Technical Report 1282

Assessment of Assembling Objects (AO) for Improving Predictive Performance of the Armed Forces Qualification Test

Lance Anderson
ICF International

Richard R. Hoffman III and Brian Tate
U.S. Army Research Institute

Jessica Jenkins, Carolyn Parish, and Alicia Stachowski
ICF International

J. Douglas Dressel
U.S. Army Research Institute

April 2011

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 for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

ICF International

Technical review by

Tonia Heffner, U.S. Army Research Institute
Nehama Babin, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(dd-mm-yy)*	2. REPORT TYPE	3. DATES COVERED *(from - to)*
April 2011	Final	Nov 2009 - Oct 2010

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Assessment of Assembling Objects (AO) for Improving Predictive Performance of the Armed Forces Qualification Test	91WAW-09-D-0016, DO 0002
	5b. PROGRAM ELEMENT NUMBER 665803

6. AUTHOR(S)	5c. PROJECT NUMBER
Lance Anderson (ICF International); Richard Hoffman and Brian Tate (U.S. Army Research Institute for the Behavioral and Social Sciences); Jessica Jenkins, Carolyn Parish, and Alicia Stachowski (ICF International); J. Douglas Dressel (U.S. Army Research Institute for the Behavioral and Social Sciences)	D730
	5d. TASK NUMBER 377
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
ICF International 9300 Lee Highway Fairfax VA, 22031	029625.0.002.00

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
United States Army Research Institute ATTN: DAPE-ARI-RS 2511 Jefferson Davis Highway Arlington, Virginia22202-3926	ARI
	11. SPONSORING/MONITORING AGENCY REPORT NUMBER Technical Report 1282

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Richard Hoffman

14. ABSTRACT *(Maximum 200 words):*

The purpose of this investigation was to examine whether the Assembling Objects test (AO) should be added to the AFQT predictor composite. The investigation included a literature review and analysis of data from Army Classification project. Analyses included regressions with 20 different criteria, subgroup analyses, and bias analyses. Findings suggest that the Army should consider including the AO subtest in the AFQT predictor composite. Adding the AO subtest to the AFQT composite would likely increase the prediction of performance and job knowledge, while potentially decreasing the subgroup predictor differences between Hispanic and white subgroups. Although adding the AO subtest to the AFQT composite may increase the subgroup predictor differences of female and black subgroups relative to their respective majority comparison groups, results suggest that any additional subgroup differences would be borne out in performance differences, and that the revised AFQT composite would thus be fair and unbiased to minority groups. Future research should examine the degree to which these findings differ across job clusters, where cluster members are similar in terms of abilities needed to perform the work.

15. SUBJECT TERMS

Assembling Objects, Armed Forces Qualifying Test, performance, prediction, test, spatial ability

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			Ellen Kinzer Technical Publication Specialist 703-545-2445
Unclassified	Unclassified	Unclassified	Unlimited	79	

Technical Report 1282

Assessment of Assembling Objects (AO) for Improving Predictive Performance of the Armed Forces Qualification Test

Lance Anderson
ICF International

Richard Hoffman and Brian Tate
U.S. Army Research Institute

Jessica Jenkins, Carolyn Parish, and Alicia Stachowski
ICF International

J. Douglas Dressel
U.S. Army Research Institute

Personnel Assessment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

April 2011

Army Project Number
665803D730

Personnel and Training Analysis Activities

ACKNOWLEDGEMENTS

The researchers would like to thank Dr. Raymond Morath of ICF International for his comments on an earlier draft of this report. We also thank the Soldiers, Non-commissioned Officers, Officers, and Army Civilians who participated in this research and contributed to making the data collection possible.

ASSESSMENT OF ASSEMBLING OBJECTS (AO) FOR IMPROVING PREDICTIVE PERFORMANCE OF THE ARMED FORCES QUALIFICATION TEST

EXECUTIVE SUMMARY

Research Requirement:

The purpose of the present investigation was to determine whether adding the Assembling Objects (AO) subtest of the Armed Services Vocational Aptitude Battery (ASVAB) to the Armed Forces Qualification Test (AFQT) would improve the AFQT's validity as an applicant screening tool. This research was undertaken by the U.S. Army Research Institute for the Behavioral Social Sciences (ARI) to inform Army decisions regarding potential revisions to the use of the AFQT as a selection and assignment tool. This research was intended to support the Army's goal of maintaining an applicant-screening program that is valid and supports an army that is representative of the U.S. population.

The AFQT is used by the Armed Services to determine enlistment eligibility, and is calculated from a set of four subtests of the ASVAB – Work Knowledge, Paragraph Comprehension, Mathematics Knowledge, and Arithmetic Reasoning. Currently, the AFQT score reflects two components of general aptitude, mathematical aptitude and verbal aptitude, with two subtests measuring each component. Past empirical research on a spatial subtest of the ASVAB, AO, indicates that this subtest may be an excellent candidate for addition to the AFQT. The present research defines AO's utility in terms of the extent to which it contributes to the prediction of Soldier performance and attrition criteria over AFQT scores as well as the extent to which AO demonstrates any subgroup differences.

