability for inclusion of the measure in the assessment battery, in addition to considering whether the measure was publicly available and/or would need modification to be applicable to the FA48s. To make determinations about the suitability of including the measures in an FA48 assessment battery, several criteria regarding the properties of the measures were taken into consideration.

Specifically, information related to the criterion validity, content and construct validity, and ease of administration were considered in concert to assign usability ratings. A measure received a rating of 3 (high) if it demonstrated high predictive validity for relevant outcomes in military and/or multicultural settings, had satisfactory content and construct validity, and was deemed easy to administer. Notably, a few exceptions were made based on ease of administration and/or applicability of the measure despite limited criterion-related validity if the measure was developed within the military and/or if it was developed very recently and had qualities superior to other pre-existing measures. A measure received a rating of 2 (medium) if all criteria for high were met, but validity studies that specifically contained a cross-cultural component could not be identified; concerns emerged with construct or content validity (e.g., sparse information regarding measure development identified), but other measure properties were satisfactory, such as predictive validity and ease of administration (including relevance of items to FA48s); and if predictive validity evidence was sparse, but the measure was sound in all other areas (relevant, easy to administer, strong content/construct validity). Finally, a measure received a rating of 1 (low) if there were serious concerns with content or construct validity and/or no available predictive validity evidence in addition to other flaws, such as item relevance or strong correlations with impression management measures; it was unfeasible to administer (e.g., a group-based scenario); developed for an irrelevant population and/or could not be modified; and/or was pay-for-use and therefore unavailable.

The review of the literature resulted in a measure database that included approximately 150 measures to be considered for use in assessing the five categories of attributes deemed important for FA48 performance. Within each category, effort was made to identify multiple measures of varying types (where possible) for each attribute. The usability ratings were used to determine the quality of available measures for each of the attributes deemed critical for FA48

performance. The measures prioritized for use were available to the public, developed and/or utilized within the military, and did not require substantial modification. For many of the attributes, pre-existing measures internal to ARI were the most feasible and appropriate, as ARI has detailed information regarding their effectiveness for predicting Army outcomes (e.g., stress tolerance and tolerance for ambiguity).

Attribute Measure Recommendations

A set of existing measures was identified for use in the assessment battery that provided coverage of most critical attributes (i.e., 20 of the 26 KSAs that met inclusion criteria). Additionally, measures were identified for some attributes that did not meet inclusion criteria, but were chosen for inclusion based upon their previous prediction of officer performance, including self-efficacy and extraversion.

Following further review and discussion among the research team, several additional attributes with accompanying ARI measures were identified for inclusion given their promise for predicting performance in the FA48 context (i.e., based on predictive validity in other Functional Areas). These additional attributes included figural reasoning, divergent thinking, and multitasking. The cognitive ability (i.e., figural reasoning) and divergent thinking measures, which assess thinking styles relevant to FA48 duties such as creative and critical thinking, have been used and validated repeatedly in selection research.

Table 6 summarizes how the attributes, within each attribute category, are covered in the assessment battery across all measures selected for inclusion. The measures were determined to both (1) provide the most efficient coverage across the attributes and (2) meet the accessibility criteria described above. The included attribute measures fell into three types--ethical, biodata, and cognitive measures--that are described below and subsequently used to categorize the results of the concurrent validation detailed in the next chapter.

Table 6. *Conceptual Coverage of Attributes in Selected Measures*

KSA	Biodata Measures													Cognitive Measures		Ethical Measures	
	CA	WC	OC	BF	TA	RT	CF	PL	E	SE	AO	ST	MT	FR	CT	MIS	MSED
Culture competence and intelligence																	
Cultural perspective-taking	X																
Cultural intelligence	X																
Cultural open-mindedness	X																
Ethical Conduct																	
Integrity																X	
Ethical decision-Making																	X
Self-regulation																X	
Moral awareness																X	X
Interpersonal and relationship-oriented skills																	
Written communication		X															
Oral communication			X														
Behavioral flexibility				X													
Political skill	X	X	X														
Leading in challenging environments																	
Tolerance for ambiguity					X												

28

Table 6. *(Continued)*

KSA	CA	WC	OC	BF	TA	RT	CF	PL	E	SE	AO	ST	MT	FR	CT	MIS	MSED
						Biodata Measures								**Cognitive Measures**		**Ethical Measures**	
Strategic thinking						X											
Strategic communication			X			X											
Adaptability					X							X					
Cognitive flexibility							X										
Cognitive complexity						X											
Peer leadership								X									
Temperament																	
Extraversion									X								
Self-efficacy										X							
Work motivation											X						
Stress tolerance												X					
Multi-tasking													X				
Cognitive*																	
Figural reasoning														X			
Divergent thinking															X		

Note. Biodata subscales: CA = cultural adroitness, WC = written communication, OC = oral communication, BF = behavioral flexibility, TA = tolerance for ambiguity, RT = reflective thinking, CF = cognitive flexibility, PL = peer leadership, E = extraversion, SE = self-efficacy, AO = achievement orientation, ST = stress tolerance, MT = multitasking. Cognitive measures: FR = Figural Reasoning scale. CT = Consequences Test. Ethical measures: MIS = Moral Identity Scale, MSED = Military-Specific Ethical Dilemmas scale. *The cognitive attributes and requisite measures were added to the battery following additional consideration of the anticipated importance of these attributes for predicting FAO performance as described in the paragraphs below.

Biodata Measures. Existing biodata scales developed by ARI were included to assess many of the critical attributes for FA48 performance across attribute categories. Many of these measures have demonstrated predictive validity for a multitude of officer outcomes and are currently utilized for the selection of officers through multiple commissioning sources.[1] These scales have been described at length based upon their use in other, published ARI studies and technical reports; as such, we refer the reader to Zaccaro and colleagues' work (2012, 2015) for further information on the biodata measures used in this study.

Ethical Measures. Three measures were included in the assessment battery to capture attributes related to ethical conduct: the Military-Specific Ethical Dilemmas (MSED) measure, the Multidimensional Ethics Scale (MES), and the Moral Identity Scale (MIS). The full ethical measures are listed in Appendix F.

Military-Specific Ethical Dilemmas (MSED). The MSED comprises two scenarios that were developed from military commanders' detailed accounts of moral dilemmas that they confronted while deployed in high-intensity operations. One dilemma involves disciplining a good friend and subordinate for repeated risky decision-making for which participants are presented the choice to either privately reprimand the subordinate or pursue a full court martial. The second dilemma involves the decision to afford civilian refugees under duress access and safe haven in a military camp; participants choose to either let them in or turn them away.

For each scenario, participants select their preferred course of action and respond to three items about their choice. First, they rate their moral awareness (i.e., the extent to which their decision involved ethical and morality considerations) by responding to the item "Overall, to what extent does this decision involve ethical and morality considerations?" using a 5-point Likert-type rating scale ranging from 'not at all' to 'great extent'. Next, they evaluate the magnitude of the consequences using the item "The possible harm resulting from this option would be…" and a 5-point Likert-type rating scale ranging from 'very minor' to 'very severe'. Finally, respondents determine the social consensus associated with their choice of action by completing the prompt "Most people would consider this option to be…" using a 5-point Likert-type rating scale ranging from 'very inappropriate' to 'very appropriate'.

Morality. The Multidimensional Ethics Scale (MES; Reidenbach & Robin, 1990) is used to assess morality with a 5-point, eight-item semantic-differential scale ("How well do the following characteristics describe this option: Just, Fair, Morally Right, Acceptable to my family, Culturally acceptable, Traditionally acceptable, Violates an unspoken promise, and Violates an unwritten contract"). The content validity of this measure was considered adequate. Predictive validity has been demonstrated in military and civilian samples against aspects of morality, including moral awareness, moral judgment, and moral intent (Blais & Thompson, 2013, Krosch et al., 2012, Thompson et al., 2006, Thompson et al., 2018).

Moral Identity Scale (MIS). The MIS was included to assess attributes related to ethical conduct including moral awareness, integrity, and self-regulation. Specifically, the MIS measures one's self-conception organized around a set of moral traits described as a self-regulatory mechanism. The measure demonstrates adequate content and construct validity, as

---

[1] Contact ARI for additional information regarding biodata scales.

well as predictive validity in military and non-military samples for outcomes including leader behavior, ethical leadership, and ethical behavior. Cross-cultural research indicates that the measure is valid for measuring moral traits as valued in Western culture (Aquino & Reed, 2002, Mayer et al., 2012, Olsen et al., 2006, Reed & Aquino, 2003, Shao et al., 2008, Vitell et al., 2016).

Cognitive Measures. Additional measures were selected for inclusion to assess cognitive ability and creative thinking, as these attributes have been shown to predict officer performance in a variety of settings and assess the type of thinking styles that are relevant to FA48 duties.

Figural Reasoning. ARI provided a timed cognitive ability assessment to capture critical thinking. This cognitive assessment comprised an overall measure of figural reasoning, as well as experimental measures of associated components, including figural classification, figural series, figural analogy, and pattern matrix. The experimental measures were included for evaluation of psychometric properties and validity for predicting FA48 performance.

Consequences Test. ARI provided a measure of creative thinking that presents respondents with a series of novel scenarios (e.g., people no longer need or want sleep) and asks them to record all the possible consequences that they can imagine. This assessment is being evaluated for several Army talent management initiatives, including as a part of the Captains Career Course, the Army War College, and select Functional Areas.

According to Table 6, the measures described above provide explicit coverage of all attributes that meet the step 1 inclusion criteria with the exceptions of strategic communication and ethical leadership. While no usable, standalone measure of strategic communication was identified, an individual scoring high in both strategic thinking and communication may be considered proficient in strategic communication. Additionally, an appropriate and usable measure of ethical leadership was unable to be identified in the literature; as such conceptual overlap and frequency of mentions in the SME interviews support inclusion of ethical decision-making and self-regulation to capture KSAs related to ethical conduct. Further, the MSED and MIS measures assess another component of ethical conduct (i.e., moral awareness) resulting in comprehensive coverage of the ethical conduct KSA category.

Additional Predictor Items. To supplement the selected attribute measures described above, new biodata scales were developed by the research team to assess additional predictors that emerged from the literature review and SME interviews as indicative of success in the FA48 role, including prior international experience, family support, and self-development initiative. All items developed were reviewed by multiple members of the research team for accuracy and clarity and revised accordingly. The response scales were designed to be as specific as possible as to leave little room for differences in interpretation across participants. While performance as a regular Army officer was mentioned by SMEs as indicative of success, it was also considered a performance indicator and is described in the Criterion Measures section below. As such, the following newly created biodata scales were included to capture predictors of FA48 job performance in the assessment battery: (1) prior international experience, (2) family support, and (3) self-development initiative. All items developed for these scales are presented in Appendix D.

Prior International Experience. Both the codeable units from the interview notes as well as the academic literature on cross-cultural performance were reviewed to generate items regarding international experience. In particular, SMEs indicated that prior cultural experiences, such as living in another country, likely contribute to FAO success. Further, the academic literature suggests that individually-initiated cultural experiences, including non-work experiences like vacationing in foreign countries or studying abroad, contribute to accelerated professional development for expatriates (Caligiuri & Tarique, 2012). Similarly, prior international work assignments have been shown to predict cross-cultural adjustment for expatriates (Chen et al., 2010). Given these findings, four biodata items were developed to assess aspects of prior international experience. SMEs also indicated that the pursuit of voluntary cultural learning experiences (e.g., learning multiple languages) was a common characteristic among high performers; thus, a biodata item was developed to capture the number of languages spoken and maintained.

Family Support. The concept of family support and compatibility also emerged from the SME interviews as an important predictor of FAO success. Participants mentioned that families often struggle to adapt to unfamiliar cultural settings, citing unrealistic expectations of living conditions. As family members are often the primary support for FAOs located overseas, such struggles can be a large contributor to stress. As such, SMEs suggested that support from family and family compatibility with particular locations and regions are key to successful FAO performance. To capture this concept, four biodata items were developed around family support, adaptability, and willingness to relocate to different regions of the world.

Self-Development Initiative. Finally, themes regarding taking initiative and giving back to the profession emerged from the SME interviews as attributes associated with high performing FA48's. Codeable units reflecting these themes included developing the office over and above assigned tasks and engaging in self-directed learning to deepen regional expertise. Two biodata items were developed to capture these concepts.

Criterion Measures

Performance Rating Forms

Results from the qualitative assessment of FA48 performance were used to inform the selection and development of criterion measures in the form of performance rating forms. Based upon the frequency counts for sub-themes within the Performance Evaluation section and the mixed sentiment among participants with regard to who can best judge FAO performance, a 360-degree approach to performance evaluation was preferred to incorporate multiple perspectives and overcome the limitations associated with individual sources of ratings (Campion et al., 2019). As such, performance rating forms were developed for self, peers, and supervisors.

Performance Indicators. The codeable units extracted from SME interviews that reflected FA48 performance requirements were used to develop behavioral statements representing the coded FA48 performance indicators. Additionally, cross-cultural military studies were reviewed to ensure that all relevant aspects of military performance in a multicultural environment were captured in the behavioral statements (e.g., Abbe & Gallus, 2012; Hardison et al., 2009). Many of these studies pertained to junior or enlisted officers on

32

deployment and included more general cross-cultural performance elements (e.g., basic cultural skills) that were not considered or included in the battery as, by the time they enter the FA48 role, officers already possess advanced cultural expertise. Due to the multicultural nature of the FAO role, all performance dimensions encapsulate behavior in a cross-cultural context; therefore, cultural expertise is interwoven throughout the behavioral statements across performance dimensions.

Positively-worded behavioral indicators were developed to be context-specific and reflective of performance in and across FA48 roles. Multiple research team members generated and evaluated statements in an iterative process focused on producing clear, succinct, and accurate statements, and removing any redundant or poorly worded items. Though 16 themes were identified as behaviors within the coded Performance Indicators, conceptually similar behavior themes were identified and further consolidated, resulting in eight higher-level performance dimensions. Each dimension consisted of two to six behavioral indicators, ensuring adequate coverage of the criterion space.

Past attempts to develop criterion measures in similar contexts have suffered from issues of restriction of range, as SMEs have been unwilling to rate participants lower than the highest one or two performance categories. To increase the variance in performance ratings for the concurrent validation effort, a 5-level proficiency scale was developed where 1 indicated "Below Average" and 5 indicated "Truly Exceptional." Descriptions were included for each proficiency anchor to help raters differentiate performance on each dimension (e.g., "Truly Exceptional" was described as "In the top 4% of FAOs you know at this skill"). Finally, a ninth item asked for a general rating of overall performance compared to peers of the same rank. The full performance rating form and associated proficiency scale can be found in Appendix E.

Biodata Criterion Measures

Biodata items were identified or developed to assess additional indicators of performance that were identified in the qualitative analysis, including aspects of ethical conduct. Pre-existing biodata items concerning disciplinary incidents and indicators of high performance (e.g., awards received, ratings received on previous OERs) were included in both self-ratings of performance as well as peer and supervisor performance ratings. Performance as a regular Army officer was frequently mentioned by SMEs as indicative of FAO success; thus, several self-report biodata items (previously developed by ARI) were included to capture OER ratings.

Using the codeable units from SME interviews reflecting the performance requirements of the FA48 role, additional biodata items were developed to represent the Red Flags coded in the qualitative assessment (see Table 4). Ethical conduct items were developed for self-evaluations to assess aspects of ethical decision-making, including potential ethics violations or instances of questionable judgment. Items were also developed to assess ethical conduct from peer and supervisor perspectives and included perceived moral character, ethical decision-making, and instances of known ethics violations or disregard for regulations (e.g., misuse of funding). Further, biodata items were developed to capture other specific performance indicators that emerged from the qualitative analysis but were not captured in the rating form, including use of FAO written work products by senior leadership (e.g., inclusion in strategy documents), nature of FAO assignments (including assignment location and selection for special

assignments), and reputation amongst peers. These items were administered in the self, peer, and supervisor evaluations. Finally, biodata items developed to capture themes of *taking initiative* and *self-directed learning* were also administered in the peer and supervisor evaluations as potential indicators of high performance. The process for item generation was the same as that described above for the developed predictor measures. All biodata criterion measures are presented in Appendix D. The final set of criterion measures for self, peer, and supervisor evaluations are detailed below.

FA48 Performance Rating Scale. The performance rating scale served as the primary criterion measure, assessing eight dimensions of FA48 performance with two to six behavioral indicators per dimension. Ratings were collected from FA48s (self), peers, and supervisors using a 1-5 proficiency scale where 1 indicated "Below Average" and 5 indicated "Truly Exceptional*."* For each performance dimension, an optional comment box was provided where the rater could provide additional information about their rating. This strategy was intended to elicit more honest ratings and alleviate potential concerns about providing low ratings (i.e., resulting in greater response variability).

Biodata and Archival Measures. Newlydeveloped biodata items regarding ethical conduct, assignment experience, work products, and reputation amongst peers were administered in self, peer, and supervisor evaluations. Additional biodata items regarding self-development initiative were administered in peer and supervisor evaluations; notably, these same items were included in the FA48 assessment battery as predictor measures. Finally, self-evaluations included a series of archival performance items regarding ratings received on past OERs.

Summary

The results of SME interviews and the KSA survey served to finalize the attribute measures that warrant inclusion in an assessment battery for FA48 selection. Table 6 depicts the attributes by category and the measures with high usability ratings that were identified to capture those attributes. Additionally, the inclusion of cognitive measures (e.g., Figural Reasoning scale) will provide insight into the applicability of these related attributes, which have previously been shown to predict officer performance, in the FA48 context. The SME interview results also guided identification and development of additional predictor items and criterion measures of FA48 performance to be included in the assessment battery. The finalized assessment battery was subsequently administered to current FA48s for concurrent validation to evaluate the psychometric properties of the measures. Supervisor and peer ratings of performance were collected for FA48s who completed the survey to provide preliminary validation of the assessment.

CHAPTER 6: CONCURRENT VALIDATION OF ASSESSMENT BATTERY

Concurrent Validation Analysis Overview

Purpose

To evaluate the validity of the developed assessment battery for predicting FA48 officer performance, a concurrent validation was conducted. The primary objective of this effort was to empirically validate a standardized selection test consisting of measures of attributes that predict FA48 job performance and ultimately to help select FA48s. Assessment measures administered to active-duty FA48 officers were linked to current job performance criteria provided by the FA48 officer, their rater, and their peers to inform the best combination of assessment scores.

Methodology

The Office of Primary Responsibility for FA48 (Strategic Leadership Division, G-3/5/7) helped identify and contact active duty FA48 officers with at least one year of experience and from a range of Areas of Concentration (AOC). The assessment battery containing predictor measures and self-rated performance criteria was administered to current FA48s online using the survey platform Verint, which must be accessed on a CAC-enabled computer. Potential participants were invited to participate via an email that included random ID codes and passwords for the participants to use to access Verint. The ID code and password fields in Verint were restricted such that only numerals could be entered, and the 7-digit limit was short enough to prevent participants from entering most PII (e.g., social security numbers).

