Technical Report 1377

Cadet Training and Personality Metrics Longitudinally Predict Officer In-unit Performance: *R* = .37

Peter J. Legree
U.S. Army Research Institute

Justin Purl
Human Resources Research Organization

Robert N. Kilcullen
Mark C. Young
U.S. Army Research Institute

September 2019

United States Army Research Institute
for the Behavioral and Social Sciences

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

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE L. ZBYLUT, Ph.D.
Director

Technical review by:

Kristophor G. Canali, U.S. Army Research Institute
Elizabeth D. Salmon, U.S. Army Research Institute
Alisha M. Ness, U.S. Army Research Institute
Leonard A. White, U.S. Army Research Institute

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
September 2019	Interim	01 September 2015 – 30 September 2019

4. TITLE AND SUBTITLE
Cadet Training and Personality Metrics Longitudinally Predict Officer In-unit Performance: *R* = .37

5a. CONTRACT/GRANT NUMBER
W911NF-15-C-0063

5b. PROGRAM ELEMENT NUMBER
622785

6. AUTHOR(S)
Peter J. Legree, Justin Purl, Robert N. Kilcullen;
Mark C. Young

5c. PROJECT NUMBER
A790

5d. TASK NUMBER

5e. WORK UNIT NUMBER
1011

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street (Bldg. 1464 / Mail Stop: 5610)
Fort Belvoir, Virginia 22060-5610

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street (Bldg. 1464 / Mail Stop: 5610)
Fort Belvoir, Virginia 22060-5610

10. SPONSOR/MONITOR'S ACRONYM(S)
ARI

11. SPONSORING/MONITORING
Technical Report 1377

12. DISTRIBUTION AVAILABILITY STATEMENT
Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ARI Research POC: Dr. Peter J. Legree, Selection and Assignment Research Unit

14. ABSTRACT

Using data collected from U.S. Army officers before commissioning, we used a longitudinal design to validate ROTC Cadet Order of Merit list (OML) scores and training metrics against supervisor ratings of officer performance that were obtained up to eight years later. Analyses documented the substantial predictive validity of the ROTC OML and training metrics, *R* = .33. Analyses also showed that the cadet personality measures are valid predictors of the subsequent supervisor ratings of officer performance, *R* = .28. The combined predictive validity of the ROTC cadet training and personality metrics against the officer performance ratings was substantial, *R* = .37. These results demonstrate the potential utility of using metrics collected from ROTC cadets to predict their subsequent performance as U.S. Army officers. The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) supported this research project.

15. SUBJECT TERMS
Profile Similarity Metrics, Training Metrics, Personality

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

Standard Form 298 (Rev. 8-98)

Technical Report 1377

Cadet Training and Personality Metrics Longitudinally Predict Officer In-unit Performance: *R* = .37

Peter J. Legree
U.S. Army Research Institute

Justin Purl
Human Resources Research Organization

Robert N. Kilcullen
Mark C. Young
U.S. Army Research Institute

Selection and Assignment Research Unit
Tonia S. Heffner, Chief

September 2019

CADET TRAINING AND PERSONALITY METRICS LONGITUDINALLY PREDICT OFFICER IN-UNIT PERFORMANCE: *R* = .37

EXECUTIVE SUMMARY

Research Requirement:

The purpose of this research was to establish the predictive validity of the USACC Cadet Order of Merit List (OML) scores against officer performance metrics. The predictive validity of OML scores is important because these scores are used to assign higher scoring candidates to the active duty component and critical positions, and as an outcome to validate measures that are used to award ROTC scholarships.

Approach:

Supervisor ratings of the performance of 1,068 junior officers was collected and merged with an archive dataset that contained the Cadet OML scores, OML component scores, and personality data for those same officers. The time delay between the collection of the cadet and officer data was approximately 4.3 years. We used these data to estimate the long-term validity of the Cadet OML and personality metrics against officer performance ratings that were collected from in-unit supervisors using a longitudinal design.

Findings:

Despite the four-year time delay, analyses demonstrated substantial validity estimates for the ROTC cadet training metrics, *R* = .33. Analyses also showed that cadet personality measures predicted the subsequent supervisor ratings of officer performance, *R* = .28. The combined validity of the ROTC cadet training and personality metrics predicted the subsequent officer performance ratings, *R* = .37.