Procedure:

The investigation included a review of literature on the AO subtest, and analyses on extant data to determine whether the AO subtest could aid the existing AFQT configuration in predicting performance and other criteria.

We drew the sample for this investigation from a database developed for the *Validating Future Force Performance Measures (Army Class)* research project. The primary purpose of the Army Class project is to provide the Army with recommendations on the predictor measures that demonstrate the greatest potential to inform entry-level Soldier selection and classification decisions (Moriarty, Campbell, Heffner, & Knapp, 2009). The current investigation focused on criteria gathered at two points in time:

- Initial Entry Training (IET) including either one-station unit training (OSUT) or advanced individual training (AIT) with data collection beginning in the fall of 2007 and continuing through the summer of 2008.
- In-Unit (IU), targeting the same sample of Soldiers at about 18-20 months time in service (TIS).

We identified the following variables as predictors for the regression analyses: the Assembling Objects (AO) subtest and the four subtests used to form the AFQT composite [Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), and Mathematics Knowledge (MK)].

We identified or calculated 20 criterion measures for use in the investigation. These 20 measures covered four types of criteria: performance ratings, job knowledge, attitudes, and attrition.

We conducted analyses to evaluate and clean the data, identify or calculate criterion variables, evaluate Supplemental predictor composites (based on adding AO to the existing AFQT composition), and evaluate Substitutional predictor composites (based on replacing AFQT subtests with the AO). We evaluated the different types of predictor composites by comparing them to the Baseline composite consisting of the four subtests currently used to form the AFQT. We compared the composites according to their relationships with each criterion, the impact on subgroup differences, and the level of bias or differential prediction associated with the composite.

Findings:

The findings showed that adding the AO subtest to the current subtests included in the AFQT improved prediction of job knowledge and performance ratings and had little or no impact on subgroup differences. Bias analyses revealed that adding the AO subtest had no effect on the level of differential prediction for any of the subgroup comparisons.

Substituting the AO subtest for any one of the subtests used to form the AFQT improved or had a negligible impact on prediction of 19 out of 20 criterion scores examined. Substituting the AO subtest for any of the other subtests led to mixed results in regard to subgroup differences, and had no impact on the level of differential prediction for any of the subgroup comparisons.

Utilization and Dissemination of Findings:

The Army may wish to consider including the AO subtest in the AFQT predictor composite. Adding the AO subtest to the AFQT composite would likely increase the prediction of performance and job knowledge. Although our analyses showed an increase in the differences among subgroups when the AO subtest was added to the AFQT composite, this finding should be interpreted with caution. The incumbent sample used in this investigation is affected by differential range restriction across subgroups. That is, the variance of the AFQT scores is reduced (or restricted) in the incumbent sample (relative to the applicant population) and the degree of that range restriction differs across subgroups. This makes it difficult to predict applicant subgroup differences based on incumbent data. Thus, we recommend that future investigations examine subgroup differences using an applicant sample. It is interesting to note, however, that even when we use this incumbent sample to examine the impact of any additional subgroup differences, we found that those differences were reflected in performance and job

knowledge criteria, suggesting that the revised AFQT composite would be fair and unbiased to minority groups.

Using the AO subtest as a supplement to the current AFQT composite, or as a replacement for one of the subtests would likely provide a greater improvement in the prediction of job performance for some jobs than it would for others. In theory, the jobs that are most likely to see improved prediction with the addition of the AO subtest are jobs whose tasks require spatial aptitude (e.g., Light Wheel Vehicle Mechanic [91B]). To determine whether spatial aptitude requirements moderate the predictive power of the AFQT, we suggest conducting a series of moderated multiple regressions. The moderator in these regressions could be defined by coding each job for the amount of spatial aptitude required. This research might also examine how prediction of technical training is differentially affected in these types of jobs. Documentation of improved prediction in technical training failure/attrition would provide great benefit to the Army in terms of improved assignment and reductions in Soldier attrition.

Future research should also systematically examine the financial utility of adding the AO subtest to the AFQT. Changes in criterion outcomes and subgroup differences could be subjected to economic modeling to help better understand the overall impact of adding the AO subtest. This research might also consider the impact of combining the subtest scores in a non-compensatory manner.