The assessment instrument was divided into three sections. Sections one and two measured intellectual or cognitive skills associated with critical thinking (i.e., figural reasoning measures) and creative thinking (i.e., Consequences Test). Section three measured traits and temperaments (i.e., biodata and ethical measures), and asked respondents to report on their own job performance, awards, test scores. The specific measures included in each portion of the assessment are detailed in the previous chapter. FA48 respondents were also asked to provide demographic information (e.g., gender, race/ethnicity, rank, AOC) as well as contact information for their supervisor rater and up to five peers whom they consider knowledgeable about their job performance. The supervisors and peers listed by the participants were then contacted by email and asked to provide unofficial, off-the-record ratings of incumbents' performance to serve as criterion measures for estimating the validity of the measures in the assessment battery. The measures included in the performance rating forms are also detailed in the previous chapter.

Sample Characteristics. The target population of the data collection for the concurrent validation was incumbent active duty FA48 officers at the rank of COL or below who had at least one year of experience in a FA48 role, as well as their supervisor raters and up to five peers who could speak to their job performance. Over 800 FA48 officers were invited to voluntarily participate in this research effort as well as to refer up to 400 supervisor raters and up to 600 peer raters. Predictor data and self-rated criterion data was collected from a total of 170 FA48s, though the total number of responses received for each measure varied from 62 to 170. The online assessment was kept open for more than six weeks to obtain as many responses as

possible. Multiple reminder emails were sent throughout the data collection period to encourage participation.

Of the FA48 respondents who indicated gender, 90% were male ($n = 102$) and 10% were female ($n = 11$). Of the officers who indicated race/ethnicity, 73.5% identified as white ($n = 83$), 8.8% selected Hispanic or Latino, 6.2% selected Asian, 6.2% selected more than one race, 4.4% selected Black or African American, and 1% selected American Indian or Alaska Native. Given the small sample sizes for many of the race/ethnicity categories, a minority status variable was created where white was considered non-minority ($n = 83$), and all other groups were considered minority ($n = 30$). The FA48 respondents' current ranks ranged from O4 to O6, with the majority of officers ranked O5 (48%, $n = 55$). The respondents were commissioned in a variety of branches with the majority commissioned in Infantry (19%, $n = 33$), followed by Armor (16%, $n = 21$) and Military Intelligence Corps (14%, $n = 18$). The respondents were also working in a variety of AOCs, the majority FA48Es (31%, $n = 35$), followed by FA48Bs (20%, $n = 23$). When asked about their current assignment, 33% ($n = 38$) of respondents indicated Overseas U.S. Country Team, 32% ($n = 36$) indicated Political-Military, 26% ($n = 30$) indicated Army Operational/Institutional, and 11% ($n = 10$) officers selected Other. The majority of officers had been in their current assignments for 1 to 5 years (38%, $n = 65$) or less than a year (27.5%, $n = 47$). When asked about tenure as an FA48 officer, 44% of respondents indicated 5 to 10 years ($n = 50$), 40% indicated 10 or more years ($n = 46$), and 16% indicated 1 to 5 years ($n = 18$).

Additionally, criterion data from supervisors and peers was collected from 161 participants (32 supervisors and 129 peers) who rated a total of 98 of the participating FA48 officers. 195 performance ratings were received in total (20% from supervisors and 80% from peers). However, three ratings were dropped due to the raters' indication that they did not observe enough of the FA48's performance to rate any aspects, bringing the total number of usable officer ratings down to 192 and the number of individual officers rated down to 97. Similar to the assessment battery, the data collection period for performance ratings was extended to maximize response rates. Multiple reminder emails were sent to peer and supervisor raters over a two-month period to boost the number or responses.

The ranks of the peer and supervisor raters ranged from O2s to O8s, with 30 raters reporting their rank as Other (e.g., civilian). The majority of supervisor raters were O6s (50%), followed by Other (25%), and O5s (14%). A total of 71% of supervisor raters described their opportunity to observe the FA48 officer's performance as "enough to judge all aspects." When asked about their relationship with the rated officer, the majority of supervisor raters indicated that they were previously (63%) or currently (23%) the officer's rater. For peer raters, the majority were O5s (46%) and O4s (24.5%). Most peer raters indicated having observed the FA48 enough to judge all (45%) or most (35%) aspects of the officer's performance, though many indicated having observed the officer enough to judge only some aspects of performance (20%). When asked about their relationship with the officer being rated, 70% of peer raters indicated having previously worked with him/her and 22% indicated currently working with him/her.

Data Merging and Cleaning

Upon launching the assessment battery, it was discovered that several of the FA48s invited to participate were unable to access the assessment via Verint. For this reason, the

assessment was also uploaded to a different, easier-to-access Verint system. Upon completion of data collection, a comprehensive data cleaning process was completed to ensure consistency in scoring and reporting for each survey item across the two Verint systems. Detailed data codebooks were created for both systems and utilized to streamline response scoring and data labeling to merge the predictor data from the two sources. Once a complete dataset was constructed for the predictor data, response options were re-coded as needed (e.g., item responses were reverse-coded such that high scores reflected high levels of the focal construct, "Cannot Say" responses were re-coded as missing). Individuals who completed less than half of the assessment battery were removed from further analysis. Response frequencies were examined for each item to identify any out-of-range responses. Examination of response patterns across items revealed one case where the same response option was selected 150 consecutive times; this case was removed from the dataset.

Peer and supervisor performance ratings were collected via PDF forms; thus, rating data was extracted from the PDFs into two separate data files. Within each data file, names of the officers rated were thoroughly reviewed to correct spelling inconsistencies. This was necessary to match officer names with their unique ID codes and subsequently merge rating data with assessment scores. Unique ID codes were also created for the supervisor and peer raters to fully de-identify the data files. Before merging the peer and supervisor rating data into one dataset, a variable was created to indicate rating source (i.e., peer or supervisor). Response options were then re-coded as needed (i.e., item responses were reverse-coded such that high scores reflected high levels of the focal variable, "Cannot Rate" responses were recoded as missing). Three ratings were removed from the dataset where the rater indicated that he / she did not observe the FA48 enough to rate any aspects of their performance.

As several of the officers who participated in the assessment received multiple ratings of their performance, rating data was aggregated to reflect average ratings for each performance measure, a process described in more detail in later sections. A new variable was created to assign a rating number to different ratings of the same officer. The maximum number of performance ratings received for an individual officer was five. Performance rating data was then aggregated in SPSS by the rated officer's ID number. Manual checks against the raw data were performed to ensure there were no errors in this process. Comments provided in the optional comment box were not retained in the final dataset, as they were not included in the analyses. The aggregated performance measures were then merged with the predictor assessment scores using the officer's unique ID codes. Again, manual checks were performed to ensure that data merged correctly.

Psychometric Properties of Predictor and Criterion Measures

This section describes the basic psychometric properties of the predictor scales (i.e., biodata, cognitive, and ethical measures), as well as the criteria used for validity analyses described later in this report. For all predictor variables, we examined (a) distributional properties (e.g., mean, standard deviation), (b) reliability estimates, (c) correlations among the predictor scales, and (4) correlations with criterion measures.

Psychometric Properties of Predictor Scales

Biodata Measures. Table 7 presents the distributional properties of the biodata measures. Among the previously established ARI biodata scales, score variance and internal consistency reliability estimates were acceptable; means ranged from 3.41 to 4.19 on a 5-point scale. The highest mean (4.19) was found for cultural adroitness, while the social desirability measure exhibited the lowest mean (3.40). Standard deviations ranged from 0.46 to 0.83, with extraversion demonstrating the highest variance. All pre-established measures reached an acceptable internal consistency reliability of $\alpha = 0.75$ or higher when poorly performing items were dropped. The majority of measures demonstrated internal consistency reliabilities of 0.80 or higher. Coefficient alpha is reported for the final set of items retained for each biodata scale, which were used to calculate scale scores.

Of the newly developed biodata predictor measures, only the family support scale reached acceptable internal consistency reliability; the coefficient alphas for international experience and self-development initiative were below the acceptable threshold to justify creating scale scores. It is possible that the items developed to assess these predictors are capturing related, yet distinct, constructs. Given these findings, only the full family support measure was retained for further analysis. The remaining newly created biodata scales were examined at the item level to identify individual items that warranted inclusion in future analyses. Items that exhibited little to no variance in responses were excluded, while items with adequate variance in responses were retained as potential predictors of performance. Specifically, three items regarding prior international experience, subsequently referred to as international travel, international experiences, and fluent languages, and one item regarding self-development initiative, subsequently referred to as takes initiative, were retained. See Appendix D for the full list of items.

To reduce overall assessment length and burden on survey respondents going forward, all biodata scales were further refined by identifying the best performing items within each scale. Items with item-total correlations below 0.40, and in some cases 0.50, were removed to achieve the smallest set of items that demonstrated acceptable reliability without significantly reducing coefficient alpha. Measures were reduced to 10 items or fewer wherever possible. Items were removed from 12 scales in total. Only written communication (7 items) and extroversion (5 items) scales were unable to be reduced any further.

Table 7. *Descriptive Statistics and Internal Consistency Reliability Estimates for Biodata Scales*

Predictor	k	*n*	M	SD	Skew	Coef α
Existing Biodata Scales						
Achievement orientation	10	130	3.98	0.65	-0.59	0.88
Behavioral flexibility	10	149	4.05	0.48	-0.41	0.84
Cognitive flexibility	9	147	3.88	0.60	-0.20	0.84
Cultural adroitness	11	147	4.19	0.46	-0.53	0.76
Extroversion	5	141	3.78	0.83	-0.41	0.86

Table 7. *(Continued)*

Predictor	k	*n*	M	SD	Skew	Coef α
Multi-tasking	11	128	3.45	0.73	-0.17	0.90
Oral communication	5	147	3.99	0.61	-0.29	0.75
Peer leadership	8	139	3.74	0.62	-0.02	0.89
Reflective thinking	10	137	4.15	0.52	-0.23	0.85
Self-efficacy	6	139	3.83	0.77	-1.06	0.85
Stress tolerance	10	135	3.55	0.62	-0.34	0.83
Tolerance for ambiguity	8	129	3.89	0.57	-0.19	0.82
Written communication	7	129	3.76	0.63	-0.30	0.75
Social desirability	6	139	3.41	0.55	-0.56	0.78
Newly Created Biodata Scales						
Self-development initiative	2	125	-	-	-	0.52
Prior international experience	4	149	-	-	-	0.43
Family support	4	145	4.22	0.76	-1.22	0.79
Takes initiative*	1	125	3.81	0.93	-0.22	-
International travel*	1	149	4.02	1.06	-0.73	-
International experience*	1	150	3.18	1.35	-0.09	-
Fluent languages*	1	149	2.94	0.86	0.5	-

Note. *Individual items retained for analysis. k = number of items; M = mean; SD = standard deviation; Coef α = coefficient alpha; Scale scores were not calculated for measures that did reach acceptable coefficient alpha.

Cognitive Measures. Table 8 presents the distributional properties of the scales used to measure cognitive ability and creative thinking. Score variance for the primary figural reasoning measure was somewhat limited. The mean score was 24.77 with a maximum possible score of 30, and the standard deviation was 3.40. This lack of variance is unsurprising given the high-ranking nature of the officers assessed. This finding suggests evidence of restricted range in cognitive ability and related concepts, which may influence validity for predicting FA48 performance. The experimental figural reasoning scales demonstrated greater variance, with means ranging from 3.90 to 5.00 with a maximum possible score of 10. Regarding divergent thinking, the Consequences Test mean was fairly high at 8.10 with a standard deviation of 2.30. On average, participants provided eight valid responses to each of the five prompts. Again, this finding is unsurprising, as officers who have reached higher ranks in the Army are more likely to score high on these types of cognitive skills. The correlation between divergent thinking and figural reasoning was not significant. Intercorrelations among all cognitive measures can be found in Appendix G.

Table 8. *Descriptive Statistics and Internal Consistency Reliability Estimates for Cognitive Measures*

Predictor	k	*n*	M	SD	Skew	Coef α
Divergent thinking	5	125	8.06	2.30	0.52	0.82
Figural reasoning scale	30	163	24.77	3.45	-1.07	0.79
Experimental figural reasoning scales						
Pattern matrix (Form A and B)	10	126	5.01	1.73	0.04	0.35, 0.18
Figural classification (Form A and B)	10	106	3.89	2.09	0.16	0.18, 0.69
Figural series (Form A and B)	10	116	4.76	2.04	-0.06	0.52, 0.70
Figural analogy (Form A and B)	10	113	4.08	1.72	-0.14	0.38, 0.28

Note. k = number of items; M = mean; SD = standard deviation; Coef α = coefficient alpha

Ethical Measures. Table 9 presents the distributional properties of the ethical measures. Score variance and internal consistency reliability estimates were acceptable. Coefficient alphas were above 0.80 for moral identity and the MES for both ethical decision-making scenarios. Moral identity demonstrated a mean of 3.80 on a five-point scale and a standard deviation of 0.68. Responses to each of the ethical decision-making scenarios were examined in detail to identify any potential response patterns across the two scenarios. Regarding preferred course of action for the subordinate dilemma, 34% of respondents opted to reprimand the subordinate in private, intended to involve more serious/harmful consequences, while 66% opted to relieve the subordinate of command and have him court-martialed. Regarding the refugee dilemma, 25% of respondents opted to let the refugees into camp, intended to involve more serious/harmful consequences, while 75% opted to turn the refugees away. These results indicate that more individuals were accepting of privately reprimanding a subordinate than were accepting of letting refugees into camp.

Table 9. *Descriptive Statistics and Internal Consistency Reliability Estimates for Ethical Measures*

Predictor	k	*n*	M	SD	Skew	Coef α
Moral identity	13	62	3.76	0.68	-0.70	0.90
Subordinate dilemma						
MES	8	66	4.06	0.62	-0.21	0.81
MC	1	133	3.38	0.97	-0.07	-
SC	1	133	3.80	1.21	-0.85	-
MA	1	133	3.8	1.10	-0.64	-
Refugee dilemma						
MES	8	67	3.38	1.00	-0.14	0.88
MC	1	133	4.08	0.82	-0.56	-
SC	1	133	3.24	1.23	-0.22	-
MA	1	133	4.08	1.17	-1.15	-

Note. K = number of items; M = mean; SD = standard deviation; Coef α = coefficient alpha; MC = magnitude of consequences; SC = social consensus; MA = moral awareness; MES = Multidimensional Ethics Scale.

A cross-tabulation was conducted to further explore response patterns across the two dilemmas (*n* = 133). The results of this analysis were not significant, though sample size may be a factor. Only 42% of individuals who chose to let the refugees into camp also chose to reprimand the subordinate in private. There was greater overlap in the other response options, as 69% of individuals who turned the refugees away also chose to have the subordinate court martialed. These findings, presented in Table 10, suggest that the ethical scenarios are different enough to elicit different decision outcomes. The descriptive statistics for the ethical dilemma measures show that the MES mean for the refugee dilemma was lower than the subordinate dilemma (*M* = 3.38 and 4.06, respectively). Means for moral awareness and magnitude of consequences were higher for the refugee dilemma than the subordinate dilemma, whereas the opposite was true regarding social consensus. These findings indicate that respondents perceived greater moral and ethical considerations, as well as greater consequences, associated with the refugee dilemma. A correlation matrix including both sets of responses to the ethical dilemmas showed significant positive relationships between morality measures across dilemmas. Specifically, the MES and moral awareness measures for both dilemmas were significantly positively correlated, as were the magnitude of consequences and social consensus measures, though less strongly. This pattern of correlations indicates that participants consistently perceive ethical / moral considerations associated with their course of action for both scenarios. Intercorrelations among the various ethical measures can be found in Appendix G.

Table 10. *Ethical Dilemmas Descriptives by Course of Action*

Course of Action	MC		SC		MA		MES	
	M	SD	M	SD	M	SD	M	SD
Subordinate dilemma								
Private reprimand (*n* = 45)	3.33	0.95	3.02	0.94	3.18	1.11	3.86	0.63
Court martial (*n* = 88)	4.07	1.09	3.56	0.95	4.10	1.14	4.15	0.61
Refugee dilemma								
Let them in (*n* = 33)	4.79	0.65	3.94	0.83	3.64	1.03	4.17	0.61
Turn them away (*n* = 100)	3.84	1.21	4.12	0.82	3.11	1.26	3.05	0.95

Note. M = Mean, SD = standard deviation, MC = magnitude of consequences; SC = social consensus; MA = moral awareness, MES = Multidimensional Ethics Scale.

Correlations Among Predictors

To assess potential redundancy among the predictor measures, we examined the correlations among all individual scales (see Appendix G). Correlations between the predictors and self-rated performance criteria were also examined and are described in detail later in this report. The magnitude of correlations among predictor measures were considered, in part, as criteria for inclusion in validity analyses.

Correlations among Biodata Measures. We examined intercorrelations among all biodata scales for evidence of convergent and discriminant validity (see Appendix G). Many biodata scales were moderately to highly correlated, with the strongest relationship found between oral communication and peer leadership ($r = 0.66$). Additionally, strong relationships were found between stress tolerance, behavioral flexibility and tolerance for ambiguity ($r = 0.52$ - 0.56). Achievement orientation was highly correlated (i.e., above 0.50) with measures of peer leadership, oral communication, tolerance for ambiguity, and reflective thinking ($r = 0.52$ - 0.55). Cultural adroitness demonstrated strong correlations with behavioral flexibility and extraversion ($r = 0.50$ and 0.55, respectively). The magnitude of these correlations suggests a high degree of conceptual and statistical overlap between many of the biodata scales.

Additionally, consistent with previous ARI research efforts (e.g., Zaccaro et al., 2012, 2015), we examined the relationships between biodata scales and a measure of socially desirable responding. Several biodata measures exhibited moderately positive correlations with the social desirability measure. Of note, stress tolerance demonstrated the strongest relationship with a correlation above 0.50. To account for socially desirable responding, we applied a statistical adjustment to the biodata measures that were significantly correlated with the social desirability scale. Scores were adjusted until they exhibited as close to a 0 correlation as possible. The adjusted scales included behavioral flexibility, cultural adroitness, reflective thinking, self-efficacy, and tolerance for ambiguity. The adjusted biodata scales were included in validity analyses detailed in later sections. Intercorrelations among these scales are also presented in Appendix G.

Biodata Correlations with Cognitive and Ethical Measures. The biodata measures exhibited minimal correlations with cognitive measures, showing no evidence of redundancy in measurement. The strongest correlations were found between achievement orientation and divergent thinking ($r = 0.24$) and between peer leadership and figural reasoning ($r = 0.21$). A few sporadic correlations in the 0.20 range were significant between biodata measures (e.g., tolerance for ambiguity, reflective thinking, and cognitive flexibility) and various cognitive measures.