Utilization and Dissemination of Findings:

These results support the continued use of the USACC Cadet OML score to assign newly commissioned officers to critical positions. In addition, results support the use of OML as a criterion to validate personality scale scores and other metrics that are used to award ROTC scholarships to those individuals who are most likely to become high performing U.S. Army officers. Results also carry implications for refining algorithms used by USACC to compute the Cadet OML score.

CADET TRAINING AND PERSONALITY METRICS LONGITUDINALLY PREDICT OFFICER IN-UNIT PERFORMANCE: R = .37

CONTENTS

LIST OF TABLES

CADET TRAINING AND PERSONALITY METRICS LONGITUDINALLY PREDICT OFFICER IN-UNIT PERFORMANCE: *R* = .37

Introduction

Personality metrics are widely used for personnel selection applications because they have modest predictive validity against occupational performance outcomes, yet carry minimal adverse impact (Hogan, 2005; Hough & Oswald, 2000; Ones & Anderson, 2002). In addition, job knowledge is known to be a strong predictor of job performance (Hunter, 1986), and the U.S. Army often uses training performance metrics to inform personnel assignment decisions to help ensure that high-performing Soldiers will be assigned to critical positions. Consistent with these perspectives, the U.S. Army is evaluating the use of personality measures to identity individuals who are likely to excel in Reserve Officer Training Corps (ROTC) programs, perform well as commissioned officers, and pursue long-term military careers (Putka et al., 2009).

As part of the ROTC program, the U.S. Army Cadet Command (USACC) uses pre-commissioning data to develop Cadet Order of Merit List (OML) scores. The Cadet OML metric reflects approximately 20 discrete sources of information including: college grade point average, military science grade point average, multiple physical fitness indicators, professorial ratings of leadership potential, and ratings of leadership performance and potential in military training exercises. However, USACC has modified the computation of the Overall OML score during the past decade, and therefore it is not possible to analyze all the individual OML variables in a multi-cohort database such as the one used for our analyses. Nonetheless, Cadet OML scores reflect three principle components, Academic OML, Leadership OML, and Fitness OML, and these values can be estimated for most of the officers in our current dataset.

Cadet OML score is critical to an individual's career trajectory because the U.S. Army uses this metric to identify newly commissioned officers for active duty and assign these individuals to critical positions. In addition, the Army is actively investigating the use of personality measures to award ROTC scholarships to individuals who are likely to perform well in ROTC pre-commissioning programs and obtain a superior Cadet OML score. This is because of implicit expectations that the Cadet OML score predicts officer tendencies to perform well in the military (Putka et al., 2009). To evaluate the capacity of personality measures to predict Cadet OML, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has collected personality data from most ROTC cadets during summer training since 2013.

While analyses show that personality data predict Cadet OML (Legree, Kilcullen, Putka & Wasko, 2014), the capacity of training and personality measures to predict occupational performance over an extended timespan has not been carefully evaluated in the military for officer populations. To address this issue, we used a longitudinal design to validate ROTC pre-commissioning and personality metrics against supervisor ratings of officer in-unit performance that had been collected several years after the cadets had been assigned to operational units..

It is important that the validity of OML scores against officer performance metrics be evaluated to justify use of ROTC performance measures (i.e., Cadet OML score) to: (a) assign higher scoring cadets to the active duty component and critical positions, and (b) use Cadet OML

score as an outcome to validate and use personality scales and other metrics for the award of ROTC scholarships. It also follows that the results of these validity analyses are relevant to the continued use of training and personality metrics for personnel selection applications. To explore the potential of personality scales to predict in-unit performance outcomes, we also evaluated the use of profile similarity metrics (PSMs) to improve the validity of personality scales. When interpreting our findings, it is also relevant that personality scales developed to predict ROTC program continuance, as opposed to ROTC program performance, are less likely to predict in-unit performance outcomes.

Distance Scores versus Profile Similarity Metrics (PSMs) for Personality Scales

Most rating-based personality scales use a conventional scoring algorithm that computes a respondent's scale score as the mean of the item rating with ratings for reversed items corrected for item direction. Table 1 contains examples of reversed and non-reversed items and illustrates the computation of conventional and distance scores for these items. Inspection of Table 1 shows that there is a perfect negative correlation between conventional and distance item scores, $r = -1$. Therefore, conventional and distance scores are redundant, and conventional personality scores constitute transformed distance metrics.