ASSESSMENT OF ASSEMBLING OBJECTS (AO) FOR IMPROVING PREDICTIVE PERFORMANCE OF THE ARMED FORCES QUALIFICATION TEST

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

APPENDIX FIGURES

ASSESSMENT OF ASSEMBLING OBJECTS (AO) FOR IMPROVING PREDICTIVE PERFORMANCE OF THE ARMED FORCES QUALIFICATION TEST

CHAPTER 1: INTRODUCTION

Project Background

The Armed Forces Qualification Test (AFQT) is used by the Armed Services to determine enlistment eligibility, and is calculated from a set of four subtests of the Armed Services Vocational and Aptitude Battery (ASVAB) including Paragraph Comprehension, Work Knowledge, Arithmetic Reasoning, and Mathematics Knowledge. Currently, the AFQT score reflects two components of general aptitude, mathematical aptitude and verbal aptitude, with two subtests measuring each component. The goal of the present research is to assess the potential benefits of adding a subtest of spatial aptitude to the AFQT, either as a replacement to one or more existing subtests or as a supplement to the four existing subtests. Past empirical research on a spatial subtest of the ASVAB, Assembling Objects (AO), indicates that this subtest may be an excellent candidate for addition to the AFQT. For example, Alderton, Wolfe, and Larson (1997), Busciglio (1990) and Wolfe (1997) have found that adding the AO subtest to the AFQT may aid in predicting personnel performance and might reduce adverse impact.

The investigation included a review of the literature on the AO subtest and analyses on extant data to determine whether the AO subtest could aid in predicting performance. This report provides a thorough documentation of the investigation. Chapter 2 (Literature Review) provides an overview of existing research on the AO subtest as well as the subtests that comprise the AFQT. This chapter provides information on test background, test properties, predictive validity, and gender and racial subgroup differences. Chapter 3 (Methodology) describes the procedures used to select, clean, and analyze the data examined in this investigation. Chapter 4 (Results and Discussion) provides the results of the analyses and includes tables on demographics, descriptive statistics, subgroup differences, prediction equations, and bias analyses. Each finding is followed by a brief discussion of its implications. Chapter 5 (Conclusions) provides a summary of the important points, lists important limitations, and includes recommendations for further research.

CHAPTER 2: LITERATURE REVIEW

In this chapter, we provide a brief overview of spatial aptitude tests in general, followed by an overview of existing research on the AO subtest as well as the current AFQT subtests. For both AO and AFQT, we provide information on background, test properties, predictive validity, and gender and racial subgroup differences.

Spatial Aptitude Tests

The AFQT does not currently include a measure of spatial aptitude; however, a variety of research has demonstrated the relationship between spatial aptitude and performance. For example, previous research on spatial aptitude has demonstrated that it predicts outcomes including performance on psychomotor tasks (Kyllonen & Chaiken, 2003), creativity (González, Campos, & Pérez, 1997), virtual reality simulation performance (Manrique, 1998), training performance (Bertua, Anderson, & Salgado, 2005), and job performance (Bertua et al., 2005; Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003). Research has also suggested that the addition of a spatial aptitude test to the AFQT may result in an improvement in predictive validity (Wolfe, 1997). Such encouraging findings suggest that it may be beneficial to add a measure of spatial aptitude, such as the AO subtest, to the AFQT.

Assembling Objects

Background

The AO subtest is currently administered as one of the ten components of the ASVAB. Although AO is the most recent subtest to be added to the ASVAB, it has been in development for several decades. It was originally developed under the Army's Project A initiative in the 1980s, which aimed to improve the selection and classification of enlisted personnel (Busciglio & Palmer, 1996). Although AO has now been included on the ASVAB for eight years, it is currently used for limited classification purposes. Of all the Services, only the Navy currently uses AO, and the subtest is used in the qualification of only two Navy jobs.

Based upon an evaluation to synthesize findings from Project A and similar research undertakings in other services, AO was selected as one of nine subtests to be included in the Enhanced Computer-Administered Test (ECAT) battery. This battery is a collection of promising subtests that were identified as possessing the potential to improve upon the existing ASVAB. The Technical Advisory Selection Panel (TASP) that evaluated and selected tests for the ECAT aimed to include at least two tests of spatial aptitude. AO was selected as one of these tests due to encouraging findings in terms of its internal consistency reliability, retest reliability, uniqueness with respect to the ASVAB, likelihood of reduced adverse impact, construct validity, and predictive validity (Alderton, Wolfe, & Larson, 1997).

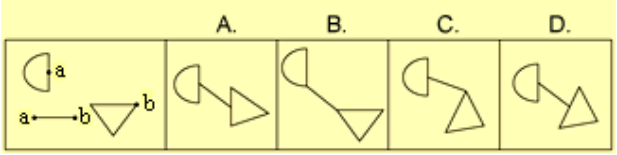
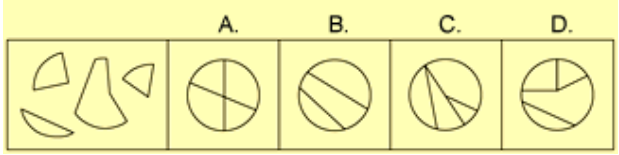
After extensive research on the ECAT subtests, only AO was selected for subsequent inclusion in the ASVAB. AO was incorporated into the ASVAB during the last content revision in 2002. At that time, two additional subtests, Numerical Operations (NO) and Coding Speed (CS), were dropped from the ASVAB (Office of the Secretary of Defense, 2010).