The biodata measures exhibited a pattern of meaningful relationships with ethical measures, with little to no evidence to suggest redundancy in measurement. Several biodata measures were moderately correlated with moral identity. The strongest relationships were found between moral identity and peer leadership and achievement orientation. Cognitive flexibility was consistently correlated with responses to the MES for both ethical dilemmas ($r = 0.28$ and 0.47). Interestingly, a larger number of significant correlations were found between biodata measures and responses to the subordinate ethical dilemma (i.e., moral awareness and the MES). In general, biodata measures exhibited significant correlations with ratings of moral awareness and the MES, rather than course of action, magnitude of consequences, or social consensus. Correlations between several biodata measures and moral awareness for the subordinate dilemma ranged from 0.18 to 0.25. Cultural adroitness was significantly correlated with the MES for the subordinate dilemmas, but not the refugee dilemma. The same was true for tolerance for ambiguity, written communication, and reflective thinking. Consistent with the response analysis for the ethical dilemmas, these patterns of correlations indicate potential differences in ethical decision-making processes and outcomes across the two scenarios. The two scenarios may elicit very different moral and ethical considerations for this particular group.

Subgroup Analyses

Independent samples t-tests were conducted to check for significant mean differences in all predictor measures by minority status. Individuals who identified as white were considered non-minority, while all other groups were considered minority for the purpose of this analysis, as the sample sizes for specific race / ethnicity groups were too small to analyze separately.

Significant mean differences between subgroups were found for only four measures: peer leadership, tolerance for ambiguity, reflective thinking, and figural reasoning (original measure). In all four instances, mean scores were higher for individuals who identified as white. The largest mean difference found between groups among the biodata scales was 0.28 for peer leadership. The mean difference for the original 30-question figural reasoning measure was 1.85. It should be noted that the sample size for the non-minority group was much larger in comparison to the minority group. If equal variances are not assumed, only peer leadership and figural reasoning show significant differences. Subgroup analyses could not be conducted on the basis of gender, as the sample size for females was extremely low.

Psychometric Properties of Criterion Measures

Self-Rated Criteria. Table 11 presents the distributional properties of the self-rated performance criteria. Score variance and internal consistency reliability estimates were acceptable, with some exceptions. The newly developed biodata criterion measures (i.e., ethical conduct, assignment experience, work efforts) did not achieve acceptable internal consistency

reliability. Regarding the ethical conduct biodata measure, little to no variance was found in item-level responses. The vast majority of respondents reported no participation in any sort of unethical behavior. The assignment experience and quality of work efforts scales were examined at the item level to determine which items, if any, warranted inclusion in future analyses. Items that exhibited little to no variance were excluded, while items with adequate variance in responses were retained as potential performance indicators. Specifically, one quality of work efforts item, subsequently abbreviated as work efforts, and two assignment experience items, assignment location and special assignments, were retained.

Regarding the eight-dimension performance rating scale, adequate variance was found for self-ratings ($M = 3.70$, $SD = 0.67$), though most participants rated themselves as "Above Average" or "Well Above Average" on many of the dimensions. Internal consistency reliability was high, indicating the eight performance dimensions were highly correlated.

Table 11. *Descriptive Statistics and Internal Consistency Reliability Estimates for Self-Rated Criteria*

Criterion Measure	k	*n*	M	SD	Skew	Coef α
Performance rating scale	8	124	3.72	0.67	-0.45	0.88
Performance compared to peers	1	127	3.21	0.60	-0.19	-
Past four OERs	1	127	2.95	1.00	-0.71	-
Ethical conduct	3	128	-	-	-	0.44
Assignment experience	3	150	-	-	-	0.46
Quality work efforts	2	110	-	-	-	0.39
Assignment location*	1	147	3.88	1.17	-0.69	-
Special assignments*	1	143	2.76	1.37	0.17	-
Work efforts*	1	116	3.26	1.10	-0.45	-

Note. *Individual items retained for analysis. k = number of items; M = mean; SD = standard deviation; Coef α = coefficient alpha; Scale scores were not calculated for measures that did not reach acceptable coefficient alpha.

An exploratory factor analysis (EFA) was conducted to further evaluate the psychometric properties of the performance rating scale (see Table 12). The performance dimensions were factor analyzed using principal axis factoring with oblimin rotation. The EFA revealed a single factor with an Eigenvalue of 4.37 that accounted for 54.7% of the variance. Eigenvalues for each of the remaining factors were less than one. The scree plot clearly indicated that the first factor accounted for most of the total variability in the data, while the remaining factors accounted for a very small portion of the variability. Factor loadings for each of the dimensions ranged from 0.54 to 0.78. The communalities of the dimensions were relatively high, indicating that the individual dimensions of FA48 job performance are highly related to one another. The single item assessing overall performance compared to peers was treated separately from the eight-dimension performance scale, as the response scale for that item was slightly different. The mean for this item was 3.20 on a 4-point scale (SD = 0.60). Finally, little variance was found for archival items regarding past performance ratings on OERs. Only one item (*"Indicate how many times you*

received each rating on your past 4 OERs. "Above Center Mass" or "Most Qualified")
demonstrated enough variance to warrant inclusion in future analyses ($M = 2.95$, $SD = 1.00$).

Table 12. *Exploratory Factor Analysis of the Self-Rated Performance Scale*

Performance Dimension	Factor Loading	Communality
Thinks strategically	0.54	0.30
Influences stakeholders	0.75	0.56
Adapts to unfamiliar cultural settings	0.74	0.54
Manages stress effectively	0.61	0.37
Demonstrates personal discipline and ethical behavior	0.64	0.41
Communicates effectively	0.78	0.61
Works effectively with Army soldiers	0.73	0.54
Interacts effectively with foreign partners	0.75	0.56

Other-Rated Criteria. Table 13 presents the distributional properties of the other-rated performance criteria. Overall, the means were high (4+ on a five-point scale). The ethical conduct items demonstrated almost no variance, as did the assignment experience items. Only the performance rating scale and self-development initiative biodata measure achieved adequate internal consistency reliability.

Table 13. *Descriptive Statistics and Internal Consistency Reliability Estimates for Other-Rated Criteria*

Criteria	k	*n*	M	SD	Skew	Coef α
Performance rating scale	8	190	4.15	0.70	-0.65	0.92
Performance compared to peers	1	190	4.14	0.87	-0.71	N/A
Unethical conduct*	3	188	1.07	0.30	6.49	0.65
Quality of work	2	177	4.30	0.70	-1.01	0.46
Self-development	2	181	4.43	0.72	-1.08	0.72
Assignment experience	3	170	4.00	0.77	-0.65	0.66

Note. *Criterion reverse coded. k = number of items; M = mean; SD = standard deviation; Coef α = coefficient alpha.

Other-Rater Agreement

For the purposes of assessing rater agreement, supervisor and peer ratings were combined into an aggregated category of other-rater performance measures due to sample size limitations. For officers who received multiple ratings of their performance from supervisor(s) and/or peer(s), inter-rater agreement was calculated using Awg to justify using an average rating (Brown & Hauenstein, 2005). Agreement was high at both the item and scale level, with a few exceptions. Instances where Awg did not meet the acceptable threshold of 0.60 or above were

further investigated to identify any discrepancies between raters of the same officer. Notably, Awg cannot be calculated when the standardized variance is 0 (i.e., the mean rating equals the highest or lowest possible rating). In most instances where Awg was below 0.60, it was for an individual item, rather than an entire scale. See Table 14 for the inter-rater agreement results.

Respondents exhibiting inter-rater agreement below the 0.60 threshold on any item were flagged for additional analysis. Seven different rated officers were flagged for rater disagreement on one or more items. Each of these FA48 officers were rated by three or more individuals in total. Raw ratings on the flagged items for these officers were further examined to determine the source of rater disagreement. In most cases, one rater was a clear outlier from the rest; as such, these outlier ratings were dropped so that averages could be calculated based on the majority rating wherever possible. For items with no majority rating, we chose to drop the item from the overall scale (i.e., it was treated as missing).

Table 14. *Other-Rated Criteria Inter-Rater Agreement*

Performance Item / Scale	Mean Awg	Awg Range
Thinks strategically	0.82	0.53 - 1.00
Influences stakeholders	0.83	0.60 - 1.00
Adapts to culture	0.80	0.53 - 1.00
Manages stress	0.81	0.57 - 1.00
Personal discipline	0.78	0.53 - 1.00
Communicates effectively	0.83	0.53 - 1.00
Works effectively	0.80	0.53 - 1.00
Interacts effectively	0.81	0.60 - 1.00
Performance rating scale (items 1 to 8)	0.81	0.56 - 0.96
Overall performance compared to peers	0.80	0.53 - 1.00
Unethical conduct	0.71	0.54 - 0.86
Quality of work products	0.80	0.53 - 1.00
Self-development initiative	0.77	0.53 - 1.00
Assignment experience	0.84	0.65 - 1.00

Aggregated Rating Data

Following the assessment of the rater agreement, we then aggregated the data by Officer ID to create average scores for officers with multiple other-ratings. A total of 58 out of 97 total officers had received two or more ratings. The descriptive statistics on the aggregated item and scale scores are presented in Table 15. Notably, all means are high with the exception of unethical behavior. The self-development initiative measure also showed acceptable internal consistency reliability.

Table 15. *Descriptive Statistics and Internal Consistency Reliability Estimates for Other-Rated Criteria Aggregated*

Criteria	k	*n*	M	SD	Skew	Coef α
Performance rating scale	8	97	4.20	0.55	-0.53	0.90
Performance compared to peers	1	97	4.20	0.66	-0.45	N/A
Unethical conduct	3	96	1.10	0.33	7.44	0.78
Quality of work	2	78	4.30	0.52	-0.85	0.35
Self-development initiative	2	95	4.50	0.56	-0.87	0.77
Assignment experience	3	90	3.90	0.71	-0.78	0.58
Work efforts*	1	78	4.20	0.67	-0.83	-
High performer*	1	91	4.40	0.64	-1.25	-

Note. *Individual items retained for analysis. k = number of items; M = mean; SD = standard deviation; Coef α = coefficient alpha.

The internal consistency reliability for the performance rating scale was high. An exploratory factor analysis (EFA) was conducted to further evaluate the psychometric properties of the performance rating scale using aggregated data (see Table 16). The eight performance dimensions were factor analyzed using principal axis factoring with oblimin rotation. Findings were similar to those obtained for the self-rated version. The EFA revealed a single factor with an Eigenvalue of 4.77 that accounted for 59.50% of the variance. Eigenvalues for each of the remaining factors were less than one. The scree plot clearly indicated that the first factor accounted for most of the total variability in the data, while the remaining factors accounted for a very small portion of the variability. Factor loadings for each of the dimensions on the primary factor ranged from 0.57 to 0.86, with the majority above 0.70. The communalities of the dimensions were relatively high, again indicating that the individual dimensions of FA48 job performance are highly related to one another.

Table 16. *Exploratory Factor Analysis of the Other-Rated Performance Scale*

Performance Dimension	Factor Loading	Communality
Thinks strategically	0.74	0.55
Influences stakeholders	0.86	0.74
Adapts to unfamiliar cultural settings	0.76	0.58
Manages stress effectively	0.72	0.51
Demonstrates personal discipline and ethical behavior	0.57	0.33
Communicates effectively	0.81	0.66
Works effectively with Army soldiers	0.70	0.50
Interacts effectively with foreign partners	0.68	0.47

The ethical conduct biodata items demonstrated an acceptable coefficient alpha. However, these items showed little to no variance at the scale or item level, despite the design of these items to consider how best to elicit a wider range of responses regarding ethical conduct. That is, most raters indicated witnessing no unethical behavior. Similarly, regarding the nature of assignments biodata items, most raters selected the "Cannot say" response option.

Overall, at both the item and scale level, little variance exists in the other-rated performance data, suggesting strong evidence for range restriction in the criterion measures. The participants demonstrated a strong tendency to rate everyone highly, even though several preemptive actions were taken to mitigate this response trend, such as changing the wording and scale anchors from previous performance measures used at ARI. Specifically, 70% of individuals were rated higher than a 4 out of 5 on the performance rating scale. To further assess this issue, we separated supervisor ratings from peer ratings to determine whether there was a marked difference in rater perspectives, but no differences were found. Variance was minimal on each of the individual dimensions as well. Item level descriptives are presented in Appendix H.

In summary, only a few measures were retained for any further analysis, including the eight-dimension performance rating scale, the self-development initiative measure, and two items regarding quality of work efforts, subsequently referred to as work efforts and high performer. These measures exhibited slightly higher variance at the item level than the other measures, enabling us to retain them for predictive analyses.

Correlations Among Criterion Measures

To examine any potential redundancy among the criterion measures, we examined the correlations among all measures retained for analysis between self-rated and other-rated criteria. The magnitude of correlations among performance measures were considered, in part, as criteria for inclusion in validity analyses, alongside the distributional properties and internal consistency reliability.

Self-Rated Criteria. Table 17 presents correlations among the self-rated criterion variables retained for analysis. The majority of variables were significantly related to one another with correlations ranging from $r = 0.21$ between self-ratings on the performance rating scale and past OERs to 0.50 between self-ratings on the performance rating scale and overall performance compared to peers.

The largest number of significant correlations was found between self-ratings on the performance rating scale and the remaining criteria. Two items assessing assignment experience (i.e., special assignments and assignment location) were moderately correlated at 0.34, suggesting these facets are related, yet distinct concepts. The correlation between self-ratings on the performance rating scale and overall performance compared to peers indicates evidence of convergent validity, as both measures are intended to assess overall job performance. The small, yet significant, correlation between self-ratings on the performance rating scale and past OERs suggests that OERs may not completely capture the unique aspects of performance in the FA48 role. Interestingly, the work efforts item exhibited no meaningful relationships with any other

performance criteria. It may be the case that this item is not applicable to all types of FA48 roles or that the phenomenon assessed is not entirely within the FA48's control. None of the relationships found are strong enough to indicate redundancy in measurement.

Table 17. *Intercorrelations among Self-Rated Criterion Measures*

Criteria	PRS	PCP	OERs	Work efforts	Asgt loc	Spec asgts
PRS	1.00					
PCP	**0.50**	1.00				
OERs	**0.21**	**0.29**	1.00			
Work efforts	0.13	0.09	0.02	1.00		
Asgt loc	0.10	0.07	0.09	0.15	1.00	
Spec asgts	**0.22**	0.09	-0.06	0.04	**0.34**	1.00

Note. Bolded values are significant at $p < .05$. PRS = performance rating scale, PCP = performance compared to peers, OERs = past four OERs. Sample size for the criteria ranged from 109 to 125.

Other-Rated Criteria. Table 18 presents correlations among the other-rated criterion variables retained for analysis. Almost all variables were significantly related to one another with correlations ranging from $r = 0.37$ between other-rated work efforts and self-development initiative to 0.86 between other ratings on the performance rating scale and overall performance compared to peers. The only nonsignificant correlation found among other-rated performance criteria was between work efforts and reputation as a high performer. The strongest correlations were found between other ratings on the performance rating scale and the remaining criteria. The strength of the relationships among criteria provides some evidence of convergent and discriminant validity. As expected, the performance rating scale (intended to be a comprehensive measure of job performance across dimensions) exhibited the strongest correlations with other performance criteria. Performance compared to peers was highly correlated ($r = 0.69$) with reputation among peers as a high performer. Unlike the self-rated work efforts item, the other-rated work efforts item was moderately correlated with the remaining criteria, with the exception of reputation as a high performer. Overall, raters tended to evaluate FA48 performance consistently across measures/dimensions.

Table 18. *Intercorrelations among Other-Rated Criterion Measures*

Criteria	PRS	PCP	SDI	Work efforts	High performer
PRS	1.00				
PCP	**0.86**	1.00			
SDI	**0.51**	**0.46**	1.00		
Work efforts	**0.49**	**0.40**	**0.37**	1.00	
High performer	**0.63**	**0.69**	**0.47**	0.21	1.00

Note. Bolded values are significant at $p < .05$. PRS = performance rating scale, PCP = performance compared to peers, SDI = self-development initiative. Sample size for the criteria ranged from 77 to 97.

Self- and Other-Rated Performance Criteria. Table 19 presents correlations among the self-rated and other-rated criterion variables retained for analysis. A pattern of small to moderate correlations was found between self-rated and other-rated criteria. Significant positive correlations were found between measures retained from both rater perspectives (i.e., self and supervisor / peer). These measures included the eight-dimension performance rating scale and overall performance compared to peers. For the performance rating scale, self- and other-evaluations on those same eight dimensions were correlated at 0.22. A similar relationship was found for self- and other-evaluations of overall performance compared to peers ($r = 0.24$). The magnitude of these correlations is lower than what would typically be expected.

Examining the descriptive statistics for self and other-ratings on these two criteria revealed that peers/supervisors rated officers higher, on average, than they rated themselves. Among the remaining performance criteria retained, self-ratings of performance compared to peers and past OERs were positively correlated with other ratings of the officer's reputation among peers as a high performer, indicated by high performer in Table 19 below.

Table 19. *Intercorrelations among Self- and Other-Rated Criterion Measures*

Criteria	Other PRS	Other PCP	Other SDI	Other work efforts	Other high performer
Self PRS	**.22**	**.24**	.13	-.10	.21
Self PCP	**.25**	**.28**	.08	.10	**.29**
Past OERs	**.27**	**.27**	.03	.15	**.21**
Self work efforts	-.04	-.03	-.09	-.03	-.03
Self asgt loc	-.02	-.07	-.03	.06	.00
Self spec asgts	-.02	-.07	.13	.00	-.04

Note. Bolded values are significant at $p < .05$. PRS = performance rating scale, PCP = performance compared to peers, SDI = self-development initiative, Asgt loc = assignment location, Spec asgts = special assignments. Sample size for the criteria ranged from 67 to 94.

Criterion Subgroup Differences

Independent samples T-tests were conducted to check for significant mean differences in criterion measures (i.e., self-rated and other-rated) by minority status. Individuals who identified as white were considered non-minority, while all other groups were considered minority. No significant mean differences were detected for any of the self-rated criteria. Regarding other-rated performance criteria, one significant mean difference was found for the performance rating scale. The mean performance rating for the minority group was higher than that of the non-minority group ($M = 4.40$ and 4.10, respectively). No other significant differences were found between groups on the criterion variables.

Summary

The goal of this section was to describe the psychometric properties of the predictor measures and criterion variables of interest (i.e., self-rated and other-rated). Overall, the predictor measures were functioning as expected. The three sets of predictor measures showed

meaningful patterns of relationships, providing evidence of convergent and discriminant validity across measure types. The pattern of moderate to high correlations found among many of the biodata measures was not completely unexpected, suggesting a high degree of conceptual overlap among biodata constructs (e.g., behavioral flexibility, tolerance for ambiguity, and stress tolerance). These correlations were further evaluated, in conjunction with the results of psychometric analyses and prior work efforts (i.e., SME interview, KSA survey) to narrow down the set of predictor variables for inclusion in validity analyses.

The self-rated performance criterion variables looked acceptable, although some of the measures were restricted and positively skewed. The ethical conduct measure exhibited very little variance and was thus excluded from further analysis. Several self-rated criterion measures showed promise for validating the predictor measures (i.e., self-ratings on the performance rating scale, overall performance compared to peers, past OERs, and items assessing assignment experience and quality of work efforts). Examination of the other-rated criterion measures yielded less positive results. Most notably, severe restriction of range was evident across all of the other-rated performance measures. Descriptive analyses indicated that supervisors and peers tended to rate all FA48s highly within and across performance measures. Interestingly, mean ratings were higher for other-rater evaluations of performance than self-evaluations on the same measures. A subset of other-rater measures was retained for further analysis, including the performance rating scale, overall performance compared to peers, self-development initiative, and items assessing work efforts and reputation among peers as a high performer.