Table 1
Conventional and Distance Scoring Algorithms for Non-reversed and Reversed Items

Example Items	Conventional Scoring		Distance Scoring			Sum
	Respondent Rating	Corrected Rating	Respondent Rating	Key	Distance Score	
Non-reversed Item: *In my free time, I enjoy physical activities:*						
	1. Never; 2. Seldom; 3. Occasionally; 4. Frequently; 5. Often.					
Item 1 (Non-reversed)	4	4	4	5	1	5
Item 2 (Non-reversed)	5	5	5	5	0	5
Item 3 (Non-reversed)	3	3	3	5	2	5
Reversed Item: *While in school, I tend to avoid sports:*						
	1. Never; 2. Seldom; 3. Occasionally; 4. Frequently; 5. Often					
Item 4 (Reversed)	2	4	2	1	1	5
Item 5 (Reversed)	1	5	1	1	0	5
Item 6 (Reversed)	2	4	2	1	1	5
Scale Score (Item Mean):	4.17		0.83			5.0

While distance scores may be intended to quantify the overall "match" between a scoring key and a respondent's ratings, distance scores can be highly influenced by examinee tendencies to elevate or depress their ratings relative to most respondents (i.e., by using one end of the scale), or to use more or less of the available rating scale than most respondents. These tendencies can act to limit the validity of distance scores that are computed for personality scales. This observation has led to the proposal that profile similarity metrics (PSMs) may be used to increment the validity of distance scores computed for personality scales that contain a mix of reversed and non-reversed items (e.g., battery analyses using PSMs resulted in higher validity estimates than the use of distance scores, $R = .54$ vs $R = .47$; Legree, Ness, Kilcullen & Koch, 2019).

PSMs are important because they allow distance scores to be partitioned into their component parts that account for their underlying variance (e.g., the tendency to inflate ratings may diminish the validity of distance scores). From a theoretical perspective, PSMs reflect the formulaic demonstration that distance scores represent separate sources of variance that may be used to model distance score variance for personality scales.

Accordingly, we computed the following PSMs for each personality scale that contained a mix of reversed items:

PSM 1. Rating scatter, which is computed as the variance of each respondent's rating profile. Scatter scores = sd_x^2.

PSM 2. Rating elevation difference, which equals the squared difference between each respondent's mean scale rating and the mean value in the scale key: Elevation-Difference scores = $(X_{mean} - K_{mean})^2$.

PSM 3. Rating elevation equals the respondent's mean item rating, X_{mean} elevation difference, which equals the squared difference between each respondent's mean scale rating and the mean value in the scale key: Elevation-Difference scores = $(X_{mean} - K_{mean})^2$.

PSM 4. Rating shape conventional, which equals the correlation between each respondent's rating profile and the conventional scale key. Shape-Conventional scores = $r_{x,\text{conventional key}}$.

PSM 5. Rating shape consensus equals the correlation between each respondent's rating profile and the mean response value for each scale item. Shape-Consensus = $r_{x,\text{consensus key}}$.

Using large datasets ($n \approx 5000$), these PSMs have been analyzed to evaluate the possibility that conventional scores for personality scales may underestimate the validity of the personality scales against the Cadet OML metric (Legree et al., 2019). Those earlier analyses demonstrated that PSMs provide incremental validity beyond conventional distance scores against the Cadet OML metric that were both substantial and stable upon cross validation. For this project, we conducted additional analyses to determine if PSMs would also provide gains beyond the conventional personality scores against supervisor ratings of officer in-unit performance that were collected several years after the personality data had been collected.

Based on earlier results (Legree et al., 2019), we computed shape scores using either the consensus or the conventional key for each personality scale.

Research Hypotheses

Optimizing Cadet OML validity. As described above, the Cadet OML score computation has been repeatedly modified by USACC based on subject matter expert (SME) opinion. Alternatively, empirical weighting may allow better predictive validity by optimally weighting the various sources of information that incorporated to produce the Cadet OML score. Based on synthetic validity findings that SMEs can estimate the validity of predictors only modestly well in comparison to empirical analyses (Oswald & Hough, 2010), we propose the first two hypotheses:

Hyp 1. The Cadet OML score will be a valid predictor of officer performance outcomes using a longitudinal design.

Hyp 2. The OML component scores (i.e., Academic OML, Leadership OML, and Fitness OML) will add incremental validity to Cadet OML score against officer performance outcomes using a longitudinal design.