Test Properties

The AO subtest measures the "ability to figure out how an object will look when its parts are put together" (DMDC, 2008, p. 1), and it is currently the only ASVAB subtest that measures spatial aptitude (DMDC, 2008). In the paper-and-pencil version of the ASVAB, the AO subtest consists of 25 questions, and participants are given a 15-minute time limit to complete them. In the CAT-ASVAB, the AO subtest consists of 16 questions with a 16-minute time limit.

The AO subtest consists of two item types: connection items and puzzle items. The connection questions illustrate two objects, which are labeled with a letter in one spot on each of the two objects. Adjacent to the shapes is a line with letters on each end corresponding to the letters on the shapes, which indicate where the line must be connected to the two shapes. Participants must visualize how the object will look when the two shapes are connected by the line as specified by the lettering in the initial diagram. In the puzzle questions, participants are provided with pieces of a disassembled puzzle. Participants must visualize how the puzzle will look when it is assembled. Prior research indicates that connection items tend to be more difficult than puzzle items (DMDC, 2009). Table 2.2 presents sample items for each of the two item types.

Table 2.1. Sample Assembling Objects Items

Sample Connection Item	Sample Puzzle Item
	
Which figure best shows how the objects in the left box will touch if the letters for each object are matched? o A. o B. o C. o D.	Which figure best shows how the objects in the left box will appear if they are fit together? o A. o B. o C. o D.

Source: Office of the Secretary of Defense, 2010, Official Site of the ASVAB, http://www.official-asvab.com/questions/app/question_ao1_app.htm.

In general, AO has performed well psychometrically, as was demonstrated through extensive research prior to its inclusion in the ASVAB. In particular, the AO has shown high internal consistency reliability (e.g., Carey, 1994; Larson & Alderton, 1997), as well as high test-retest reliability (e.g., Alderton et al., 1997; Carey, 1994; Larson & Alderton, 1997). AO does not appear to be susceptible to coaching effects (Busciglio & Palmer, 1996); however, research on its susceptibility to practice effects has been mixed. Busciglio and Palmer (1996) found support for the impact of practice effects on AO, although the impact was comparable to that of the other two spatial subtests considered. In contrast, in an investigation conducted by Larson and Alderton (1997), AO was the only spatial subtest on the ECAT battery that did not display significant practice effects.

Predictive Validity

The AO was added to the ASVAB battery largely due to promising results in terms of its predictive ability. For example, Mayberry and Hiatt (1990) found that, of five subtests considered, AO was the best predictor of job knowledge test scores across a variety of Army jobs, or Military Occupational Specialty (MOS), as well as the best predictor of a hands-on performance test for two of the four MOS. AO also displayed incremental validity in comparison to the other existing ASVAB tests (Mayberry & Hiatt, 1990). Busciglio (1990) found that AO was the most useful of six spatial tests for predicting a variety of performance criteria. Subsequent research found that AO was a strong predictor across MOS and performance criteria, and that it remained a significant predictor after the ASVAB was added into a stepwise regression (Busiglio, 1991). Abrahams, Pass, Kusulas, Cole, and Kieckhaefer (1993) found that the AO subtest could outperform some of the subtests on the ASVAB in terms of its ability to predict final school grade and other training criteria. When averaged across the 17 MOS-specific training schools included in the study, AO provided a modest amount of incremental validity; however, the amount of incremental validity added by the subtest was particularly substantial for certain schools, suggesting that it may be more predictive for certain specializations. Carey (1994) found that AO added incremental validity to the existing ASVAB when predicting hands-on mechanical performance. Wolfe (1997) found that adding a subtest to measure each of three factors—psychomotor aptitude, working memory, and spatial aptitude—to the ASVAB added the most incremental validity, and AO served as a good example of a spatial aptitude subtest that added incremental validity. Finally, Caretta and King (2008) found that AO was significantly correlated with Air Traffic Controller (ATC) training graduation or elimination, ATC Fundamentals Course test score, and Federal Aviation Administration Certified Tower Operator (CTO) test score.

The majority of research has focused on a variety of performance variables as the criteria; however, some research has also examined the relationship of AO with attrition. Although the ASVAB has not historically been predictive of attrition (Laurence, Naughton, & Harris, 1996), the AO subtest has indeed shown significant correlations with attrition from Initial Military Training (IMT), unit attrition during the first nine months of service, and all attrition that occurred during the first 15-months of service (Putka & Bradley, 2008).