Bivariate Criterion-Related Validity Evidence

This section presents initial criterion-related validity evidence for each set of predictor measures, including the ability of these measures to predict both self-rated and other-rated performance criteria. For self-rated performance criteria, we describe validity results for biodata measures, cognitive measures (i.e., critical thinking and divergent thinking), and ethical measures (i.e., moral identity and ethical decision making). Results are based upon uncorrected correlations between the predictor measures and criteria.

Biodata, Cognitive, and Ethical Measures as Predictors of Self-Rated Performance Criteria

Biodata Measures. We examined correlations among biodata measures and self-rated performance criteria to establish initial validity evidence for predicting FA48 performance. Results of these correlations are presented in Table 20. The biodata measures showed statistically significant levels of validity for predicting most self-rated performance criteria via significant and positive correlations with one or more of the self-rated performance measures. The strongest correlations were found between the eight-dimension performance rating scale and biodata measures of oral communication, stress tolerance, achievement orientation, and cultural adroitness ($r = 0.35 - 0.47$). extroversion, behavioral flexibility, multitasking, written communication, reflective thinking, and tolerance for ambiguity were also correlated with self-rated performance ($r = 0.20 - 0.27$). A small, yet significant, positive correlation was found between social desirability and the performance rating scale ($r = 0.20$). The single item assessing overall performance compared to peers of the same rank exhibited the strongest correlations with peer leadership, stress tolerance, cultural adroitness and written communication ($r = 0.32 - 0.42$).

Achievement orientation and oral communication were also significantly correlated with this item (i.e., $r = 0.30$).

Several biodata scales correlated significantly with the single item assessing work efforts (i.e., frequency with which an individual's work efforts are included in strategy documents such as campaign plans or technical plans). The strongest correlation was found with written communication ($r = 0.37$), though oral communication and peer leadership were also significantly correlated ($r = 0.20 - 0.22$). The only significant correlations found with past ratings on OERs were for peer leadership ($r = 0.22$) and self-efficacy ($r = 0.22$). Two single biodata criterion items regarding assignment experience were retained for analyses and demonstrated a few significant correlations with biodata predictor measures. The assignment location item (i.e., "How frequently have you willfully taken assignments in dangerous or austere locations?") was positively correlated with achievement orientation, multitasking, tolerance for ambiguity, and cultural adroitness ($r = 0.19 - 0.36$). The special assignment item (i.e., "How often have you been pulled for special assignments based on particular regional knowledge or experience?") exhibited the strongest correlation with achievement orientation ($r = 0.26$).

Among the newly created biodata predictor items retained for analysis, minimal correlations were found with self-rated performance criteria. The item assessing initiative (i.e., "To what extent do you take the initiative to develop the office and/or multinational programs over and above assigned tasks?") was positively correlated with a few of the self-rated performance criteria. Specifically, correlations with the performance rating sale, overall performance item, and assignment location item ranged from 0.23 to 0.27. The international experience items did not demonstrate any notable relationships with self-rated performance criteria.

Table 20. *Correlations among Biodata and Self-Rated Criteria*

Criteria	PRS	PCP	OERs	Work efforts	Asgt loc	Spec asgts
BF	**0.26**	**0.23**	0.03	0.00	0.14	-0.02
CF	0.06	**0.22**	0.05	0.11	**0.19**	0.08
CA	**0.35**	**0.33**	-0.04	-0.03	0.06	0.04
OC	**0.47**	**0.31**	0.14	**0.20**	0.14	**0.19**
PL	**0.39**	**0.42**	**0.22**	**0.22**	0.14	**0.18**
Ext	**0.27**	0.16	-0.05	-0.03	0.09	**0.21**
RT	**0.20**	**0.22**	0.06	0.06	0.02	0.05
SE	0.10	**0.22**	**0.22**	-0.04	-0.09	-0.11
ST	**0.39**	**0.39**	0.15	0.02	0.12	0.12
AO	**0.37**	**0.31**	0.17	0.08	**0.36**	**0.26**
MT	**0.24**	0.16	0.13	0.08	**0.26**	0.05
WC	**0.22**	**0.32**	-0.06	**0.37**	0.07	0.15
TA	**0.20**	0.17	0.15	0.15	**0.23**	0.04
FS	0.07	0.04	0.03	0.06	-0.03	0.12
TI	**0.23**	**0.27**	0.10	0.15	**0.23**	0.13
IT	0.05	0.07	0.02	0.14	0.15	0.13
IE	0.05	0.08	-0.07	0.04	0.02	0.14
FL	0.14	0.00	-0.07	-0.09	-0.06	**0.18**
SD	**0.20**	0.15	0.08	-0.17	0.06	**0.19**

Note. Bolded values are significant at *p* < .05. PRS = performance rating scale, PCP = performance compared to peers, Asgt loc = assignment location, Spec asgts = special assignments. BF = behavioral flexibility, CF = cognitive flexibility, CA = cultural adroitness, OC = oral communication, PL = peer leadership, Ext = extroversion, RT = reflective thinking, SE = self-efficacy, ST = stress tolerance, AO = achievement orientation, MT = multi-tasking, WC = written communication, TA = tolerance for ambiguity, FS = family support, TI = takes initiative, IT = international travel, IE = international experiences, FL = fluent languages, SD = social desirability. Sample size for the criteria ranged from 111 to 146.

In examining the adjusted biodata scales to account for socially desirable responding, some correlations with self-rated performance measures were reduced in magnitude, though the majority remained significant. Correlations between adjusted biodata scales and self-rated performance measures can be found in Table 21.

Table 21. *Correlations among Adjusted Biodata and Self-Rated Criteria*

Criteria	PRS	PCP	OERs	Work efforts	Asgt loc	Spec asgts
aBF	**0.21**	**0.19**	0.01	0.06	0.07	-0.13
aCA	**0.31**	**0.30**	-0.07	0.01	0.02	0.02
aRT	0.16	**0.19**	0.05	0.10	0.01	0.00
aSE	0.05	**0.18**	**0.20**	0.00	-0.12	-0.17
aST	**0.34**	**0.33**	0.12	0.11	0.08	0.01
aTA	0.16	0.15	0.14	**0.20**	**0.22**	0.00

Note. Bolded values are significant at $p < .05$. PRS = performance rating scale, PCP = performance compared to peers, Asgt loc = assignment location, Spec asgts = special assignments. aBF = adjusted behavioral flexibility, aCA = adjusted cultural adroitness, aRT = adjusted reflective thinking, aSE = adjusted self-efficacy, aST = adjusted stress tolerance, aTA = adjusted tolerance for ambiguity. Sample size for the criteria ranged from 114 to136.

   Cognitive Measures with Self-Rated Criteria. We examined correlations among cognitive measures (i.e., critical thinking and divergent thinking) and self-rated criteria to establish initial validity evidence for predicting FA48 performance (see Table 22). The cognitive measures showed limited validity evidence for predicting self-rated performance, as the majority of correlations among cognitive measures and various self-rated criteria were minimal or nonsignificant. A few significant and near-significant positive correlations were found between figural reasoning measures and past ratings on OERs. Notably, significant negative relationships were found between two figural reasoning measures- the original figural reasoning test and figural series- and the performance rating scale (r = -0.22 and -0.23, respectively). While this finding is interesting, it is not completely unexpected. One potential explanation for these findings is that individuals high in cognitive ability may rate their own performance more modestly, consistent with previous research (Dunning, 2011). Additionally, descriptive analyses suggested evidence of range restriction in the cognitive measures that may account for this finding. No significant correlations were found among divergent thinking and any of the self-rated performance criteria. Again, the lack of significant relationships with performance outcomes may be explained by range restriction on the divergent thinking measure.

Table 22. *Correlations among Cognitive Measures and Self-Rated Criteria*

Criteria	SPR	PCP	OERs	Work efforts	Asgt loc	Spec asgts
DT	0.05	0.02	0.09	-0.01	-0.02	-0.04
FR-O	**-0.22**	-0.06	0.17	0.06	-0.08	-0.16
FR-PM	-0.11	-0.02	-0.03	-0.16	-0.03	0.05
FR-FC	-0.06	0.00	0.11	-0.07	**0.20**	0.00
FR-FS	**-0.23**	-0.07	**0.29**	0.05	-0.08	-0.11
FR-FA	-0.03	0.08	0.15	0.14	0.02	-0.03

Note. Bolded values are significant at $p < .05$. DT = divergent thinking, FR-O = figural reasoning original test, FR-PM = figural reasoning pattern matrix, FR-FC = figural reasoning figural classification, FR-FS = figural reasoning figural series, FR-FA = figural reasoning figural analogy; PRS = performance rating scale, PCP = performance compared to peers, OERs = past four OERs, Asgt loc = assignment location, Spec asgt = special assignments. Sample size for the criteria ranged from 89 to 141.

Ethical Measures with Self-Rated Criteria. We examined correlations among ethical measures (i.e., moral identity and ethical decision making via two ethical dilemmas) and self-rated criteria to establish initial validity evidence for predicting FA48 performance (see Table 23). The various ethical measures showed some statistically significant validity evidence for predicting self-rated performance outcomes. The moral identity measure was correlated 0.22 with the performance rating scale, though this correlation did not achieve significance at $p < .05$, likely due to low sample size ($n = 62$) for moral identity. For the subordinate ethical dilemma, both moral awareness and magnitude of consequences ratings were moderately correlated with the work efforts item ($r = 0.20$ and 0.21, respectively).

Table 23. *Correlations among Moral Identity, Subordinate Dilemma, and Self-Rated Criteria*

Criteria	PRS	PCP	OERs	Work efforts	Asgt loc	Spec asgts
Moral identity	0.22	0.07	0.03	0.10	0.08	0.07
Sub dilemma	0.00	0.13	-0.06	0.11	0.01	-0.01
MA (SD)	0.09	0.15	0.02	**0.20**	-0.09	-0.09
MC (SD)	0.11	0.04	-0.04	**0.21**	0.04	0.11
SC (SD)	0.02	0.06	**-0.18**	0.07	0.08	0.05
MES (SD)	0.03	0.10	-0.06	0.20	-0.05	-0.05

Note. Bolded values are significant at $p < .05$. SD = subordinate dilemma, MA = moral awareness, MC = magnitude of consequences, SC = social consensus, MES = multidimensional ethics scale; PRS = performance rating scale, PCP = performance compared to peers, OERs = past four OERs, Asgt loc = assignment location, Spec asgt = special assignments. Sample size for the criteria ranged from 53 to 130.

Additionally, an interesting pattern of correlations was observed among responses to the refugee dilemma and self-rated performance criteria (see Table 24). Specifically, moral awareness was positively correlated with past ratings on OERs ($r = 0.25$) and negatively correlated with assignment location (i.e., taking assignments in dangerous or austere locations, *r*

= -0.19). This indicates that those who scored higher on moral awareness earned higher ratings on past OERs. The course of action for this dilemma was negatively correlated with past ratings on OERs (r = -0.18), suggesting that individuals who opted to turn the refugees away received fewer "Above Center Mass" or "Most Qualified" ratings. Finally, the MES for the refugee dilemma was significantly positively correlated with the assignment location item (r = 0.27).

Table 24. *Correlations among Refugee Dilemma and Self-Rated Criteria*

Criteria	PRS	PCP	OERs	Work efforts	Asgt loc	Spec asgts
Ref dilemma	0.10	0.07	**-0.18**	-0.06	0.02	0.16
MA (RD)	-0.01	0.00	**0.25**	0.18	**-0.19**	-0.14
MC (RD)	-0.01	-0.04	0.00	0.07	-0.15	-0.04
SC (RD)	0.11	-0.05	-0.11	-0.04	0.02	-0.10
MES (RD)	0.02	0.01	0.21	0.14	**0.27**	-0.04

Note. Bolded values are significant at $p < .05$. RD = refugee dilemma, MA = moral awareness, MC = magnitude of consequences, SC = social consensus, MES = multidimensional ethics scale; PRS = performance rating scale, PCP = performance compared to peers, OERs = past four OERs, Asgt loc = assignment location, Spec asgt = special assignments. Sample size for the criteria ranged from 57 to 130.

Biodata, Cognitive, and Ethical Measures as Predictors of Other-Rated Performance Criteria

Despite previously-described challenges with the other-rated performance criteria—primarily, restriction of range—correlations among each set of predictor measures (i.e., biodata, cognitive and ethical) and select other-rated performance measures were examined for predictive validity evidence. Results of the correlations between each set of predictor measures and other-rated criteria are presented in Appendix I.

In summary, due to the challenges described above, primarily with range restriction, none of the predictor measures showed statistically significant levels of validity for predicting other-rated performance outcomes. Significant negative correlations were found between behavioral flexibility, extroversion, and reflective thinking and several of the other-rated performance measures. For the biodata measures, the only significant positive correlation exhibited was between family support and work efforts. The same pattern of results was found for the adjusted biodata scales. None of the cognitive or ethical measures demonstrated significant relationships with any of the other-rated performance criteria. This non-intuitive pattern of relationships (and lack of meaningful relationships) can most likely be explained by restriction of range in the performance criteria and in some of the predictor measures. For example, evidence of range restriction was also found for divergent thinking and some of the figural reasoning measures, which likely contributes further to attenuation of the correlations.

Summary

Initial validation evidence in the form of observed, uncorrected bivariate correlations between the predictor scales and self-rated criteria was positive overall. Results indicated that several of the measures were predictive of a range of self-rated performance measures. The

majority of biodata scales exhibited small to moderate positive correlations with self-evaluations via the performance rating scale, overall performance compared to peers, and past ratings on OERs. Despite some sample size limitations, meaningful correlations were found between ethical measures (i.e., moral awareness and MES) and self-rated performance criteria (primarily, ratings on past OERs). While the cognitive measures exhibited few significant relationships with self-rated criteria, this is likely explained by restriction of range in the predictor measures. Restriction of range occurred, in part, because the FA48 respondents to the assessment likely have higher levels of critical and divergent thinking skill; thus, these results do not necessarily indicate that these attributes are unrelated to FA48 performance.

Further validation evidence in the form of observed, uncorrected bivariate correlations between the predictor scales and other-related criteria was extremely limited. In addition to a low sample size for other-rated performance measures, severe restriction of range was present and afforded great challenge to interpreting correlations between the predictor measures and other-rated performance measures. Relationships among predictor measures and other-rated performance measures were minimal at best, with some biodata measures exhibiting negative correlations with performance measures.

To mitigate the restriction of range, one option is to perform a statistical correction. However, most accepted applications of correction techniques among selection researchers require an estimate of the unrestricted population variance for the measures (Hallgren, 2018). Such an estimate is not known in the current study, as only a restricted sample of individuals were assessed on the predictor and criterion measures, rather than the full population of individuals who would apply for the FA48 role. Additional correction techniques involving data imputation are not without flaws (Wiberg & Sundström, 2009), and many of them have not been studied enough to be well-understood. Specifically, data imputation techniques can result in over-correction, leading to inflation of validity estimates (Wiberg & Sundström, 2009), as well as under correction, leading to under-estimation of validity estimates. As any results obtained could not be interpreted with certainty, we opted not to proceed with statistical correction of correlations.

Based upon the findings described above, other-rated performance measures were not included in subsequent validity analyses. Instead, we identified the best-performing self-rated performance measures for inclusion in regression analyses. These analyses are described in detail in the following section.

Validity of Predictor Measures

This section presents criterion-related validity evidence for a subset of the predictor measures. Of particular interest is the ability of these measures to predict self-rated performance criteria. Despite limitations with self-rated performance data as described above, we utilized self-rated criteria for predictive analyses to estimate the relationship between the best-performing predictor scales and indicators of FA48 performance. Specifically, we used three self-rated performance measures as dependent variables, including the self-rated performance scale, overall performance compared to peers, and ratings on past four OERs, as these outcome measures exhibited acceptable psychometric properties, meaningful correlations with predictor measures,

and are more indicative of overall job performance than the biodata items that reflect specific facets of high job performance for FA48s.

Predictor Variable Inclusion Criteria

Due to a lack of statistical power as a result of the small sample size, we developed inclusion criteria to limit the set of predictors for the analyses. Specifically, we considered the distributional properties (i.e., means, variances), reliability estimates, correlations with outcome measures, and intercorrelations among predictors to prevent multicollinearity. Additionally, we accounted for the findings from previous conceptual and empirical efforts of this work (i.e., literature review, KSA survey, SME interviews) to determine the optimal set of predictors that would demonstrate the highest validity. As a part of this determination, we also aimed to include variables from each of the five categories of attributes where possible, including cultural competence and intelligence, ethical conduct, interpersonal and relationship-oriented skills, leading in challenging environments, and temperament.

Regression of Self-Rated Performance Scale on Predictor Measures

The predictor measures chosen for inclusion in this regression analysis were based on inclusion criteria as described above. First, among the biodata predictors, we narrowed the pool down to the subscales with the top 10 highest correlations with self-rated performance and then considered the intercorrelations among these predictor measures (see Appendix G for the intercorrelations of all biodata measures) as well as the attributes that emerged from the SME interviews and KSA survey as most critical. Cultural adroitness (representative of the culture competence and intelligence attribute category) was considered an important measure to include based upon the criteria listed above; given its high correlations with behavioral flexibility and extroversion, these two scales were not included. Oral communication (another top critical attribute) and peer leadership were correlated above 0.60, thus peer leadership was not included. Achievement orientation seemed to overlap highly with most biodata scales and was not considered a top critical attribute, thus this measure was not included. Stress tolerance, tolerance for ambiguity, and behavioral flexibility were all highly intercorrelated; however, stress tolerance was chosen among these for inclusion given its high correlations with self-rated performance. Given the overall criteria, we also chose to retain written communication, reflective thinking, and multitasking. Next, no cognitive scales were found to have significant correlations with performance and thus none were included. Finally, only the moral identity scale from the ethical measures was found to have significant correlations and was retained for the regression analyses.

In sum, we retained seven variables: cultural adroitness (CA), oral communication (OC), written communication (WC), reflective thinking (RT), stress tolerance (ST), multitasking (MT), and moral identity (Identity).

Results of regressing the performance rating scale on the set of predictors, using the non-adjusted biodata scales, revealed that the full set of predictors was significantly related to this self-rated criterion, $F(7, 51) = 3.36$, $p < .05$. Model R for the full set of predictor scales was 0.56 ($p < .05$; $R^2 = 0.32$). When examining the individual scales, none of the standardized regression coefficients were significant at $p < .05$, though oral communication and stress tolerance were the

strongest predictors ($p < .08$). Oral communication and stress tolerance accounted for 5% and 6%, respectively, of unique variance in the criterion.

Additionally, despite their significant positive correlations with the criterion, reflective thinking and multi-tasking exhibited minimal negative regression coefficients, suggesting evidence for multicollinearity. This finding is unsurprising, given the high degree of overlap among many of the biodata measures. We performed a second regression with these variables removed from the analysis. Table 25 presents a summary of relationships between each predictor scale and self-rated performance. This reduced set of predictors was significantly related to the criteria, $F (5, 53) = 4.80$, $p < .001$. Model R for this set of predictors was 0.56 ($p < .01$; $R^2 = 0.31$). Again, results revealed that oral communication and stress tolerance were the strongest predictors, though standardized regression coefficients only approached significance ($p < .07$ and $p < .08$, respectively). These measures accounted for 4.40% and 4.10%, respectively, of unique variance in performance ratings.