Results for Hypotheses 1 and 2 are critical for justifying (a) the use of the Cadet OML score to assign newly commissioned officers to critical positions, and (b) the use personality scale scores to predict Cadet OML score and support ROTC scholarship award. An important implication of Hypothesis 2 is that Cadet OML score represents suboptimal weighting of the OML Academic, OML Fitness, and OML Leadership component scores. Therefore, these results may carry implications for refining the algorithm used by USACC to compute the Cadet OML score.

Profile Similarity Metrics (PSMs). Past analyses demonstrate that PSMs provide incremental validity beyond distance scores computed for individual personality scales against Cadet OML score and that these results are stable upon cross validation (Legree, Ness, Kilcullen & Koch, 2019). Based on the expectation that Cadet OML will predict officer performance outcomes (Putka et al., 2009), we extend this reasoning to propose:

Hyp 3. Personality scale data collected for recruits will predict long-term performance outcomes.

Hyp 4. Optimizing PSMs for individual personality scales against Cadet OML score will improve the predictive validity of personality scales against long-term performance outcomes.

Endorsement of Hypotheses 3 and 4 supports the use of personality scales and PSMs to award ROTC scholarships.

Predicting officer performance. Finally, we propose that the use of PSMs and personality scales will add incremental validity to the OML metrics for the prediction of long-term officer performance.

Hyp 5. Personality conventional scales add incremental validity over and above OML metrics against officer performance rating.

Hyp 6. Personality PSMs add incremental validity over and above OML and conventional personality metrics against officer performance ratings.

Endorsement of Hypotheses 5 and 6 carries implications for improving the Army's capacity for the prediction of officer performance by using a broader array of pre-commissioning metrics.

Current Research

We conducted six sets of analysis to evaluate the above hypotheses and validate the cadet pre-commissioning training and personality data against supervisor ratings of officer performance. We used standard regression procedures to evaluate the predictive capacity of the Cadet OML and conventional personality metrics to predict the officer in-unit performance outcome and assess Hypotheses 1, 2, 3 and 5.

As described above, Hypotheses 4 and 6 addressed the utility of using PSMs to enhance the validity of conventional personality scale scores. For these analyses, we wanted to evaluate the capacity of these metrics to predict the officer performance ratings. However, the longitudinal nature of our dataset resulted in a much smaller sample size for several of the personality scales (Written Communication n = 605; Oral Communication n =586) than has been used to evaluate PSMs against the Cadet OML metric ($n \approx 5000$; Legree et al., 2019). This constraint resulted in the sample sizes being lower for the multiple regression models used to evaluate Hypothesis 4 (n =578) and Hypothesis 6 (n = 533). In order to conserve degrees of freedom for these analyses, we first regressed Cadet OML onto the scatter, delta and elevation metrics as well as shape scores computed using either the consensus or the conventional key for each personality scale (key choice was based on Legree et al., 2019). This procedure optimized the PSMs for predicting OML scores. We then evaluated the incremental validity of the Optimized PSM scale scores beyond OML and the conventional personality scores against the officer in-unit performance rating criterion.

Method

Participants

The analysis dataset contains predictor and criterion data for 1068 junior officers who provided permission for us to collect: (a) confidential supervisor ratings of their current job performance; (b) pre-commissioning training data from which the OML overall and component scores were computed, and (c) personality data collected from the individuals when they participated in the ROTC Summer Advanced Camp, a 9-week field training exercise just prior to the cadet's senior year. However, changes to the personality battery and incomplete response sets limited the sample size for individual analyses. The junior officers included 723 Lieutenants and 341 Captains. The sample was primarily male, 86%. These individuals are recorded as having a primary ethnic designation of: Caucasian, 79%; Hispanic, 8%; African-American, 6%; Asian or Pacific Islander, 4%; and American Indian, 1%. In addition, 3% of the sample identified a primary ethnic designation as a group other than those listed. These individuals had served for

an extended period in the military as commissioned officers when their supervisors provided ratings of their in-unit performance as junior officers (Mean = 4.32 years, SD = 1.58 years).

Measures

Supervisor performance ratings. Supervisor ratings of officer performance were collected for the following eight performance dimensions that were based on job analysis data (Paullin et al., 2014):

1. Branch Specific Technical & Tactical Duties;
2. Writing Performance;
3. Public Speaking Performance;
4. Self-control and Personal Discipline;
5. Physical Fitness;
6. Organizational ability;
7. Cross-cultural Performance;
8. Innovation.