Subgroup Differences

Gender

In general, males often tend to outperform females on spatial aptitude tests (e.g., Geary, Saults, Liu, & Hoard, 2000; Halpern, 1997; Maccoby & Jacklin, 1974; Maitland, Intrieri, Schaie, & Willis, 2000; Malinowski, 2001; Weiss, Kemmler, Deisenhammer, Fleischhacker, & Delazer, 2003; Wise, Welsh, Grafton, Foley, Earles, Sawin, & Divgi, 1992). Despite this common trend, research has suggested that the magnitude of these differences varies with the type of spatial aptitude test that is administered. For example, Linn and Petersen (1985) found that the largest gender differences occurred on tests that involve three-dimensional spatial rotation and the smallest differences occurred on tests that involve spatial visualization. Thus, it is important to examine potential gender differences on tests that specifically involve object assembly.

Research to date on gender differences on the AO subtest has been relatively promising. In general, it appears that the gap between male and female scores is reduced (or eliminated) on the AO subtest in comparison to other tests of spatial aptitude (Larson & Alderton, 1997; Russell & Peterson, 2001; Russell, Reynolds, & Campbell, 1994). This finding is consistent with previous research indicating that gender differences tend to be lower on spatial tests measuring visualization in contrast to other forms of spatial aptitude (e.g., orientation, speeded rotation) (Linn & Peterson, 1985). Additionally, it has displayed similar gender subgroup differences in comparison to the overall AFQT (Knapp & Heffner, 2009). In regard to the two question types included on the AO subtest, it appears that gender differences may be more of a concern for the connection items in comparison to the puzzle items (DMDC, 2009). Overall, although it seems that the AO subtest may produce smaller gender differences than other tests of spatial aptitude, this issue still warrants further investigation.

Race/Ethnicity

Race and ethnicity differences must also be examined carefully, as performance on aptitude tests tends to differ among racial and ethnic groups (e.g., Suzuki & Valencia, 1997). When examining potential racial differences in terms of specific cognitive abilities, Loehlin, Lindzey, and Spuhler (1975) found that the largest Black-White differences were on tests of spatial aptitude, which further highlights the need to examine potential race differences on AO.

Research to date has generally shown that Blacks and Whites perform differently on AO, with Blacks scoring significantly lower; however, this difference appears to be of similar magnitude to the difference in performance between these two groups on the AFQT (Alderton et al., 1997; Knapp & Heffner, 2009). Further, the AO subtest appears to result in a smaller difference in performance between Hispanics and Whites than those found with the AFQT (Alderton et al., 1997; Knapp & Heffner, 2009). Similarly, Carey (1994) found that Black-White and Hispanic-White differences were smaller for the AO subtest than for the two more mechanically-oriented existing ASVAB subtests, Auto and Shop Information and Mechanical Comprehension. Other research has been less promising, however. Sager, Peterson, Oppler, and Rosse (1997) conducted research on a variety of test batteries that were composed of different combinations of ASVAB and ECAT subtests. In particular, they determined the frequency that each subtest appeared in the selection of batteries that minimized subgroup differences. AO was included less than half the time in a battery to minimize Hispanic-White differences, while it never appeared in a battery to minimize Black-White differences, suggesting that AO is below average in comparison to the other ASVAB and ECAT subtests when the goal is to minimize subgroup differences (Sager et al., 1997).

Less research has examined differences between the scores of Asian and White test takers; however, based upon the limited research, this comparison appears to be less of a concern. While a number of the ASVAB and ECAT subtests, as well as the AFQT overall, were found to result in significant differences between White and Asian test takers, this difference was not found to be significant for the AO subtest (Alderton et al., 1997).

AFQT Subtests

Background

In addition to examining the AO subtest, the current research initiative also requires consideration of the properties of the existing AFQT. The AFQT is computed from four of the ASVAB subtests: Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), and Mathematics Knowledge (MK). The ASVAB was first administered in 1968, and both AR and WK have remained on every version of the ASVAB since that time. MK was added during the first ASVAB revision in 1976, while PC was added during the second revision in 1980 (Office of the Secretary of Defense, 2010). The 1980 version of the AFQT consisted of AR, WK, PC, and Numerical Operations (NO), a subtest that was dropped from the ASVAB in 2002. Due to issues on the NO subtest such as considerable practice effects, research in the mid-1980s supported the replacement of the NO with MK (Maier, & Hiatt, 1986a; Maier, & Hiatt, 1986b). In an assessment of alternative AFQT composites, Monzon and Foley (1988) recommended the scoring composite that is now in place for the AFQT because it displayed comparable predictive validity to the former AFQT composite, while displaying reduced gender and racial subgroup differences.

Test Properties

WK and PC comprise the verbal domain of the AFQT. WK measures the "ability to select the correct meaning of words presented in context and to identify the best synonym for a given word," while PC measures the "ability to obtain information from written passages." AR and MK comprise the mathematics domain of the AFQT. AR measures the "ability to solve arithmetic word problems," while MK measures "knowledge of high school mathematics principles" (Office of the Secretary of Defense, 2010).