We repeated the regression analysis for the five predictor scales using adjusted biodata measures (i.e., cultural adroitness and stress tolerance). Overall results were similar to those found with non-adjusted biodata measures, $F (5, 53) = 4.28$, $p < .02$ (see Table 25). Model R for the five predictor scales was 0.54 ($p < .01$; $R^2 = 0.29$). Oral communication was the strongest predictor, contributing 5% of unique variance in the criterion. While stress tolerance remained the second strongest predictor, the magnitude of the relationship was weakened when we accounted for socially desirable responses. Stress tolerance uniquely contributed 3% of the total variance in performance ratings.

Table 25. *Validity Evidence for Predicting Self-Rated Performance*

Predictor / Criterion	Performance rating scale (unadj BD)	Performance rating scale (adj BD)
	Br	Br
Cultural adroitness	0.10	0.10
Oral communication	*0.26*	**0.28**
Stress tolerance	*0.25*	0.20
Written communication	0.10	0.09
Moral identity	0.16	0.17
Full model R	**0.56**	**0.54**

Note. Bolded values are statistically significant, $p < .05$, bolded italicized significant at $p < .1$; Model R = multiple correlation for full model; Br = standardized regression coefficient for the full model (all predictors); unadj BD = unadjusted biodata; adj BD = adjusted biodata. Sample size for the criteria ranged from 62 to 147.

Regression of Overall Performance Compared to Peers on Predictor Measures

As with selecting the predictors to be included in the analyses of self-rated performance, the set of variables chosen were based upon the inclusion criteria described above. Applying these same considerations, we retained five biodata subscales, including oral communication,

written communication, cultural adroitness, stress tolerance, and reflective thinking. Notably, no cognitive nor ethical predictors were included, given their lack of significant correlations with self-rated performance compared to peers.

Table 26 presents a summary of the relationships between each of the selected predictors and the self-rated item assessing overall performance compared to peers. Results of regressing performance compared to peers onto the set of predictors (with non-adjusted biodata scales) revealed that the full set of predictors was significantly related to this self-rated performance criterion, $F (5, 118) = 7.86, p < .001$. Model R for the full set of predictor scales was 0.50 ($p < .01, R^2 = 0.25$). When examining the individual scales, stress tolerance and written communication were the strongest predictors ($p < .01$), accounting for 5.50% and 5.30%, respectively, of unique variance in the criterion.

To determine whether results differed when we accounted for response distortion, we repeated the regression analysis using adjusted biodata measures (i.e., cultural adroitness, stress tolerance, and reflective thinking). The results remained largely the same; overall, the set of predictors was significantly related to the criterion, $F (5, 118) = 6.32, p < .001$. Model R for the full set of predictors was 0.46 ($p < .001$; $R^2 = 0.21$). Though the relationships were slightly weakened, stress tolerance and written communication still remained the strongest predictors. Both measures demonstrated significant standardized regression coefficients ($p < .01$), accounting for 3% and 4.50%, respectively, of the variance in the criterion.

Table 26. *Validity Evidence for Predicting Overall Performance Compared to Peers*

Predictor / Criterion	Performance compared to peers (unadj BD) Br	Performance compared to peers (adj BD) Br
Cultural adroitness	0.11	0.12
Oral communication	0.07	0.12
Stress tolerance	**0.28**	**0.20**
Written communication	**0.24**	**0.23**
Reflective thinking	0.03	0.02
Full model R	**0.50**	**0.46**

Note. Bolded values are statistically significant; Model R = Multiple correlation for full model; Br = Standardized regression coefficient for the full model (all predictors); unadj BD = unadjusted biodata; adj BD = adjusted biodata. Sample size for the criteria ranged from 128 to 147.

Regression of Past OER Ratings on Predictor Measures

As with selecting the predictors to be included in the analyses of self-rated performance and performance compared to peers, the set of variables were chosen based upon the inclusion criteria described above. Applying these same considerations, from the cognitive predictors, we retained the figural reasoning scales that exhibited positive correlations with this criterion measure, including the original figural reasoning and figural series scales. From the set of

biodata predictors, only the self-efficacy scale was retained based upon the application of the inclusion criteria. Finally, from the ethical measures, we retained the refugee dilemma items that were significantly correlated, including the course of action, moral awareness, and MES items.

Table 27 presents a summary of relationships between each selected predictor scale and the self-rated item assessing ratings on previous OERs. Results of regressing past OER ratings on the set of predictors (with unadjusted biodata scales) revealed that the full set of predictors was significantly related to this self-rated performance criterion, $F(6, 56) = 2.74$, $p < .05$ Model R for the full set of predictor scales was 0.48 ($p < .02$, $R^2 = 0.23$). When examining the individual scales, the strongest predictors were figural series (a figural reasoning measure) and the MES for the refugee dilemma. The figural series measure demonstrated a significant standardized regression coefficient at $p < .02$, contributing 8% of unique variance in the criteria. While the regression coefficient for the refugee dilemma MES was not significant at $p < .05$, it approached significance and explained 4.10% of the variance in past OER ratings. Of note, the sample size for MES measures was significantly lower than the other predictors ($n = 67$). It is possible that this relationship would be significant with a sample size comparable to the other predictors.

To determine whether results differed when we accounted for response distortion, we repeated the regression analysis using the adjusted self-efficacy biodata measure. Results are not reported for this analysis, as there were no differences in findings.

Table 27. *Validity Evidence for Predicting Past Ratings on OERs*

	Past four OERs (unadj BD)
Predictor / Criterion	Br
COA refugee dilemma	0.04
Moral awareness (RD)	0.20
MES (RD)	*0.24*
Figural reasoning- original	0.00
Figural reasoning - figural series	**0.31**
Self-efficacy	0.20
Full model R	**0.48**

Note. Bolded values are statistically significant, $p < .05$, bolded italicized significant at $p < .10$; model R = multiple correlation for full model; Br = standardized regression coefficient for the full model (all predictors); unadj BD = unadjusted biodata; adj BD = adjusted biodata; COA = course of action; MES = Multidimensional Ethics Scale; RD = refugee dilemma. Sample size for the criteria ranged from 67 to 139.

Summary

Despite several limitations, which are described in detail in the following chapter, multiple scales showed value in predicting self-rated performance criteria. Specifically, the biodata scales showed the most promise in predicting self-evaluations of FA48 performance as indicated by the eight-dimension rating scale and overall performance compared to peers.

Written communication, oral communication and stress tolerance accounted for the largest portions of unique variance in these two criteria. Cognitive and ethical measures also showed indication of predictability, demonstrating significant or near significant relationships with past ratings on OERs. The figural series measure uniquely accounted for 8% of the variance in past ratings on OERs, suggesting that cognitive abilities such as critical thinking show promise for predicting FA48 performance. While we were unable to examine the validity of predictor scales for other-rated performance measures, the results of the analyses we were able to conduct provided insight into the measures that should be included in a standardized selection tool for FA48s. In the next chapter, we outline a series of recommendations for next steps, including additional data collection, to inform the development of an assessment battery for longitudinal validation with FA48 applicants.

CHAPTER 7: STUDY OVERVIEW

Summary

The context of this research effort was the job role of Foreign Area Officers (FA48s) in the U.S. Army, as they require a unique and specialized skillset for successful job performance. As the use of valid, standardized selection assessments for officer assignments is not common, current selection procedures for FA48s likely do not include the unique KSAs needed for optimal FA48 job performance. As such, the overarching goal of this effort was to identify the most critical attributes for FA48 job performance and subsequently develop and validate a selection tool to measure these attributes as a part of the application process.

To gain deeper insight into what job performance entails for an FA48, as well as the needed KSAs to enable high levels of job performance, this research effort began with a literature review followed by interviews with FA48 subject matter experts and the administration of a KSA survey. In particular, findings from these efforts were utilized to (1) determine the attributes that are most important for FA48 performance, (2) determine the attribute measures that should be included in an assessment battery, and (3) develop or select FA48 criterion measures to validate the assessment battery.

Following these procedures, the finalized assessment battery was assembled and subsequently administered to current FA48s for concurrent validation to evaluate the psychometric properties of the measures. Supervisor and peer ratings of performance were collected for FA48s who completed the survey to provide preliminary validation of the assessment. That is, to evaluate the validity of the developed assessment battery for predicting FA48 officer performance, a concurrent validation was conducted. Assessment measures administered to active-duty FA48 officers were linked to current job performance criteria provided by the FA48 officer, their rater, and their peers to inform the best combination of assessment scores.

Finally, as described in the previous chapter, the results of the concurrent validation study provided evidence for multiple scales to predict self-rated performance criteria, including biodata scales (written communication, oral communication and stress tolerance), as well as cognitive (figural series, critical thinking) and ethical measures. Overall, the results of the analyses conducted provide insight into the measures that should be included in a standardized selection tool for FA48s, as well as additional steps that could be taken to further inform the optimal set of attribute measures for selecting FA48s.

Limitations

Though these research efforts afford the Army insight into improved selection procedures for the FA48 role, we did face several limitations related to the data collection administration, sample, methodology, and analyses.

Data Collection Administration

First, we faced challenges related to accessibility for participants who were recruited to take the surveys, largely due to the sensitive nature of the data collected (i.e., performance data). The full assessment was hosted on the Army survey platform Verint; however, many FAOs were unable to access this platform given the nature of their requisite assignments. As such, the research team created a work around for those participants, resulting in the use of two different survey administration methods. When the data collected from these methods were consolidated however, the research team discovered that some of the measures were removed from the alternate data collection method. This resulted in a much smaller response rate to some of these critical measures, particularly the ethical predictors.

Next, we faced the challenge of assessment length. As the battery contained many measures, it was time consuming for the participants to complete. Additionally, the battery began with the most time-consuming cognitive measures; as such, participation dropped off following the completion of these measures. With this drop off, far fewer responses were recorded for the remaining scales. This trend suggests early fatigue, which may affect the quality of the data, particularly for the non-cognitive measures.

Sample

Relatedly, several limitations emerged regarding the sample, starting with the low participation rate for the reasons described above. Additionally, the findings showed evidence of restriction of range for both the predictor and criterion data. Though reasons related to range restriction for specific variables are described in the requisite sections of the previous chapter, we subsequently describe overall impacting trends regarding the participant samples for the predictor and criterion data collections.

Predictor. Regarding the restriction of range in the predictor variables, we note that the population of interest is likely to score highly on the predictors, as they have already been selected for the FA48 job role and have been promoted to rank O4 or higher. Past ARI studies show that the traits and attributes included herein are predictive of promotion to higher ranks (Zaccaro et al., 2012, 2015). This is particularly true for the cognitive predictors, as this population of Army officers are highly educated and are likely at the high end of the cognitive ability spectrum. Further exacerbating this trend, high performers are most likely to have self-selected to participate in the voluntary study.

Criterion. Past attempts to develop criterion measures in similar contexts have suffered from issues of restriction of range as SMEs have been unwilling to rate participants lower than the highest one or two performance categories. Despite efforts described throughout this report to mitigate such a response trend, it still persisted. Both peer and supervisor raters tended to rate FA48 officers on the high end of the rating scale (i.e., well above average or truly exceptional). This trend emerged for both the performance rating scale and the newly developed biodata measures. With a low sample size and inability to estimate unrestricted population variance for the measures, we opted not to utilize statistical correlations to account for range restriction, though such an approach is possible if results are interpreted with caution.

Methodology

The primary limitation regarding the methodology relates to the measures. Specifically, due to challenges of accessibility, the research team opted to use existing ARI scales wherever possible. As such, the assessment battery comprised a high proportion of biodata scales, which carries the inherent limitations associated with this type of self-report measure (e.g., social desirability response trends). Though we made efforts to prevent this impact, some impact was still found and needed to be accounted for in the interpretation of the results. Ideally, given unlimited time and resources, the research team would have preferred to develop a variety of measure types to assess the constructs of interest.

Analyses

Two limitations emerged related to the analyses. First, given the lack of variance in the performance ratings from peers and supervisors, the findings are primarily based upon the self-rated performance measures as the primary criterion. Notably, this data was collected at the same time point via the same survey as the predictor measures, which lends theoretical limitations regarding causality. Second, analyses suffered from a lack of power resulting from the small sample size. As such, the research team had to strategically determine which predictors to include in regression analyses and were unable to conduct exploratory analyses and moderated regressions to check for differences related to demographics or job-related characteristics.

Recommendations for Next Steps

Given the limitations as described above, the results of the concurrent validation should be interpreted and applied with caution. However, the findings do offer several benefits and contributions related to the primary aims of these research efforts. First, most of the predictor measures demonstrated adequate psychometric properties, particularly the biodata. Next, analyses of the ethical measures showed value for their predictability of FA48 performance and also uncovered some interesting results that warrant further analysis. Finally, the results of the concurrent validation may guide the development of a longitudinal assessment battery via additional data collection to support further analysis / decision making. Relatedly, future efforts can be designed to create criterion measures that are specific to FAO roles or AOCs, illuminating important performance differences across the AOCs.

The primary recommendation for next steps in this study involves the collection of additional data. Such efforts will serve to further evaluate the predictability of ethical and cognitive measures, particularly if greater variance in outcome measures can be obtained. Based on the findings, such data collection efforts should further explore the moral identity and ethical dilemma measures. While the moral identity measure approached significance for predicting self-rated performance criteria, the sample size was very limited. Further research should further examine the value of this measure for predicting FA48 performance.

Regarding the ethical dilemmas, future research should aim to develop an appropriate scoring mechanism for the FA48 population and context. Specifically, additional data collection with a larger sample size could further examine the way participants respond across dilemmas in terms of both the course of action preferred as well as the moral/ethical concerns associated.

Such an analysis could inform response patterns that would be considered "red flags," which could then inform development of "exclusion criteria" for FA48 selection. For example, if participants exhibit little to no moral awareness in regard to their preferred course of action (regardless of which course of action is preferred), they may not be well-suited for the FA48 environment where ethical decision-making is a key attribute for success. Further research should also examine ethical dilemma response patterns in relation to job performance or other important outcomes to determine if there is a desired response, or set of responses, for each scenario.

Additional data collection could also examine differences in predictive validity of the predictor measures based upon region of the world and/or area of concentration to determine whether some constructs are more important for performance in certain regions versus others. Finally, further evaluation is needed for cultural attributes (e.g., cultural adroitness) and strategic thinking measures against criterion, as these were found in our early work to be highly important KSAs for FA48 performance.

Overall, we recommend moving forward with consideration of both results from this concurrent validation, with caution, as well as all prior work to identify the most critical KSAs for FA48 performance (i.e., literature review, SME interviews, KSA survey) toward a longitudinal validation effort with FA48 applicants. In such an effort, we make several key recommendations.

First, we recommend the reduction of the overall assessment length to reduce fatigue in participants. This can be accomplished by reducing the number of cognitive measures to only those deemed most important, as well as excluding measures that are conceptually overlapping. Regarding the biodata scales, we suggest the use of revised biodata scales, removing the items with lower item-total correlations. Additionally, the overall number of biodata scales in the battery should be reduced based upon intercorrelations, as many scales seem to have conceptual overlap. For example, conceptual overlap between stress tolerance and tolerance for ambiguity suggests that some of these scales can be removed. Stress tolerance was predictive of self-rated performance in a regression analysis, but highly correlated with the lie scale and may be better to replace with a similar measure. Further, biodata scales for constructs not identified as highly important in prior work might be removed (e.g., self-efficacy, multitasking). Conversely, we recommend increased focus upon variables in the culture competence and intelligence category, as the cultural adroitness scale seemed to be a promising predictor based on correlations with several of the self-report measures. Given the multicultural nature of the FA48 performance environment, a measure of cultural competence and intelligence is crucial for selecting high-performing officers into this role. Measures such as oral communication and written communication showed evidence of predictive validity and are recommended for inclusion in a longitudinal validation with FA48 applicants, especially given the criticality ratings (i.e., as determined from SME interviews and a KSA survey) of these attributes for FA48 job performance.

Regarding the cognitive measures, given the limited evidence supporting the inclusion of divergent thinking, additional data collection is needed to confirm this decision. Specifically, we recommend further review of findings from ARI studies in other functional areas, as well as predictors of performance and promotion at different officer ranks. If such findings support the

inclusion of divergent thinking, it may be worthwhile to include in the assessment battery. Additionally, based upon the data collected, the research team found it difficult to make a clear determination regarding figural reasoning. Though slight evidence emerged for one of the experimental figural reasoning scales predicting self-rated performance, additional input is needed to determine whether this measure should be included in the longitudinal assessment. At minimum, the number of figural reasoning items should be reduced to the best-performing set.

Finally, regarding ethical measures, more work is needed overall to support the necessary analysis of these measures. Given the high importance of ethical behavior for the FA48 role, supported by high criticality ratings for ethical attributes, we recommend inclusion of ethical measures in the assessment battery. Specifically, given the interesting nature of the findings, several potential options are available for exploring how participants respond across the ethical dilemmas. The moral identity measure also showed promise, although the limited sample size likely impacted the magnitude of relationships with performance outcomes.

REFERENCES

Abbe, A. & Gallus, J. A. (2012). *The Socio-Cultural Context of Operations: Culture and Foreign Language Learning for Company-Grade Officers* (Technical Report 1316). U.S. Army Research Institute for the Behavioral and Social Sciences.

Ahmad, S. (2018). Can ethical leadership inhibit workplace bullying across East and West: Exploring cross-cultural interactional justice as a mediating mechanism. *European Management Journal, 36*(2), 223-234. http://doi.org/10.1016/201801003

Ang, S., & Van Dyne, L. (2008). Conceptualization of cultural intelligence: Definition, distinctiveness, and nomological network. In S. Ang & L. Van Dyne (Eds.), *Handbook of cultural intelligence: Theory, measurement, and applications* (pp. 3 – 15). M.E. Sharpe.

Ang, S., Van Dyne, L., & Koh, C. (2006). Personality correlates of the four-factor model of cultural intelligence. *Group and Organization Management, 31*(1), 100-123. http://doi.org/10.1177/1059601105275267

Ang, S., Van Dyne, L., Koh, C. K. S., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation, and task performance. *Management and Organization Review, 3*(3), 335 – 371. http://doi.org/10.1111/j.1740-8784.2007.00082.x

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Baard, S. K., Rench, T. A., & Kozlowski, S. W. (2014). Performance adaptation: A theoretical integration and review. *Journal of Management, 40*(1), 48-99. http://doi.org/10.1177/0149206313488210

Blais, A., & Thompson, M. (2013). What would I do? Civilians' ethical decision making in response to military dilemmas. *Ethics & Behavior, 23*(3), 237–249. http://doi.org/10.1080/10508422.2012.748634

Brouer, R. L., Douglas, C., Treadway, D.C. & Ferris, G.R. (2012). Leader political skill, relationship quality, and leadership effectiveness: A two-study model test and constructive replication. *Journal of Leadership & Organizational Studies 20*(2) 185–198. http://doi.org/10.1177/1548051812460099

Brown, R. D., & Hauenstein, N. M. (2005). Interrater agreement reconsidered: An alternative to the rwg indices. *Organizational Research Methods, 8*(2), 165-184. http://doi.org/10.1177/1094428105275376

Brown, M. E., Trevino, L. K., & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes, 97*(2), 117−134. http://doi.org/10.1016/j.obhdp.2005.03.002

Caligiuri, P., & Tarique, I. (2012). Dynamic cross-cultural competencies and global leadership effectiveness. *Journal of World Business, 47*(4), 612-622. http://doi.org/10.1016/j.jwb.2012.01.014

Campion, E. D., Campion, M. C., & Campion, M. A. (2019). Best practices when using 360 feedback for performance appraisal. In A. H. Church, D. W. Bracken, J. H. Fleenor, & D. S. Rose (Eds.), *The handbook of strategic 360 feedback* (pp. 19–59). Oxford University Press.