The Overall Supervisor Performance Rating metric, our principle outcome measure, was computed by averaging the supervisor ratings collected across the eight highly correlated dimensions.[1] Performance ratings for each dimension were collected on a 7-point scale to increase rating variance and minimize the presence of ceiling effects. The 7-point rating scale contained anchors that ranged from "well below average" to "truly exceptional."

Order of Merit List (OML) scores. The Cadet OML metric and associated scores were provided by the U.S. Army Cadet Command (USACC) for Army officers who completed ROTC and became commissioned officers from June 2011 through June 2017. While the Cadet OML score represents data from numerous sources, we restricted our analyses to the overall Cadet OML and the OML component scores because the algorithm used to compute the OML scores had been repeatedly modified across these cohorts during the past decade. The OML overall scores and its principal components (Academic, Fitness, and Leadership) represent our primary predictor variables due to their centrality to the military's personnel assignment system.

Personality scale scores. Nearly all officers commissioned through ROTC participate in ROTC Summer Advanced Camp approximately nine-months before they become commissioned officers (Legree et. al., 2014; Legree et al., 2019; Wasko, Putka, Legree & Kilcullen, 2019). Since 2012, most cadets have completed a personality battery during this exercise. The personality scores have since been validated against Cadet OML scores that were subsequently obtained. Data for the following scales were available for most of the officer sample: Achievement Orientation, Army Identification, Fitness Motivation, Hostility to Authority, Peer Leadership, General Self-Efficacy, Stress Tolerance, Oral Communication, Written Communication, and Social Desirability (Lie scale).

All items within the personality battery used a 5-point Likert scale, and each scale was conventionally scored. In addition, each scale that contained reversed items (Army Identification, Fitness Motivation, Oral Communication, Written Communication, and Social

[1] Factoring the eight performance rating dimensions resulted in a single dimension accounting for 73.3% of the variance ($\lambda = 5.86$); These factor scores were nearly redundant with the average supervisor rating, $r = .994$. As additional data are collected, analyses will be conducted to more closely compare findings across data collection methods and subsamples.

Desirability) was scored using both PSMs and conventional methods. Table 2 describes CBEF scale constructs.

Table 2
CBEF Scales and Definitions

Scale	Definition
Achievement	The willingness to give one's best effort and to work hard towards achieving difficult objectives.
Army Identification	Identification with, and interest in being, a U.S. Army Soldier.
Fitness Motivation	Enjoyment from physical exercise and willingness to stay physically fit.
General Self-Efficacy	Feeling that one has successfully overcome past work obstacles.
Hostility to Authority	Suspicious of the motives and actions of legitimate authority figures. Views rules and directives from authority as illegitimate.
Peer Leadership	Seeks positions of authority. Comfortable with being in charge of a group and accepts responsibility for the group's performance.
Stress Tolerance	Degree of emotional control and composure under pressure.
Tolerance for Injury	Degree of enjoyment from risky and hazardous activities.
Written Communication	Degree of comfort with written communication.
Oral Communication	Degree of comfort with oral communication.
Social Desirability (Lie)	Degree of socially desirable responding.

Procedure

As described above, the participants completed a paper questionnaire that contained the personality measures when they participated in the ROTC Summer Advanced Camp as cadets. Cadet OML scores were computed by USACC and obtained for research purposes.

Electronic and in-person procedures were used to collect supervisory performance rating data. For electronic data collection, Army datasets were mined to identify supervisors for individual officers. These supervisors were then emailed a link and requested to provide performance ratings for each of the eight job dimensions listed above. For in-person data collection, unit rosters were used to identify supervisors for individual officers. These supervisors were requested to provide performance ratings for each of the job dimensions using a paper questionnaire.

Results

Descriptive Statistics

Table 3 reports descriptive statistics and bivariate validities for the OML variables and the conventional personality scale metrics against the supervisor ratings of officer in-unit performance.

Research Hypotheses

Cadet OML hypotheses. The first two hypotheses were designed to investigate the validity of the Cadet OML score and evaluate the possibility that its components might be reweighted to improve its validity. Assessing these hypotheses required both correlational and regression analyses.