Table 2.3 presents the number of items and time limit for each of the four AFQT subtests. The subtests are administered in the order they are presented in the table. Within the larger ASVAB, General Science (GS) is administered first, followed by the four AFQT subtests, and then the remaining subtests: Electronics Information, Auto and Shop Information, Mechanical Comprehension, and Assembling Objects. Thus, all AFQT subtests are administered relatively early in the test-taking process.

Table 2.3. Number of Items and Time Limit for AFQT Subtests

	Paper-and-Pencil Administration		Computer Administration	
	Number of Items	Time Limit (in minutes)	Number of Items	Time Limit (in minutes)
Arithmetic Reasoning (AR)	30	36	16	39
Word Knowledge (WK)	35	11	16	8
Paragraph Comprehension (PC)	15	13	11	22
Mathematics Knowledge (MK)	25	24	16	20

Source: Office of the Secretary of Defense, 2010

AFQT scores are calculated using the following formula: 2VE + AR + MK, with VE being a scaled score of WK + PC. Scores are reported as percentiles from 1 to 99. Scoring is currently based off a reference group of a nationally representative sample of 18 to 23 year olds who took the ASVAB during a 1997 norming investigation (Office of the Secretary of Defense, 2010).

All AFQT subtests tend to display high internal consistency (e.g., Earles & Ree, 1992). Alderton and colleagues (1997) found that PC had fairly low test-retest reliability, which was well below the reliability of any of the other ASVAB or ECAT subtests. Conversely, test-retest reliability for WK and the two math subtests was quite high. Bivariate correlations tend to illustrate expected patterns based upon the underlying domains being measured. Specifically, correlations between the two verbal subtests and between the two math subtests are typically relatively high (though not high enough to suggest they are measuring the same construct), whereas correlations between the math-verbal dyads are lower in comparison but retain their significance (e.g., Alderton et al., 1997).

Predictive Validity

A variety of research has supported the overall predictive validity of the AFQT subtests (e.g., Stermer, 1988; Welsh, Kucinkas, & Curran, 1990; Wolfe, Moreno, & Segall, 1997). In comparing the two verbal subtests, previous research generally suggests that they have relatively comparable levels of predictive validity. For example, the two subtests have been found to perform similarly when predicting outcomes of training test performance and graduation from air traffic controller training (Carretta & King, 2008), mechanical hands-on performance (Carey, 1994), military technical school grade (Abrahams, Pass, Kusulas, Cole, & Kieckhaefer, 1993), and final school grade (DMDC, 2006). Alternatively, Earles and Ree (1992) found that WK is a slightly better predictor of course grades than PC across a variety of job types.

Mirroring the research on the two verbal subtests, it is not clear which of the two math subtests possesses more predictive ability. Carretta and King (2008) found that AR and MK were similarly predictive of air traffic controller training performance and graduation from the training program, and DMDC (2006) found the two subtests to have similar validity when predicting final school grade. Abrahams et al. (1993) also found comparable validity coefficients between the two math subtests on average; however, the regression weight was larger for MK than AR when predicting final school grade. In contrast, Carey (1994) found that AR was slightly more predictive of mechanical hands-on performance than MK, and similarly, Earles and Ree (1992) found that AR was most predictive of military technical school grade. Across a variety of jobs, AR and MK were found to be the best predictors of course grades in comparison to the other ASVAB subtests, with AR showing a slightly stronger relationship after correction for range restriction. AR also had a slightly stronger relationship with grade than MK for each individual job category. Driskell, Hogan, Salas, and Hoskin (1994) found that MK had the strongest relationship with training performance in comparison to several other ASVAB subtests and personality scales; however, none of the other AFQT subtests were included in the research.

Thus, overall it seems clear that the four AFQT subtests are predictive of a variety of performance criteria; however, in each content domain, there is no clear trend in terms of which subtest may be more predictive than the other.

Subgroup Differences

Fairness is often examined in terms of predictive bias, which considers whether the same predictor-criterion relationship holds for all subgroups of interest. Research indicates that there is no support for gender or racially-based predictive bias for the four AFQT subtests (DoD, 1994; Wise et al., 1992). Another concern is whether differences exist between subgroups on the actual test scores. Findings have been less conclusive regarding such subgroup differences on the AFQT.

Gender

In general, most research has found that women tend to outperform men on tests of verbal aptitude, while men tend to outperform women on tests of mathematical aptitude (e.g., Geary et al., 2000; Halpern, 1997; Maccoby & Jacklin, 1974; Maitland et al., 2000; Malinowski, 2001; Weiss et al., 2003; Wise et al., 1992). For example, Held, Alderton, Foley, and Segall (1993) found that men scored higher than women on the Arithmetic Reasoning (AR) subtest.