Caughron, J. J., Antes, A.L., Stenmark, C.K., Thiel, C.E., Wang, X. & Mumford, M.D. (2011). Sensemaking strategies for ethical decision making. *Ethics & Behavior, 21*(5), 351–366. http://doi.org/10.1080/10508422.2011.604293

Chen, G., Kirkman, B. L., Kim, K., Farh, C. I. C., & Tangirala, S. (2010). When does cross-cultural motivation enhance expatriate effectiveness? A multilevel investigation of the moderating roles of subsidiary support and cultural distance. *Academy of Management Journal, 53*(5), 1110–1130. http://doi.org/10.5465/AMJ.2010.54533217

Chiu, C.-y., & Hong, Y.-y. (2005). Cultural Competence: Dynamic Processes. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 489–505). The Guilford Press.

Craft, J. L. (2013). A review of the empirical ethical decision-making literature: 2004–2011. *Journal of Business Ethics, 117*(2), 221-259. http://doi.org/10.1007/s10551-012-1518-9

De Hoogh, A. H .B. & Den Hartog, D. N. (2008). Ethical and despotic leadership, relationships with leader's social responsibility, top management team effectiveness and subordinates' optimism: A multi-method study. *The Leadership Quarterly, 19*(3), 297–311. http://doi.org/ 10.1016/j.leaqua.2008.03.002

Dragoni, L., Oh, I., Vankatwyk, P. & Tesluk, P. E. (2011). Developing executive leaders: The relative contribution of cognitive ability, personality, and the accumulation of work experience in predicting strategic thinking competency. *Personnel Psychology, 64*(4), 829–864. http://doi.org/10.1111/j.1744-6570.2011.01229.x

Dunning, D. (2011). The Dunning-Kruger Effect: On being ignorant of one's own ignorance. In J. Olson and M. P. Zanna (Eds.) *Advances in experimental social psychology* (Vol. 44, pp. 247-296). Elsevier.

Endsley, M. R., Garland, D. J., Wampler, R. L., & Matthews, M. D. (2000). *Modeling and measuring situation awareness in the infantry operational environment (*Research Report 1753). U.S. Army Research Institute for the Behavioral and Social Sciences.

Erkutlu, H. & Chafra, J. (2016). Impact of behavioral integrity on organizational identification: The moderating roles of power distance and organizational politics. *Management Research Review, 39*(6), 672-691. http://doi.org/10.1108/MRR-01-2015-0011

Ewen, C., Wihler, A., Blickle, G., Oerder, K., Ellen III, B. P., Douglas, C. & Ferris, G. R. (2013). Further specification of the leader political skill–leadership effectiveness relationships: Transformational and transactional leader behavior as mediators. *The Leadership Quarterly, 24*(4), 516–533. http://doi.org/10.1016/j.leaqua.2013.03.006

Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C. & Frink, D. D. (2005). Development and validation of the Political Skill Inventory. *Journal of Management, 31*(1), 126-152. http://doi.org/10.1177/0149206305503100410

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. http://doi.org/10.1037/0021-9010.92.2.347

Franklin, J. A. (2010). *The U.S. Army Foreign Area Officer Program*. Retrieved September 13, 2017, from www.disam.dsca.mil/pubs/Vol%2010-3/Franklin.pdf

Friedman, R., Hong, Y.-Y., Simons, T., Chi, S.-C., Oh, S.-H., & Lachowicz, M. (2018). The impact of culture on reactions to promise breaches: Differences between East and West in behavioral integrity perceptions. *Group & Organization Management, 43*(2), 273–315. http://doi.org/10.1177/1059601116678101

Gino, F., Schweitzer, M. E., Mead, N. L. & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes, 115*(2), 191–203. http://doi.org/10.1016/j.obhdp.2011.03.001

Groves, K. S., & Feyerherm, A. E. (2011). Leader cultural intelligence in context: Testing the moderating effects of team cultural diversity on leader and team performance. *Group & Organization Management, 36*(5), 535–566. http://doi.org/10.1177/1059601111415664

Hallgren, K. A. (2018). Restriction of Range. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1781-1783). Sage Publications.

Hardison, C., Sims, C. S., Ali, F., Villamizar, A., Mundell, B., & Howe, P. (2009). *Cross-cultural skills for deployed Air Force personnel* (MG-811-AF). RAND Corporation.

Imai, L., & Gelfand, M. J. (2010). The culturally intelligent negotiator: The impact of cultural intelligence (CQ) on negotiation sequences and outcomes. *Organizational Behavior and Human Decision Processes, 112*(2), 83-98. http://doi.org/10.1016/j.obhdp.2010.02.001

Jones, G. E. & Kavanagh, M. J. (1996). An experimental examination of the effects of individual and situational factors on unethical behavioral intentions in the workplace. *Journal of Business Ethics, 15*(5), 511-523. http://doi.org/10.1007/BF00381927

Joosten, A., van Dijke, M. Van Hiel, A. & De Cremer, D. (2014). Being ''in control'' may make you lose control: The role of self-regulation in unethical leadership behavior. *Journal of Business Ethics, 121*(1), 1-14. http://doi.org/10.1007/s10551-013-1686-2

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. http://doi.org/10.1007/s10551-010-0685-9

Kellett, J., Humphrey, R., & Sleeth, R. (2006). Empathy and the emergence of task and relations leaders. *The Leadership Quarterly, 17*(2), 146–162. http://doi.org/10.1016/j.leaqua.2005.12.003

Krosch, A. R., Figner, B. & Weber, E. U. (2012). Choice processes and their post-decisional consequences in morally conflicting decisions. *Judgment and Decision Making, 7*(3), 224–234. http://doi.org/10.1017/s1930297500002217

Lakshman, C. (2013). Biculturalism and attributional complexity: Cross-cultural leadership effectiveness. *Journal of International Business Studies, 44*(9), 922–940. http://doi.org/10.1057/jibs.2013.36

Lakshman, C., & Estay, C. (2016). Attributional complexity and leadership: Test of a process model in France and India. *International Journal of Cross Cultural Management, 16*(1), 53-76. http://doi.org/10.1177/1470595815622653

Lee, L. Y. & Kartika, N. (2014). The influence of individual, family, and social capital factors on expatriate adjustment and performance: The moderating effect of psychology contract and organizational support. *Expert Systems with Applications, 41*(11), 5483–5494. http://doi.org/10.1016/j.eswa.2014.02.030

Leone, L., Van der Zee, K. I., van Oudenhoven, J. P., Perugini, M. & Ercolani, A. P. (2005). The cross-cultural generalizability and validity of the Multicultural Personality Questionnaire. *Personality and Individual Differences, 38*(6), 1449–1462. http://doi.org/10.1016/j.paid.2004.09.010

Mayer, D., Aquino, K., Greenbaum, R., & Kuenzi, M. (2012). Who displays ethical leadership, and why does it matter? An examination of antecedents and consequences of ethical leadership. *The Academy of Management Journal, 55*(1), 151–171. http://doi.org/10.5465/amj.2008.0276

McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509. http://doi.org/10.1037/0003-066X.52.5.509

Moore, C., Detert, A. R., Trevino, L. K., Baker, V. L. & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology, 65*(1), 1–48. http://doi.org/10.1111/j.1744-6570.2011.01237.x

Mor, S., Morris, M. & Joh, J. (2013). Identifying and training adaptive cross-cultural management skills: The crucial role of cultural metacognition. *Academy of Management Learning & Education, 12*(3), 453–475. http://doi.org/10.5465/amle.2012.0202

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. http://doi.org/10.1037/0021-9010.93.4.733

Olsen, O. K., Eid, J., & Johnsen, B. H. 2006. Moral behavior and transformational leadership in Norwegian naval cadets. *Military Psychology, 18* (Suppl): S37-S56. http://doi.org/10.1207/s15327876mp1803s_4

Peltokorpi, V. & Froese, F. J. (2012). The impact of expatriate personality traits on cross-cultural adjustment: A study with expatriates in Japan. *International Business Review, 21*(4), 734–746. http://doi.org/10.1016/j.ibusrev.2011.08.006

Putka, D. J. (Ed.). (2009). *Initial development and validation of assessments for predicting disenrollment of four-year scholarship recipients from the Reserve Officer Training Corps* (Study Report 2009–06). U.S. Army Research Institute for the Behavioral and Social Sciences.

Reed, A, II., & Aquino, K. F. (2003). Moral identity and the expanding circle of moral regard toward out-groups. *Journal of Personality and Social Psychology, 84*(6), 1270–1286. http://doi.org/10.1037/0022-3514.84.6.1270

Reidenbach, R. E., & Robin, D. P. (1990). Toward the development of a multidimensional scale for improving evaluations of business ethics. *Journal of Business Ethics, 9*(8), 639-653. http://doi.org/10.1007/BF00383391

Riggio, R. E., Riggio, H. R., Salinas, C. & Cole, E. J. (2003). The role of social and emotional communication skills in leader emergence and effectiveness. *Group Dynamics: Theory, Research, and Practice, 7*(2), 83–103. http://doi.org/10.1037/1089-2699.7.2.83

Rockstuhl, T., Ang, S., Ng, K.-Y., Lievens, F., & Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology, 100*(2), 464-480. http://doi.org/10.1037/a0038098

Rockstuhl, T., Seiler, S., Ang, S., Dyne, L. V., & 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. http://doi.org/10.1111/j.1540-4560.2011.01730.x

Sadri, G., Weber, T.J., & Gentry, W.A. (2011). Empathic emotion and leadership performance: An empirical analysis across 38 countries. *The Leadership Quarterly, 22*(5), 818-830. http://doi.org/10.1016/j.leaqua.2011.07.005

Shaffer, M. A., Harrison, D. A., Gregersen, H., Black, J. S., & Ferzandi, L. A. (2006). You can take it with you: Individual differences and expatriate effectiveness. *Journal of Applied Psychology, 91*(1), 109-125. http://doi.org/10.1037/0021-9010.91.1.109

Shao, R. D., Aquino, K., & Freeman, D. (2008). Beyond moral reasoning: A review of moral identity research and its implications for business ethics. *Business Ethics Quarterly, 18*(4), 513–540. http://doi.org/10.5840/beq200818436

Stenmark, C. K., Antes, A. L., Wang, X., Caughron, J. J., Thiel, C. E., & Mumford, M. D. (2010). Strategies in forecasting outcomes in ethical decision-making: Identifying and analyzing the causes of the problem. *Ethics and Behavior, 20*(2), 110-127. http://doi.org/10.1080/10508421003595935

Stenmark, C. K., Antes, A. L., Thiel, C. E., Caughron, J. J., Wang, X., & Mumford, M. D. (2011). Consequences identification in forecasting and ethical decision-making. *Journal of Empirical Research on Human Research Ethics, 6*(1), 25-32. http://doi.org/10.1525/jer.2011.6.1.25

Sun, P. Y. T. & Anderson, M. H. (2012). The importance of attributional complexity for transformational leadership studies. *Journal of Management Studies 49*(6), 1001–1022. http://doi.org/10.1111/j.1467-6486.2011.01037.x

Tang, P. I., Chen, Y. J. & Sutarso, T. (2008). Bad apples in bad (business) barrels: The love of money, Machiavellianism, risk tolerance, and unethical behavior. *Management Decision, 46*(2), 243-263. http://doi.org/10.1108/00251740810854140

Thiel, C. E., Bagdasarov, Z., Harkrider, L., Johnson, J. F., & Mumford, M. D. (2012). Leader Ethical Decision-Making in Organizations: Strategies for Sensemaking. *Journal of Business Ethics, 107*(1), 49-64. http://doi.org/10.1007/s10551-012-1299-1

Thompson, M. H., Adams, B. D., & Sartori, J. A. (2006). *Moral and ethical decision making in Canadian Forces Operations*. (DRDC CR-2006–013). Defence Research and Development Canada.

Thompson, M., Hendriks, T., & Blais, A. (2018). Military ethical decision making: The effects of option choice and perspective taking on moral decision-making processes and intentions. *Ethics & Behavior, 28*(7), 578–596. http://doi.org/10.1080/10508422.2017.1372200

Treadway, D. C., Hochwarterb, W. A., Ferris, G. R., Kacmarc, C. J., Douglas, C., Ammetera, A. P. & Buckley, M. R. (2004). Leader political skill and employee reactions. *The Leadership Quarterly, 15*(4), 493–513. http://doi.org/10.1016/j.leaqua.2004.05.004

U.S. Department of the Army. 2014). *Commissioned Officer Professional Development and Career Management* (Pamphlet 600–3). Author.

Van der Zee, K. I. & van Oudenhoven, J. P. (2000). The Multicultural Personality Questionnaire: A multidimensional instrument of multicultural effectiveness. *European Journal of Personality, 14*(4), 291-309. http://doi.org/ 10.1002/1099-0984(200007/08)14:4<291::AID-PER377>3.0.CO;2-6

Van Woerkom, M. & de Reuver, R. S. M. (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. http://doi.org/10.1080/09585190903175589

Vitell, S., King, R., Howie, K., Toti, J., Albert, L., Hidalgo, E., & Yacout, O. (2016). Spirituality, moral identity, and consumer ethics: A multi-cultural study. *Journal of Business Ethics, 139*(1), 147–160. http://doi.org/10.1007/s10551-015-2626-0

Wasko, L. E., Putka, D. J., Legree, P. L., & Kilcullen, R. N. (2019). *Validation of measures for predicting leader development and assessment course performance* (Technical Report 1375). U.S. Army Research Institute for the Behavioral and Social Sciences.

Wiberg, M., & Sundström, A. (2009). A comparison of two approaches of restriction of range in correlation analysis. *Practical Assessment, Research, & Evaluation, 14*(5), 1–9. http://doi.org/10.7275/as0k-tm88

Zaccaro, S. J., Gilbert, J. A., Zazanis, M. M. & Diana, M. (1995). *Investigating a background data measure of social intelligence* (Technical Report 1024). U.S. Army Research Institute for the Behavioral and Social Sciences.

Zaccaro, S. J., Banks, D., Kiechel-Koles, L., Kemp, C., & Bader, P. (2009). *Leader and team adaptation: The influence and development of key attributes and processes* (Technical Report 1256). U.S. Army Research Institute for the Behavioral and Social Sciences.

Zaccaro, S. J., Connelly, S., Repchick, K. M., Daza, A., Young, M. C., Kilcullen, R. N., Gilrane, V. L., Robbins, J. M., & Bartholomew, L. N. (2015). The influence of higher order cognitive capacities on leader organizational continuance and retention: The mediating role of developmental experiences. *Leadership Quarterly, 26*(3), 342-358. http://doi.org/10.1016/j.leaqua.2015.03.007

Zaccaro, S. J., Gilrane, V. L., Robbins, J. M., Bartholomew, L. N., Young, M.C., Kilcullen, R. N., Connelly, S. & Young, W. (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 & Social Sciences.

Zeni, T. A., Buckley, M. R., Mumford, M. D. & Griffith, J. A. (2016). Making "sense" of ethical decision making. *The Leadership Quarterly, 27*(6), 838-855. http://doi.org/10.1016/j.leaqua.2016.09.002

APPENDICES

Appendix A: Literature Review Search Sample Keyword Search List

- Foreign Area Officer (performance, requirements)
- Working in a foreign country, multicultural work in the military
- Army/officer roles in foreign countries
- Multicultural context
    - Operating in . . .
    - Performance in . . .
    - Success in . . .
    - attributes related to . . .
- Performance . . . .
    - In ill-defined situations
    - In ambiguous environments
    - In unstructured environments
    - In environments without supervision
    - In self-management
    - without a supervisor
- Coordinating . . .
    - Across cultural boundaries
    - Across agencies
    - Boundary spanning
    - Boundary spanning across cultures
    - Cross-cultural communications
    - Strategizing across cultures
    - Strategizing across agencies
- Leadership . . .
    - In a multicultural setting
    - Multicultural . . .
    - Cross-cultural . . .
    - Organizational . . .
    - Attributes for . . . performance
    - Critical attributes for . . .

Appendix B: SME Structured Interview Protocol

<u>Introduction:</u>
Good morning/afternoon thank you for your participation in this interview today. My name is *(Interviewer Name)* and this is *(Note Taker Name)* from the U.S. Army Research Institute (ARI).

Our goal for today is to gain a better understanding of the knowledge, skills, and abilities of the Foreign Area Officers, as well as the dimensions of FA48 performance. We'll be using your interview responses to help in the development of an assessment that will be used to select Foreign Area Officers.

The interview will take between 30 and 45 minutes to complete, and we'll start with some brief background information about you and your position, as well as your experience with FAOs.

<u>Project Summary Form:</u>
For both in-person and phone interviews: Your participation today is voluntary. There are no consequences if you choose not to participate. If at any point you do not want to answer a question, please let us know and we will skip that question. You may also stop the interview at time. Anything you say during the interview will be kept confidential and will not be linked or attributed to you in any way.
In-person interviews: To more fully explain how we will be using the information you provide today, I have a project summary for you to review. Please take a few minutes to read over it. If you decide after reviewing the form that you do not want to participate, please return the form to me and return to your normal duties. *(Wait, until the interviewee has had a chance to read over the form before proceeding).*
Do you have any questions before we begin?
Phone interviews: Prior to each scheduled interview, participants will receive a copy of the project summary to review via email. We previously sent you a project summary via email. Have you had an opportunity to review the project summary?
If participant has not reviewed the project summary: Please take a few minutes to read over it now. (Wait, until the interviewee has had a chance to read over the form before proceeding). If you decide after reviewing the form that you do not want to participate, we can end the call and you can return to your normal duties. Do you have any questions before we begin?
If participant has reviewed the project summary: If you decide after reviewing the form that you do not want to participate, we can end the call and you can return to your normal duties. Do you have any questions before we begin?

<u>Question 1: Background & Experience with FA48s</u>
1. What is your name, rank, and current position?
    a. Have you held a FA48 position?
    b. How did you become an FA48? Do you know what made you stand out and get selected for an FA48 position?
    c. What were your main duties as an FA48?
    d. How long were you an FA48?
    e. How have you interacted with FA48s in the past?
    f. Have you experienced any challenges working with FA48s in the past? If so, please describe.