Hypothesis 1. Hypothesis 1 proposed that the Cadet OML score is a significant predictor of the supervisor performance rating outcome. As reported in Table 3, the correlational analyses supported Hypothesis 1, $r = .31$, $p < .001$. This result helps justify the use of the Cadet OML score to assign newly commissioned officers to the active duty component and to assign individual officers to critical positions. In addition, this result broadly supports the operational use of personality measures that have been validated against Cadet OML to award ROTC scholarships (Young, Kilcullen, Legree & Puente, 2018).

Table 3

Descriptive Statistics and Bivariate Validities Against Supervisor Performance Ratings

Scale	Mean	SD	*n*	*r*
Officer In-Unit Performance Outcome				
Supervisor Performance Rating	4.83	1.21	1068	
ROTC OML Variables				
Cadet OML	0.00	1.00	767	.31***
Academic OML	0.00	1.94	772	.05
Leadership OML	0.00	3.69	772	.25***
Fitness OML	0.00	2.54	746	.09*
Personality Measures				
Achievement	4.15	0.49	1068	.09**
Army Identification	4.03	0.54	1068	.01
Fitness Motivation	3.96	0.62	1068	.07*
General Self-Efficacy	4.47	0.40	1031	.11***
Hostility to Authority	2.30	0.51	1068	-.03
Peer Leadership	3.80	0.60	1051	.12***
Stress Tolerance	3.19	0.51	1068	.05
Tolerance for Injury	3.72	0.67	1051	.00
Written Communication	3.33	0.66	605	.05
Oral Communication	4.03	0.38	585	.15***
Social Desirability (Lie)	0.00	0.12	1068	-.01

*$p < .05$, 2-tailed. **$p < .01$, 2-tailed. ***$p < .001$, 2-tailed.

Hypothesis 2. Hypothesis 2 proposed that the OML Academic, OML Leadership, and OML Physical Fitness component scores would add incremental validity to the Cadet OML score against the officer in-unit performance rating outcome. We used a hierarchical regression

procedure to assess Hypothesis 2. In step 1, we regressed the supervisor performance rating outcome onto Overall OML score, (R = .30, p < .001). We then added the OML component scores in step 2 to assess the second hypothesis (R =.33, ΔR^2 = .017, p < .003). This result confirmed Hypothesis 2 and indicates that the Cadet OML score could be reweighted to modestly improve its long-term predictive validity. See Table 4 for details.

Table 4
Incremental Validity for Cadet OML and OML Component Scores Against Supervisor Performance Ratings

				Change Statistics			
Step	R	R^2	Adj R^2	ΔR^2	F	df(1,2)	Sig.
1. Cadet OML	.30	.090	.089	.090	73.192	1,739	.001
2. OML Components	.33	.107	.103	.017	4.755	3,736	.003

Personality Hypotheses. The second two hypotheses were designed to explore the predictive validity of personality data collected from cadets who were attending a training exercise against ratings of their in-unit performance that were collected several years later. We also investigated the possibility that the use of PSMs would enhance the validity of these personality scales. Assessing these two hypotheses required both correlational and hierarchical regression analyses.

Hypothesis 3. Hypothesis 3 proposed that supervisor ratings of officer in-unit performance would be predicted by conventional scores using personality data collected when these officers were ROTC cadets. As detailed in Table 2, the bivariate correlations indicated that conventional scores for 5 of the 11 personality scales were valid predictors of officer performance. More importantly, the regression analyses indicated a modest level of validity for the personality scales against the officer performance outcome: R = .23, p < .001. The validity correlations are reported in Table 3, and the regression results are reported in Table 5, step 1.

Table 5
Validity for Personality Conventional Scores and PSMs Against Supervisor Performance Ratings

				Change Statistics			
Step	R	R^2	Adj R^2	ΔR^2	F	df(1,2)	Sig
1. Conventional Scores	.23	.052	.034	.052	2.847	11,566	.001
2. PSMs	.28	.076	.048	.024	2.378	6,560	.028

Hypothesis 4. Hypothesis 4 proposed that the validity of personality scales against officer performance outcomes might be improved through the use of PSMs to score the six personality scales that contained a mix of reversed items. To conduct these analyses, we first regressed the Cadet OML measure onto four PSMs for each scale that contained reversed items and saved these PSM scores. The four PSMs corresponded to the scatter, delta, and elevation metrics, as well as shape scores computed using either the consensus or the consensus keys. We then used the PSM scale scores that had been computed against Cadet OML to assess Hypothesis 4 through correlation and regression analyses (cf., Legree et al., 2019).