In a comparison of potential gender differences in the factor structure of the ASVAB, Ree and Caretta (1995) found that, although some differences were found in the more technically-oriented ASVAB subtests, the verbal and math subtests included in the AFQT displayed the same factor structure for men and women. This suggests that the scores on the math and verbal subtests have the same meaning for both genders.

In terms of overall performance on the AFQT, a recent investigation comparing opposite-sex siblings found that males score slightly higher than females on the AFQT and that males display greater variance. Almost twice as many males as females scored in the top two percent of AFQT scores. Overall, these findings suggest that males may be slightly more likely than females to score higher on the AFQT in general and that males are much more likely to achieve a very high score (Deary, Irwing, Der, & Bates, 2007).

Race/Ethnicity

In regard to race/ethnicity subgroup differences on the verbal subtests, Alderton et al. (1997) found that both verbal subtests tend to result in lower scores for Blacks, Asians, and Hispanics in comparison to Whites; however, these differences are smaller for PC in comparison to WK. Research on the math subtests found that Whites also scored significantly higher than the three other racial/ethnic groups on AR; however, fewer differences were found on MK. Specifically, Blacks scored lower than Whites on MK, although the difference was much smaller in comparison to the difference between these two groups on AR. Additionally, there was no difference between Hispanics and Whites on MK, and Asians scored higher than Whites on this subtest.

The remainder of this report will focus on our examination of data to determine the impact of revising the AFQT composite to either add AO, or use AO to replace another component score. In particular, we examine the impact in terms of subgroup differences, prediction of a variety of criteria, and predictive bias.

CHAPTER 3: METHODOLOGY

Sample

The sample for this investigation was drawn from the *Future Force Performance Measures (Army Class)* longitudinal validation. The purpose of that project, known as Army Class, is to evaluate the selection and classification potential of non-cognitive predictors. Soldiers were tracked starting at Reception Battalion and criterion data were obtained at the end of their initial military training (EOT) and approximately one year into their first unit assignment. Soldier's ASVAB scores were obtained as part of this project. Knapp and Heffner (2009) defined the longitudinal validation sample (LVS) as consisting of Soldiers with no prior service record, regardless of component or educational background. The sample consisted of 10,814 new Soldiers. Soldiers in the Army Class longitudinal validation sample belong to either one of six target MOSs or to an Army-wide subsample of various other MOSs. The six target MOSs are fairly representative of other Army occupations in terms of job characteristics, according to Moriarty, Campbell, Heffner, and Knapp (2009). These six MOSs include Infantryman (11B, 11C, 11X, and 18X), Armor Crewman (19K), Military Police (31B), Light Wheel Vehicle Mechanic (91B), Health Care Specialist (68W), and Motor Transport Operator (88M). The number of Soldiers from Army Class assigned to each of these six target MOSs range from 307 to 1790.

Measures

Predictor Measures

The predictor measures used in the regression analyses included the Assembling Objects (AO) subtest and the subtests that comprise the AFQT: Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), and Mathematics Knowledge (MK).

Criterion Measures

The criterion measures fell into four major categories: job knowledge, job performance, attitudes, and attrition. The measures also represented all aspects of can-do criteria (e.g., technical knowledge), and will-do criteria (e.g., physical fitness). These measures were assessed for the six target MOS at EOT and for all MOS during their first in-unit assignment. All non-administrative criterion measures were administered via the computer.

Job Knowledge

Job knowledge tests (JKTs) were administered to Soldiers at EOT and in-unit. The EOT JKTs (Ingerick, Diaz, & Putka, 2009) consisted of approximately 60 items and the in-unit JKTs (Moriarty et al., 2009) consisted of approximately 40 items. Both used a variety of question formats (e.g., multiple-choice, multiple-response, rank order, and drag and drop). For EOT, the JKTs were developed only for the six target MOSs. Their content was based on MOS-Specific performance requirements identified through job analysis and other job-relevant information (e.g., Soldier Manuals, Programs of Instruction). For the in-unit data collections, Soldiers in the six target MOSs were administered a MOS-Specific JKT. An additional Army-Wide JKT (now referred to as the

Warrior Tasks and Battle Drills test) that covered Soldiering skills relevant to all Soldiers in the Army also was developed. This Army-wide JKT was administered to every Soldier in the sample (including those in the target MOSs).

EOT job knowledge also was assessed via administrative data from an administrative personnel database, the Resident Individual Training Management System (RITMS). This database provided the Soldier's average EOT exam grade. Soldiers are tested on their MOS at the end of every performance block in training. This is the average of those scores.