<u>Question 2: Senior Advisor Knowledge, Skills, and Abilities</u>
2. Based on your experience, what you think the main function of the FA48 role is?
    a. What do you think are the main types of tasks that FA48s perform?
    b. Are there other things that FA48s are doing outside of the normal job responsibilities?
    c. What knowledge, skills, or abilities do you see in the highest performance FA48s?
        i. If different, what knowledge, skills, or abilities do you think help FA48s avoid failure or poor performance on the job?
    d. What prior experience or background are needed to be a high performing FA48?
    e. What personality or character traits are needed to be a high performing FA48?
    f. What else would distinguish a high performing FA48 from an average or poor performer?

<u>Question 3: Performance Evaluation</u>
3. Based on your experience, how would you know if an FA48 is performing well or poorly?
    a. What are the signs of high performance? What are the signs of underperformance?
    b. What factors are currently used to evaluated FA48 performance?
    c. What measures or indicators are used to gauge FA48 performance/success?
    d. What factors are not evaluated but should be?
    e. Who do you believe is in the best position to be able to judge FA48 performance?
        i. What performance dimensions would each of these individuals be able to judge?
    f. When evaluating the FA48's performance, how are factors beyond the control of the senior advisors taken into account?

<u>Questions 4 & 5: Critical Incidents</u>
4. Please describe an episode of successful FA48 performance
    a. Why was the FA48 so successful?
    b. What knowledge, skills and/or abilities contributed to success?
5. Please describe an episode of poor FA48 performance, or issues that led to a FAO disciplinary incident
    c. Why did the FA48 have this negative issue?
    d. What did the FA48 lack in terms of knowledge, skills, and abilities?

<u>Questions 6 & 7: Wrap-Up</u>
6. As we described at the beginning, we are using your responses to these questions to develop a selection tool for FA48s. Do you have any additional thoughts, recommendations, or tips that you would like to share that you think we should include in the *assessment*?
7. Do you have any additional thoughts, recommendations, or tips that you would like to share on the best ways to measure and *evaluate FA48 performance*?

Appendix C: KSA List with Definitions

KSA	Definition
Cultural intelligence	Demonstrates intelligence specific to the ability to grasp, reason, and behave effectively in situations characterized by cultural diversity.
Cultural perspective taking	Thinks contextually and flexibly about issues from a foreign counterpart's perspective.
Cultural flexibility	Moves across multiple cultural environments with comfort and ease.
Ethnocentrism*	Believes that one's own cultural group is superior to others.
Integrity	Aligns words with actions, possesses honest and strong moral principles; demonstrates selflessness by doing the right thing regardless of personal and professional consequences.
Self-regulation	Guides one's own thoughts, behaviors, and feelings to reach goals.
Moral awareness	Takes responsibility for behaviors, concerned about harm to others, acts in a manner consistent with high ethical standards.
Ethical leadership	Conducts oneself appropriately through personal actions and interpersonal relationships and promotes such conduct to others.
Ethical decision making	Processes and understands ethical challenges and forms appropriate judgements for how to respond based on ethical principles.
Machiavellianism*	Manipulates others in the pursuit of self-interested goals; possesses a lack of concern for conventional morality.
Political skill	Understands social interactions at work and uses this understanding to influence others; appears genuine and sincere, builds strong social networks.
Empathy	Understands and sees situations from another's point of view.
Emotion regulation	Identifies, controls, and expresses one's emotions appropriately, and handles interpersonal relationships judiciously and empathetically.
Social expressiveness	Expresses ideas clearly and engages others in social discourse.
Social sensitivity	Interprets the verbal and nonverbal communication of others correctly, understands social situations, roles, and norms.
Social control	Presents self appropriately in social situations.
Active listening	Concentrates fully when listening, understands and responds to conversations effectively.
Negotiation	Reaches positive solutions when multiple groups disagree.
Conflict management	Limits the negative aspects of conflict and reaches desired group outcomes.
Behavioral flexibility	Responds to different situations in accurate and varied ways; modifies behavior to meet the situational requirements.
Written communication	Communicates clearly and concisely in writing.
Oral communication	Communicates verbally in a clear and concise manner.

KSA	Definition
Problem-solving skills	Analyzes problems to make sound decisions and identify strategies to find solutions.
Strategic thinking	Develops complex mental models of an organization as well as the broader environment to identify problems/opportunities and effective policies and operations that will lead to solutions.
Sensemaking	Develops an understanding of a dynamic situation by creating a mental representation of important elements in the situation.
Forecasting	Makes accurate predictions based on observations of the situation at hand.
Cognitive complexity	Thinks about situations in varied and nuanced ways, understands that a situation likely has multiple facets and responds accordingly.
Cognitive flexibility	Entertains new approaches to solving problems, creates new plans and ideas, and initiates and accepts change and innovation.
Tolerance for ambiguity	Handles working in unstructured, complex, or rapidly changing work environments.
Adaptability	Adapts rapidly to new information, changing conditions and strategy, or unexpected obstacles, processes, and requirements.
Peer leadership	Desires to obtain positions of authority and influence. Feels comfortable being in charge of a group and is willing to make tough decisions and accept responsibility for the group's performance.
Conscientiousness	Strives to do what is right, especially to do one's work or duty well and thoroughly.
Extraversion	Engages in outgoing, talkative, energetic behavior and is assertive in social situations.
Openness	Tries new experiences, is imaginative and open minded.
Emotional stability	Remains stable and balanced, doesn't easily experience and react to negative emotions.
Curiosity	Desires to learn new information and wants to understand novel ideas and concepts.
Learning goal orientation	Works toward goals of acquiring new knowledge or skills and is motivated to learn.
Work motivation	Gives one's best effort and works hard toward achieving difficult objectives.
Stress tolerance	Maintains one's composure under pressure and remains calm and in control of one's emotions instead of feeling anxious and worries.
Self-efficacy	Feels that one has successfully overcome work obstacles in the past and that one will continue to do so in the future.

Note. *Attribute is negative; a lack of this attribute is expected to be needed for successful FAO performance.

Appendix D: Sample of Newly Developed Predictor Measures

Scale	Item	Response Options
Prior international experience	How many overseas assignments have you held?	None, One, Two, Three, Four or more
	To what extent have you traveled to other countries outside of job responsibilities?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All
	To what extent have you engaged in international experiences before becoming a FAO, such as study abroad programs?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All
	How many languages do you maintain well enough to communicate with foreign nationals?	None, One, Two, Three, Four or more
Family support / compatibility	My family/significant other is supportive of my career as a FA48 officer.	Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree, Not Applicable
	My family/significant other is able to quickly adjust to new cultural environments.	Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree, Not Applicable
	My family/significant other is willing to frequently relocate to different parts of the world.	Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree, Not Applicable
	How often have you taken assignments where your family/significant other did not accompany you?	Never, Once, Twice, Three times, Four or more times, Not Applicable
	My family/significant other and I thoroughly discuss what the living situation might be like before relocating to a new country.	Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree, Not Applicable

E1. *Self-Rated Performance Rating Scales*

Instructions: Below are 8 dimensions of FA48 performance. Using the rating scale, please select the option that best describes your performance relative to other FAOs you know. Example behaviors are provided to help you make meaningful distinctions between performance dimensions.

Rating Scale Response Options and Descriptions:
- Below Average = In the bottom 25% of FAOs you know at this skill
- Average = In the middle 50% of FAOs you know at this skill
- Above Average = In the top 25% of FAOs you know at this skill
- Well Above Average = In the top 10% of FAOs you know at this skill
- Truly Exceptional = In the top 1% of FAOs you know at this skill
- N/A = Cannot rate

1. Thinks Strategically
Examples:
- Uses historical, cultural, economic and/or political knowledge of a region to identify security challenges or opportunities for the United States.
- Maintains focus on long term multinational objectives despite opportunities for short-term gains.

2. Influences Stakeholders
Examples:
- Balances goals and directives from multiple stakeholders and builds consensus.
- Influences foreign and domestic stakeholders without formal authority.

3. Adapts to Unfamiliar Cultural Settings
Examples:
- Adapts behavior to fit the cultural context (national or organizational).
- Implements creative solutions for overcoming obstacles.
- Uses alternative cultural lenses when interpreting events.

4. Manages Stress Effectively
Examples:
- Is calm and focused in high-pressure or unfamiliar cultural settings.

5. Demonstrates Personal Discipline and Ethical Behavior
Examples:
- Maintains professional appearance and demeanor in a multicultural / diplomatic environment; meets Army standards (e.g., fitness level, service requirements, etc.)
- Makes ethical decisions in the absence of supervision.
- Avoids temptation to choose easier, less ethical options without oversight or direct management.

6. Communicates Effectively
Examples:
- Communicates complex ideas regarding multinational matters effectively
- Adjusts communication style to meet audience needs in multicultural settings (e.g., differing national or organizational cultures).
- Uses regional knowledge of culture and language to articulate US objectives.

7. Works Effectively with Army Soldiers
Examples:
- Provides accurate and trusted advice to chain of command.
- Gains trust and respect of partners and colleagues.

8. Interacts Effectively with Foreign Partners
Examples:
- Develops effective cross-cultural relationships and networks to advance multinational goals and objectives.
- Obtains useful information from foreign partners.
- Provides reliable information to foreign partners to foster shared understanding.

Overall Performance: Compared to your peers of the same rank, would you say you are:
- One of the best (top 5%)
- Better than most (upper third)
- Average (middle third)
- Below average (lower third)

E2. *Self-Rated Performance Rating Biodata Scales*

Scale	Item	Response Options
Quality of work	How often are your work efforts included in strategy documents such as campaign plans or technical plans?	Hardly ever, Rarely, Occasionally, Frequently, Almost Always, Cannot Say
	To what extent are you recognized amongst peers as someone who is a high performer?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say
Ethical conduct	How often have you been formally disciplined for an ethics violation?	Never, Once, Twice, Three Times, Four or more times, Cannot Say
	How often have you been informally counseled about ethics violations or questionable ethical decisions?	Never, Once, Twice, Three Times, Four or more times, Cannot Say
	How often has a decision you made been called into question on ethical grounds?	Never, Once, Twice, Three Times, Four or more times, Cannot Say
Assignment experience	How frequently have you willfully taken assignments in dangerous or austere locations?	Never, Once, Twice, Three Times, Four or more times, Cannot Say
	How frequently have you sought assignments overseas?	Hardly ever, Rarely, Occasionally, Frequently, Almost always, Cannot say
	How often have you been pulled for special assignments based on particular regional knowledge or experience?	Never, Once, Twice, Three Times, Four or more times, Cannot Say

E3. *Supervisor- and Peer-Rated Performance Rating Scales*

Instructions: Please select the option for each dimension that best describes the performance of the FAO you are rating. Example behaviors are provided to help you make meaningful distinctions between performance dimensions.

Rating Scale Response Options and Descriptions:
- Below Average = In the bottom 25% of FAOs you know at this skill
- Average = In the middle 50% of FAOs you know at this skill
- Above Average = In the top 25% of FAOs you know at this skill
- Well Above Average = In the top 10% of FAOs you know at this skill
- Truly Exceptional = In the top 1% of FAOs you know at this skill
- N/A = Cannot rate
- [Open-ended] Optional comment / explanation

1. Thinks Strategically
Examples:
- Uses historical, cultural, economic and/or political knowledge of a region to identify security challenges or opportunities for the United States.
- Maintains focus on long term multinational objectives despite opportunities for short-term gains.

2. Influences Stakeholders
Examples:
- Balances goals and directives from multiple stakeholders and builds consensus.
- Influences foreign and domestic stakeholders without formal authority.

3. Adapts to Unfamiliar Cultural Settings
Examples:
- Adapts behavior to fit the cultural context (national or organizational).
- Implements creative solutions for overcoming obstacles.
- Uses alternative cultural lenses when interpreting events.

4. Manages Stress Effectively
Examples:
- Is calm and focused in high-pressure or unfamiliar cultural settings.

5. Demonstrates Personal Discipline and Ethical Behavior
Examples:
- Maintains professional appearance and demeanor in a multicultural / diplomatic environment; meets Army standards (e.g., fitness level, service requirements, etc.)
- Makes ethical decisions in the absence of supervision.
- Avoids temptation to choose easier, less ethical options without oversight or direct management.

6. Communicates Effectively

Examples:

- Communicates complex ideas regarding multinational matters effectively
- Adjusts communication style to meet audience needs in multicultural settings (e.g., differing national or organizational cultures).
- Uses regional knowledge of culture and language to articulate US objectives.

7. Works Effectively with Army Soldiers

Examples:

- Provides accurate and trusted advice to chain of command.
- Gains trust and respect of partners and colleagues.

8. Interacts Effectively with Foreign Partners

Examples:

- Develops effective cross-cultural relationships and networks to advance multinational goals and objectives.
- Obtains useful information from foreign partners.
- Provides reliable information to foreign partners to foster shared understanding.

Overall Performance: Compared to peers of the same rank, would you say the officer is:

- One of the best (top 5%)
- Better than most (upper third)
- Average (middle third)
- Below average (lower third)

E4. *Supervisor- and Peer-Rated Performance Rating Biodata Scales*

Scale	Item	Response Options
Quality of work products	How often are the officer's work efforts included in strategy documents such as campaign plans or technical plans?	Hardly ever, Rarely, Occasionally, Frequently, Almost Always, Cannot Say
	To what extent is the officer recognized amongst peers as someone who is a high performer?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say
Ethical conduct	To what extent have you questioned the integrity or ethicality of the officer's actions?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say
	To what extent have you questioned the officer's moral character?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say
	To what extent have you known the officer to disregard ethics standards?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say
	To what extent have you witnessed or heard of the officer misusing resources in any way (e.g., funding, equipment, vehicles)?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say
Assignment experience	How frequently has the officer willfully taken assignments in dangerous or austere locations?	Never, Once, Twice, Three Times, Four or more times, Cannot Say
	How frequently has the officer sought assignments overseas?	Hardly ever, Rarely, Occasionally, Frequently, Almost always, Cannot say
	How often has the officer been pulled for special assignments based on particular regional knowledge or experience?	Never, Once, Twice, Three Times, Four or more times, Cannot Say
Self-development initiative	To what extent does the officer take the initiative to develop the office and/or multinational programs over and above assigned tasks?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say
	To what extent does the officer engage in proactive learning to deepen understanding of a country or region?	Great Extent, Large Extent, Moderate Extent, Slight Extent, Not at All, Cannot Say

Appendix F: Ethical Measures

F1. *Moral Identity Scale*

Instructions: Listed below are some characteristics that may describe a person. The person with these characteristics could be you or it could be someone else. For a moment, visualize in your mind the kind of person who has these characteristics. Imagine how that person would think, feel, and act. When you have a clear image of what this person would be like, answer the following questions.

Stimulus Traits: Caring, Compassionate, Fair, Friendly, Generous, Hardworking, Helpful, Honest, Kind

Rating Scale Response Options: Strongly agree, Agree, Neither Agree nor disagree, Disagree, Strongly disagree

Items:
1. It would make me feel good to be a person who has these characteristics.
2. Being someone who has these characteristics is an important part of who I am.
3. A big part of my emotional well-being is tied up in having these characteristics.
4. I would be ashamed to be a person who has these characteristics.*
5. Having these characteristics is not really important to me.*
6. Having these characteristics is an important part of my sense of self.
7. I strongly desire to have these characteristics.
8. I often buy products that communicate the fact that I have these characteristics.
9. I often wear clothes that identify me as having these characteristics.
10. The types of things I do in my spare time (e.g., hobbies) clearly identify me as having these characteristics.
11. The kinds of books and magazines that I read identify me as having these characteristics.
12. The fact that I have these characteristics is communicated to others by my membership in certain organizations.
13. I am actively involved in activities that communicate to others that I have these characteristics.

*Reverse-coded item.

F2. *Military-Specific Ethical Dilemmas Scale*

Instructions: Please read the scenarios below and respond to the corresponding questions.

<div align="center">Scenario 1</div>

Subordinate Dilemma

Imagine that you are the commander of a unit on peacekeeping duty in a foreign country. There are two factions in this country, and you are trying to keep them from fighting. Your orders are to avoid fighting or siding with either faction. One of your subordinates is somebody who has been good friend for many years. Recently, he has been getting sympathetic to one of the factions. One day, you find out that he has deployed soldiers into this faction's area for protection. This is directly contrary to your orders and to your mission. He needlessly put soldiers' lives at risk, in an immediate zone of danger. He probably felt strongly that he was saving civilians' lives, and was hoping that you wouldn't find out about it. In a case like this, military rules say that he should be relieved of command and sent for a court-martial. However, you could reprimand him privately instead.

What would you do?
- I reprimand him privately.
- I relieve him of command and have him court-martialed.

<div align="center">Scenario 2</div>

Refugees Dilemma

Imagine that you are the commander of a unit on peacekeeping duty in a foreign country. There are two factions in this country, and you are trying to keep them from fighting. Your orders are to avoid fighting or siding with either faction. One of the factions starts to shell the town you are in. Thousands of bombs fall within 36 hours. Suddenly, hundreds of people from the other faction are outside your camp, trying to get away from the bombing. You contact headquarters for permission to let them in and the response is strict: Don't let them in. The concern is that our country must maintain impartiality to be effective in keeping the peace: letting people into our camp makes it look as if we are supporting their faction. Also, if we let a few in, thousands more will try to get in as well. We don't have enough resources to be able to keep them all safe, well fed, and free from diseases.

What would you do?
- I let them in.
- I turn them away.

Following each scenario, the following Moral Judgment items are presented.