The bivariate correlations indicated that while the conventional scores for two of the six personality scales significantly predicted the officer performance outcome, three scales significantly predicted officer performance when scored using PSMs. The largest validity gain was obtained for the Written Communication scale (r_{PSM} = .15 vs $r_{Conventional}$ = .05). Moreover, the difference between the two validity estimates for the Written Communication scale is significant according to the Steigler (1980) procedure designed to test the difference between these two correlations computed for a common sample (z = 2.373, $p < .05$). The gain in Written Communication primarily reflects increasing the weight of shape scores that were computed using an alternate scoring key. When interpreting results, it is relevant that the Army Identification and Social Desirability scales are used primarily to predict ROTC continuance, not ROTC performance. The bivariate validities for these personality scales are reported in Table 6.

Table 6
Comparison of Personality Scale Validity Against Supervisor Ratings Using PSMs or Conventional Scores

Scale	PSM Score[1]	Conventional Score
Written Communication	.13**	.05
Fitness Motivation	.09**	.07*
Oral Communication	.15***	.16**
Tolerance for Injury	.04	.00
Army Identification	-.04	.01
Social Desirability	-.03	-.01

[1]PSM scale scores computed as optimally weighted composites against Cadet OML.
*$p < .05$ level (2-tailed). **$p < .01$ level (2-tailed). ***$p < .001$ level (2-tailed).

As expected, the regression analyses indicated that the PSMs provided incremental validity beyond the conventional personality scale scores: R = .28, ΔR^2 = .024, $p < .05$. The regression analyses are reported in Table 5, step 2.

OML and Personality hypotheses. The final two hypotheses were designed to explore the predictive validity of combining the OML and personality metrics to predict the officer performance outcome. Although this hierarchical regression model represents a conceptual extension of the model used to assess Hypothesis 2 (Table 4), we recomputed all steps of this model due to missing data to evaluate Hypotheses 5 and 6 (Table 7).

Hypothesis 5. Hypothesis 5 proposed that conventional personality scale scores would add incremental validity to the OML metrics against the officer performance rating outcome. The first two steps of the hierarchical regression model provide the base model to test this

hypothesis: Significant results were obtain for Cadet OML (R = .24, p = .001, step 1) and the OML components (R =.28, ΔR^2 = .009, p < .007, step 2). To evaluate Hypothesis 5, we added the conventional personality scale scores (R =.33, ΔR^2 = .109, p = .129). Therefore, Hypothesis 5 was not supported.

Table 7
Validity for Cadet OML, Components and Personality Scores Against Officer Performance Ratings

				Change Statistics			
Step	R	R^2	Adj R^2	ΔR^2	F	df(1,2)	Sig
1. Cadet OML	.24	.050	.057	.059	33.438	1,531	.001
2. OML Components	.28	.081	.074	.021	4.115	3,528	.007
3. Conventional Personality	.33	.109	.083	.028	1.497	11,517	.129
4. PSMs Personality	.37	.137	.102	.028	2.798	6,511	.011

Hypothesis 6. Hypothesis 6 proposed that using PSMs to score the personality scales would add incremental validity to Cadet OML scores and the conventional personality scale scores against the officer performance rating outcome. To evaluate Hypothesis 6, we added the PSM personality scale scores (R =.37, ΔR^2 = .137, p = .011). Therefore, Hypothesis 6 was supported and suggests that a refined set of personality scales and scoring algorithms adds incremental validity to the traditional Cadet OML metrics for the prediction of the officer in-unit performance rating outcome.

Discussion

Implications

The most important results of this research correspond to the demonstrations that: (1) the Cadet OML metric is a valid predictor of supervisor ratings of officer performance that were collected one to eight years after the cadets had become commissioned officers, $R = .30$; and (2) reweighting the OML components would significantly increase the operational validity of the OML metric, $R = .33$. Moreover, these values are likely lower-bound validity estimates due to restriction of range as well as the impact of assigning officers to a wide-range of occupations. Therefore, we conclude that the Cadet OML metrics is a modest to strong predictor of subsequent officer performance.