Moral awareness: Overall, to what extent does this decision involve ethical and morality considerations?
- Not at all
- Slight extent
- Moderate extent
- Large extent
- Great extent

Magnitude of the consequences: The possible harm resulting from this option would be…
- Very minor
- Minor
- Moderate
- Severe
- Very severe

Social consensus: Most people would consider this option to be…
- Very inappropriate
- Somewhat inappropriate
- Neither inappropriate nor appropriate
- Somewhat appropriate
- Very appropriate

Morality: How well do the following characteristics describe this option?
    Response options: Not at all, Slight extent, Moderate Extent, Large Extent, Great extent
- Just
- Fair
- Morally right
- Acceptable to my family
- Culturally acceptable
- Traditionally acceptable
- Violates an unspoken promise
- Violates an unwritten contract

Table G1. *Intercorrelations among Biodata Measures*

	BF	CF	CA	OC	PL	E	RT	SE	ST	AO	MT	WC	TA	FS	TI	IT	IE	FL	SD
BF	1.00																		
CF	**0.35**	1.00																	
CA	**0.55**	**0.37**	1.00																
OC	**0.48**	**0.36**	**0.39**	1.00															
PL	**0.37**	**0.43**	**0.33**	**0.66**	1.00														
E	**0.47**	**0.18**	**0.49**	**0.45**	**0.36**	1.00													
RT	**0.45**	**0.56**	**0.39**	**0.39**	**0.33**	**0.24**	1.00												
SE	**0.23**	0.01	0.14	0.13	**0.21**	0.12	**0.19**	1.00											
ST	**0.56**	0.16	**0.44**	**0.42**	**0.34**	**0.36**	**0.20**	**0.31**	1.00										
AO	**0.47**	**0.52**	**0.39**	**0.53**	**0.54**	**0.40**	**0.55**	**0.18**	**0.31**	1.00									
MT	**0.53**	**0.22**	**0.32**	**0.37**	**0.36**	**0.39**	**0.29**	**0.21**	**0.38**	**0.38**	1.00								
WC	0.10	**0.37**	**0.22**	**0.25**	**0.38**	0.16	**0.25**	0.04	0.09	**0.32**	**0.23**	1.00							
TA	**0.55**	**0.43**	**0.31**	**0.40**	**0.40**	**0.31**	**0.43**	**0.19**	**0.52**	**0.52**	**0.46**	**0.24**	1.00						
FS	0.13	0.15	0.16	0.10	0.14	**0.18**	0.04	0.08	0.02	**0.21**	0.11	0.01	0.01	1.00					
TI	**0.25**	**0.48**	**0.32**	**0.33**	**0.34**	0.16	**0.35**	0.05	**0.23**	**0.47**	**0.32**	**0.24**	**0.29**	0.14	1.00				
IT	0.14	0.05	0.16	0.08	**0.23**	0.08	0.05	-0.02	0.12	**0.20**	**0.24**	0.08	0.15	**0.18**	0.07	1.00			
IE	0.14	0.11	**0.21**	**0.19**	**0.18**	**0.22**	**0.19**	-0.01	0.13	**0.22**	0.09	**0.23**	0.16	0.14	0.07	**0.36**	1.00		

BF	CF	CA	OC	PL	E	RT	SE	ST	AO	MT	WC	TA	FS	TI	IT	IE	FL	SD
FL	-0.01	-0.03	0.16	**0.19**	0.07	**0.19**	-0.12	**-0.17**	0.05	0.03	0.07	0.16	-0.09	0.06	0.01	0.07	0.15	1.00
SD	**0.34**	-0.05	**0.25**	0.13	0.02	0.12	**0.23**	**0.28**	**0.50**	0.14	0.12	-0.10	**0.22**	0.02	0.12	-0.05	0.02	0.06	1.00

Note. Bolded correlations are significant at the 0.05 level. Biodata subscales: BF = behavioral flexibility, CF = cognitive flexibility, CA = cultural adroitness, OC = oral communication, PL = peer leadership, E = extraversion, RT = reflective thinking, SE = self-efficacy, ST = stress tolerance, AO = achievement orientation, MT = multitasking, WC = written communication, TA = tolerance for ambiguity, FS = family support, TI = takes initiative, IT = international travel, IE = international experiences, FL = fluent languages, SD = social desirability. Sample size ranged from 121 to 150.

Table G2. *Intercorrelations among Adjusted Biodata Measures*

	BF	CA	RT	SE	ST	TA	SD
BF	1.00						
CA	**0.50**	1.00					
RT	**0.40**	**0.35**	1.00				
SE	0.15	0.08	0.14	1.00			
ST	**0.48**	**0.38**	0.11	**0.21**	1.00		
TA	**0.52**	**0.27**	**0.40**	0.14	**0.47**	1.00	
SD	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Note. Bolded correlations are significant at the 0.05 level. Biodata subscales: BF = behavioral flexibility, CA = cultural adroitness, RT = reflective thinking, SE = self-efficacy, ST = stress tolerance, TA = tolerance for ambiguity, SD = social desirability. Sample size ranged from 125 to 139.

Table G3. *Intercorrelations among Cognitive Measures*

	DT	FR-O	FR-PM	FR-FC	FR-FS	FR-FA
DT	1.00					
FR-O	0.14	1.00				
FR-PM	-0.05	**0.37**	1.00			
FR-FC	0.04	**0.31**	**0.40**	1.00		
FR-FS	0.16	**0.41**	**0.28**	**0.44**	1.00	
FR-FA	**0.31**	**0.35**	**0.26**	**0.33**	**0.31**	1.00

Note. Bolded correlations are significant at the 0.05 level. DT = divergent thinking, FR-O = figural reasoning-original, FR-PM = figural reasoning - pattern matrix, FR-FC = figural reasoning - figural classification, FR-FS = figural reasoning - figural series, FR-FA = figural reasoning - figural analogy. Sample size ranged from 95 to 163.

Table G4. *Intercorrelations among Ethical Measures*

	MI	SD	MA-SD	MC-SD	SC-SD	MES-SD	RD	MA-RD	MC-RD	SC-RD	MES-RD
MI	1.00										
SD	-0.09	1.00									
MA-SD	0.14	**0.32**	1.00								
MC-SD	0.18	**0.26**	**0.38**	1.00							
SC-SD	-0.11	**0.36**	0.09	0.07	1.00						
MES-SD	0.07	0.22	**0.37**	0.03	**0.32**	1.00					
RD	-0.06	0.10	-0.11	0.06	0.10	-0.11	1.00				
MA-RD	0.16	0.14	**0.39**	0.12	-0.03	0.20	**-0.35**	1.00			
MC-RD	-0.08	0.09	**0.23**	**0.28**	-0.13	-0.11	0.10	**0.30**	1.00		
SC-RD	0.16	**-0.17**	-0.07	0.08	**0.24**	0.15	**-0.19**	-0.05	**-0.27**	1.00	
MES-RD	0.07	-0.16	-0.07	0.05	0.04	**0.47**	**-0.51**	0.07	-0.17	**0.50**	1.00

Note. Bolded correlations are significant at the 0.05 level. MI = moral identity, SD = subordinate dilemma, MA-SD = moral awareness (SD), MC-SD = magnitude of consequences (SD), SC-SD = social consensus (SD), MES-SD = Multidimensional Ethics Scale (SD), RD = refugee dilemma, MA-RD = moral awareness (RD), MC-RD = magnitude of consequences (RD), SC-RD = social consensus (RD), MES-RD = Multidimensional Ethics Scale (RD). Sample size ranged from 61 to 133.

Table G5. *Correlations among Biodata and Cognitive Measures*

	DT	FR-O	FR-PM	FR-FC	FR-FS	FR-FA
BF	-0.05	0.01	0.08	-0.02	0.03	-0.11
CF	0.19	0.14	**0.20**	0.16	-0.03	0.11
CA	-0.03	-0.11	0.05	0.01	-0.10	-0.12
OC	0.18	0.04	0.13	-0.05	-0.08	0.01
PL	0.16	0.05	0.05	0.05	-0.04	**0.21**
E	-0.03	-0.05	0.14	-0.09	-0.07	0.03
RT	0.12	0.09	0.12	**0.20**	-0.07	0.05
SE	0.15	0.04	-0.07	0.05	-0.06	**0.19**
ST	0.11	-0.04	-0.04	0.10	0.17	0.09
AO	**0.24**	0.01	0.06	0.09	-0.05	0.06
MT	0.08	0.04	-0.06	-0.06	-0.18	-0.03
WC	0.02	0.01	0.07	0.08	-0.18	0.16
TA	0.16	0.16	0.16	**0.20**	0.14	0.07
FS	-0.09	-0.04	0.05	-0.09	-0.04	-0.04
TI	0.15	-0.03	0.02	-0.13	-0.14	0.01

	DT	FR-O	FR-PM	FR-FC	FR-FS	FR-FA
IT	0.04	-0.07	-0.17	-0.10	-0.17	-0.11
IE	0.06	0.03	0.01	-0.11	-0.10	-0.02
FL	0.06	-0.05	-0.05	-0.13	-0.11	-0.02

Note. Bolded correlations are significant at the 0.05 level. Biodata subscales: BF = behavioral flexibility, CF = cognitive flexibility, CA = cultural adroitness, OC = oral communication, PL = peer leadership, E = extraversion, RT = reflective thinking, SE = self-efficacy, ST = stress tolerance, AO = achievement orientation, MT = multitasking, WC = written communication, TA = tolerance for ambiguity, FS = family support, TI = takes initiative, IT = international travel, IE = international experiences, FL = fluent languages. Cognitive measures: DT = divergent thinking, FR-O = figural reasoning-original, FR-PM = figural reasoning - pattern matrix, FR-FC = figural reasoning - figural classification, FR-FS = figural reasoning - figural series, FR-FA = figural reasoning - figural analogy. Sample size ranged from 98 to 163.

Table G6. *Correlations among Biodata and Ethical Measures*

	MI	SD	MA-SD	MC-SD	SC-SD	MES-SD	RD	MA-RD	MC-RD	SC-RD	MES-RD
BF	0.07	0.05	**0.23**	-0.04	0.05	0.22	-0.02	0.07	0.05	0.00	0.13
CF	0.04	0.13	**0.22**	0.04	-0.07	**0.47**	-0.01	0.08	0.05	-0.03	**0.28**
CA	0.10	0.04	0.17	-0.03	0.06	**0.28**	0.03	0.08	0.06	0.07	0.09
OC	**0.27**	0.01	**0.20**	0.08	-0.01	0.21	0.02	0.05	-0.01	0.04	0.23
PL	**0.37**	0.09	**0.19**	0.11	0.03	0.12	0.01	**0.25**	0.14	-0.06	0.22
E	**0.26**	0.02	0.12	0.03	0.10	0.08	**0.20**	0.09	0.04	0.05	-0.08
RT	0.15	0.15	**0.25**	0.07	0.04	**0.29**	-0.17	0.13	0.02	-0.04	0.14
SE	**0.29**	0.05	**0.18**	-0.04	0.07	0.10	-0.07	**0.28**	0.01	-0.01	-0.06
ST	-0.11	-0.01	0.07	-0.11	0.12	0.21	0.11	-0.04	0.03	-0.03	0.15
AO	**0.35**	0.08	0.17	**0.23**	0.06	0.24	0.02	0.15	0.00	0.06	**0.28**

	MI	SD	MA-SD	MC-SD	SC-SD	MES-SD	RD	MA-RD	MC-RD	SC-RD	MES-RD
MT	**0.26**	0.16	**0.24**	0.11	-0.03	0.04	0.07	0.10	-0.01	**-0.18**	-0.04
WC	0.08	0.13	**0.18**	0.11	0.14	**0.25**	-0.02	0.15	0.03	-0.05	0.15
TA	**0.30**	0.07	**0.23**	0.00	0.04	**0.35**	0.02	**0.27**	0.06	-0.17	0.11
FS	**0.33**	-0.04	0.04	0.14	-0.07	0.02	**0.18**	0.06	-0.01	0.03	-0.09
TI	**0.28**	0.06	-0.01	0.11	0.04	0.21	-0.06	0.02	**-0.18**	0.03	**0.33**
IT	0.06	-0.01	0.00	0.13	0.04	0.12	-0.06	-0.03	-0.06	-0.05	**0.31**
IE	0.07	-0.10	-0.01	-0.02	0.03	0.21	0.04	0.00	**-0.17**	0.08	0.10
FL	-0.08	0.02	-0.03	-0.10	-0.03	0.11	0.16	-0.03	0.03	-0.06	0.16

Note. Bolded correlations are significant at the 0.05 level. Biodata subscales: BF = behavioral flexibility, CF = cognitive flexibility, CA = cultural adroitness, OC = oral communication, PL = peer leadership, E = extraversion, RT = reflective thinking, SE = self-efficacy, ST = stress tolerance, AO = achievement orientation, MT = multitasking, WC = written communication, TA = tolerance for ambiguity, FS = family support, TI = takes initiative, IT = international travel, IE = international experiences, FL = fluent languages. Ethical measures: MI = moral identity, SD = subordinate dilemma, MA-SD = moral awareness (SD), MC-SD = magnitude of consequences (SD), SC-SD = social consensus (SD), MES-SD = Multidimensional Ethics Scale (SD), RD = refugee dilemma, MA-RD = moral awareness (RD), MC-RD = magnitude of consequences (RD), SC-RD = social consensus (RD), MES-RD = Multidimensional Ethics Scale (RD). Sample size ranged from 58 to 150.

Table G7. *Correlations among Ethical and Cognitive Measures*

	DT	FR-O	FR-PM	FR-FC	FR-FS	FR-FA
MI	-0.21	-0.08	0.10	-0.15	-0.21	0.05
SD	**0.26**	0.11	0.07	0.14	-0.05	0.15
MA-SD	-0.09	0.03	0.03	-0.02	-0.10	0.00
MC-SD	-0.03	-0.06	-0.06	-0.04	-0.10	0.01
SC-SD	**0.24**	-0.10	0.02	0.07	0.05	-0.03
MES-SD	-0.14	0.04	0.04	-0.24	-0.06	-0.13
RD	-0.07	-0.07	-0.06	0.07	-0.05	-0.06
MA-RD	0.10	0.13	-0.08	-0.10	-0.02	0.08
MC-RD	-0.08	0.16	0.11	**0.24**	**0.22**	**0.31**
SC-RD	0.05	**-0.19**	-0.11	-0.19	**-0.22**	-0.16
MES-RD	0.12	0.02	-0.11	-0.18	-0.05	-0.11

Note. Bolded correlations are significant at the 0.05 level. Ethical measures: MI = moral identity, SD = subordinate dilemma, MA-SD = moral awareness (SD), MC-SD = magnitude of consequences (SD), SC-SD = social consensus (SD), MES-SD = Multidimensional Ethics Scale (SD), RD = refugee dilemma, MA-RD = moral awareness (RD), MC-RD = magnitude of consequences (RD), SC-RD = social consensus (RD), MES-RD = Multidimensional Ethics Scale (RD). Cognitive measures: DT = divergent thinking, FR-O = figural reasoning-original, FR-PM = figural reasoning - pattern matrix, FR-FC = figural reasoning - figural classification, FR-FS = figural reasoning - figural series, FR-FA = figural reasoning - figural analogy. Sample size ranged from 51 to 163.

Appendix H: Item Level Descriptives for Other-Rated Performance Data

	n	M	SD	Skew
Thinks strategically	95	4.23	0.68	-0.59
Influences stakeholders	96	4.11	0.73	-0.42
Adapts culture	91	4.20	0.66	-0.64
Manages stress	94	4.15	0.72	-0.44
Personal discipline	97	4.38	0.62	-0.70
Communicates effectively	96	4.17	0.74	-0.67
Works effectively	92	4.12	0.67	-0.52
Interacts effectively	89	4.32	0.69	-1.03
Overall performance	97	4.19	0.66	-0.45
Dangerous assignment location	61	3.62	1.07	-0.32
Overseas assignments	89	4.32	0.53	-0.38

Appendix I: Correlations between Predictor Measures and Other-Rated Criteria

Table I1. *Correlations among Adjusted Biodata and Other-Rated Performance Measures*

	Performance rating scale	Performance compared to peers	Self-development initiative	Work efforts	High performer
Behavioral flexibility	**-0.26**	-0.19	-0.15	**-0.23**	-0.11
Cognitive flexibility	-0.02	0.06	-0.09	0.13	0.10
Cultural adroitness	-0.16	-0.13	-0.11	-0.17	-0.10
Oral communication	-0.07	0.07	-0.02	-0.20	0.10
Peer leadership	-0.02	0.08	-0.09	-0.14	0.12
Extroversion	-0.18	-0.08	-0.05	**-0.23**	-0.03
Reflective thinking	-0.19	-0.04	**-0.25**	-0.18	0.01
Self-efficacy	-0.11	-0.09	-0.02	-0.17	0.01
Stress tolerance	-0.09	-0.07	0.01	-0.01	0.09
Achievement orientation	-0.02	0.04	-0.11	0.05	-0.04
Multi-tasking	-0.06	0.01	-0.05	0.07	-0.03
Written communication	0.08	0.08	-0.03	-0.07	0.09
Tolerance for ambiguity	0.07	0.15	0.04	-0.01	0.09
Family support	0.01	0.01	0.02	**0.23**	-0.12
Takes initiative	0.00	0.08	0.02	0.06	-0.16
International travel	-0.07	-0.05	-0.03	0.02	-0.01

	Performance rating scale	Performance compared to peers	Self-development initiative	Work efforts	High performer
International experiences	0.08	-0.04	0.09	0.06	-0.08
Fluent languages	0.11	0.00	0.04	0.11	0.00
Social desirability	-0.12	-0.09	0.06	-0.16	-0.01

Note. Bolded correlations are significant at the 0.05 level. Sample size ranged from 74 to 95.

Table I2. *Correlations among Adjusted Biodata and Other-Rated Performance Measures*

	Performance rating scale	Performance compared to peers	Self-development initiative	Work efforts	High performer
Behavioral flexibility	**-0.24**	-0.16	-0.18	-0.18	-0.11
Cultural adroitness	-0.16	-0.13	-0.15	-0.14	-0.11
Reflective thinking	-0.16	-0.02	**-0.27**	-0.15	0.01
Self-efficacy	-0.08	-0.06	-0.04	-0.13	0.01
Stress tolerance	-0.04	-0.03	-0.05	0.05	0.11
Tolerance for ambiguity	0.10	0.17	0.03	0.03	0.09

Note. Bolded correlations are significant at the 0.05 level. Sample size ranged from 72 to 94.

Table I3. *Correlations among Cognitive Measures and Other-Rated Performance Measures*

	Performance rating scale	Performance compared to peers	Self-development initiative	Work efforts	High performer
Divergent thinking	0.07	0.18	0.05	0.07	0.18
Figural reasoning - original	0.02	0.02	0.13	0.19	0.18
Figural reasoning - pattern matrix	-0.02	0.02	-0.15	-0.09	0.01
Figural reasoning - figural classification	0.04	0.07	-0.16	-0.08	0.13
Figural reasoning - figural series	-0.04	-0.03	-0.02	0.00	0.10
Figural reasoning - figural analogy	0.09	0.02	-0.10	-0.20	0.09

Note. No correlations are significant at the 0.05 level. Sample size ranged from 63 to 94.

Table I4. *Correlations among Subordinate Dilemma Ethical Measures and Other-Rated Performance Measures*

	Performance rating scale	Performance compared to peers	Self-development initiative	Work efforts	High performer
Moral identity	0.12	0.08	0.07	-0.24	-0.16
Subordinate dilemma	0.01	0.07	-0.05	0.12	0.09
Moral awareness (SD)	0.09	0.09	0.08	0.12	0.09
Magnitude of consequences (SD)	0.16	0.19	0.18	0.18	-0.04
Social consensus (SD)	-0.06	-0.10	-0.06	-0.12	0.06
Multidimensional Ethics Scale (SD)	-0.04	-0.14	-0.02	-0.12	0.03

Note. No correlations are significant at the 0.05 level. SD = subordinate dilemma. Sample size ranged from 38 to 95.

Table I5. *Correlations among Refugee Dilemma Ethical Measures and Other-Rated Performance Measures*

	Performance rating scale	Performance compared to peers	Self-development initiative	Work efforts	High performer
Refugee dilemma	-0.02	-0.02	0.02	0.10	0.06
Moral awareness (RD)	0.13	0.11	0.06	-0.07	-0.09
Magnitude of consequences (RD)	-0.06	-0.04	-0.09	-0.16	-0.03
Social consensus (RD)	0.04	0.03	0.14	0.06	-0.08
Multidimensional Ethics Scale (RD)	0.04	-0.01	0.13	-0.04	-0.04

Note. No correlations are significant at the 0.05 level. RD = refugee dilemma. Sample size ranged from 39 to 95.