The analyses also showed that a variety of personality scales were significant predictors of officer performance data that were collected on average 4.3 years[2] after the cadets became commissioned officers, $R = .28$. Combining the Cadet OML and the PSM-based scores for the personality data resulted in a significantly higher validity than was observed when only the Cadet OML data were analyzed, $R = .37$. This final result demonstrates that ROTC training and Cadet personality data are modestly predictive of future officer in-unit performance.

These results support the continued use of personality data to help USACC award ROTC scholarships to those individuals who are most likely to become high performing officers when assigned to operational units after becoming commissioned officers (Young et al., 2018). These results also suggest that these metrics could be used to improve the officer branch assignment process for newly commissioned officers.

Future Directions

From a statistical perspective, the generality of our results was primarily limited by the difficulty in collecting supervisor performance ratings of officer in-unit performance and the limited sample sizes that were therefore available to evaluate theoretical expectations for the personality scales and the associated PSMs. As described in the introduction, we addressed this limitation by optimizing the PSMs for the personality scales against the available Cadet OML scores, and then using the PSM personality scale scores to evaluate Hypotheses 4 and 6. We expect that somewhat stronger results would be obtained if the personality PSMs had been directly optimized on the supervisor performance ratings. Therefore, the potential validity of the PSMs that were computed for the personality scales may have been underestimated by these analyses.

In addition, the personality data were collected during the ROTC Summer Advanced Camp for scales that were selected to predict ROTC program continuance and cadet performance, as opposed to predicting officer in-unit performance (Legree et al., 2019). While there is substantial conceptual overlap between the scales that are likely to predict cadet and officer performance metrics, we also expect that the overlap would be lower between those scales that would be likely to predict officer performance versus those that were chosen to

[2] Although the total time lag was approximately five years because the personality data were collected approximately nine months before the cadets became commissioned officers.

predict ROTC cadet program continuance (e.g., Army Identification and Social Desirability). This observation has two ramifications. First, the inferential statistics for the personality scales that are reported in Tables 4 and 6 represent lower bounds on the results that would be obtained had the personality scales been chosen specifically to predict officer in-unit performance (i.e., the estimates for the inferential statistics were likely diminished by the inclusion of scales chosen to predict program continuance). Second, the analyses did not include a larger variety of personality scales that might have been selected to more optimally predict officer performance. Therefore, the overall predictive validity of the personality scales against the officer in-unit performance outcome is likely underestimates the potential validity of personality scales to predict officer in-unit performance. These analyses will be updated as additional data become available.

References

Hogan, R. (2005). In defense of personality measurement: Old wine for new whiners. *Human Performance, 18*, 331-341.

Hough, L. M., & Oswald, F. (2000). Personnel selection: Looking toward the future-remembering the past. *Annual Review of Psychology, 51*, 631-664.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge and job performance. *Journal of Vocational Behavior, 29*, 340-362.

Legree, P. J., Kilcullen, R. N., Putka, D. J., & Wasko, L. E. (2014). Identifying the Leaders of Tomorrow: Validating Predictors of Leader Performance. *Military Psychology, 26*, 292-309.

Legree, P. J., Ness, A. M., Kilcullen, R. N., Koch, A. J. (2019). *Enhancing the Validity of Rating-Based Tests.* (ARI Technical Report 1371). Fort Belvoir, VA: U.S. Army Research Institute.

Ones, D. S., & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organizational Psychology*, *75*, 255-276.

Oswald, F. L., & Hough, L. M. (2010). Validity in a jiffy: How synthetic validation contributes to personnel selection. *Industrial and Organizational Psychology*, *3*, 329-334.

Paullin, C., Sinclair, A. L., Moriarty, K. O., Campbell, R. C., Legree, P. J., Kilcullen, R. N. (2014). Delineating Officer Performance and its Determinants. *Military Psychology, 26,* 259-277.

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

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251. Retrieved at http://quantpsy.org/corrtest/corrtest2.htm.

Young, M. C., Kilcullen, R. N., Legree, P. J., & Puente, K. (2018, October). *Enhancing the Selection of 4-Year Army ROTC Scholarship Recipients*. Paper presented at the 60th Conference of the International Military Testing Association, Kingston, Canada.

Wasko, L. E., Putka, D. J., Legree, P. J., & Kilcullen, R. N. (2019). *Validation of Measures for Predicting Leader Development and Assessment Course Performance*. (ARI Technical Report 1375). Fort Belvoir, VA: U.S. Army Research Institute.