Job Performance

Job performance rating scales were developed for the EOT and in-unit data collections (see Moriarty et al., 2009 for a more detailed description). A behaviorally anchored ratings scales (BARS) format was employed for these measures. The number of dimensions per set of scales ranged from four to eight, depending on MOS. The EOT scales were completed by peers and supervisors, whereas the in-unit scales were only completed by Soldiers' supervisors. Soldiers in the target MOS were rated on both Army-wide (e.g., Exhibits Efforts, Solves Problems) and MOS specific components (Responds to Emergency Situations, Learns to Use Aiming Devices), whereas Soldiers in non-target MOSs were rated on Army-wide components only.

Attitudes

The Army Life Questionnaire (ALQ) was designed to measure Soldiers' self-reported attitudes and experiences. The original form of the ALQ was developed for a previous ARI research effort (Van Iddekinge, Putka, & Sager, 2005) with the current ALQ having slight modifications from the original (Moriarty et al., 2009). Both assessed commitment and other retention related attitudes such as satisfaction with the Army and satisfaction with MOS. The ALQ employed Likert-style response options.

Attrition

Attrition data were obtained on participating Soldiers through their first six months of service. At the time these analyses were conducted, in-unit attrition data were not available. The six-month timeframe was chosen due to the fact that it coincided approximately with the completion of initial training for most Soldiers. Attrition information was extracted for participating Soldiers from the Two Tier Attrition Screen (TTAS) database maintained by the U.S. Army Accessions Command. Attrition data were available only for Active Army Soldiers.

Criteria were selected for this investigation according to three rules:

- Use a diverse set of criteria in order to access the breadth of the criterion space
- Use criteria with a relatively large sample size in order to minimize the potential for type II error
- Aggregate MOS-specific measures in order to maximize the generalizability of the findings and to reduce the potential for both type II and type I errors.

When possible, we examined similar types of criteria at two points in time: end of training (EOT) and in-unit (IU).

We calculated the following criteria:

- Total scores for the Army-wide performance rating scales were formed by summing across dimension scores for peer, supervisor, and overall ratings.
- Performance rating scores within MOS were formed by standardizing raw scores for each performance dimension within each target MOS and then consolidating the resulting variables into a single field.
- MOS job knowledge test scores were calculated by standardizing the total raw scores on the MOS-specific job knowledge test within each target MOS and then consolidating the resulting variables into a single field.
- Attitude fit scores were calculated by summing the two Army Life Questionnaire (ALQ) attitude fit dimension scores (i.e., the Needs-Supplies Army Fit Scale, and the General MOS Fit Scale).

These scores were developed for both the EOT and IU timeframes. We also selected several extant variables for further examination as criteria in this investigation. We calculated descriptive statistics and correlations among the potential criteria and used the results to further narrow the list of criteria. When multiple criteria of the same type were available, or when correlations among two or more criteria were particularly high (e.g., > .90) we selected the criteria that were most theoretically sound. This resulted in the 20 criteria provided in Table 3.1.

Table 3.1. Criteria Identified or Calculated for Examining the Predictive Value of AO

	Title	Source	Type
End of Training			
1	MOS Job Knowledge (JKT)	EOT JKT: Army Class score standardized within MOS	Job Knowledge
2	Course Average (RITMS)	EOT RITMS: course average, standardized within MOS	Job Knowledge
3	Army-Wide Performance Ratings (PRS)	EOT PRS: Sum of AW Scores	Performance Rating
4	MOS Performance Ratings (PRS)	EOT PRS: Sum of standardized MOS scores	Performance Rating
5	Affective Commitment (ALQ)	EOT ALQ: Affective Commitment Scale	Attitude
6	Army and MOS Fit (ALQ)	EOT ALQ: Sum of two scales: the Needs-Supplies Army Fit Scale, and the General MOS Fit Scale	Attitude
7	Number of disciplinary incidents during Army training (ALQ)	EOT ALQ: Number of disciplinary incidents during Army training	Attitude
8	Physical Fitness (APFT)	EOT ALQ: What was your last APFT score?	Attitude
9	Attrition cognitions (ALQ)	EOT ALQ: Attrition cognitions scale	Attrition
10	Total Number of Recycles during BCT and/or AIT/OSUT (ATTRS)	EOT ATRRS Training Data: Total number of restarts during IMT	Attrition

Table 3.1. Criteria Identified or Calculated for Examining the Predictive Value of AO (Continued)

	Title	Source	Type
In-Unit			
11	Army Wide Job Knowledge (JKT)	IU JKT: AW Army Class total score	Job Knowledge
12	MOS Job Knowledge (JKT)	IU JKT: Overall Army Class total score standardized within MOS	Job Knowledge
13	Army-Wide Performance Ratings (PRS)	IU PRS: Sum of AW scores	Performance Rating
14	MOS Performance Ratings (PRS)	IU PRS: Sum of Standardized MOS Scores	Performance Rating
15	Affective Commitment (ALQ)	IU ALQ: Affective Commitment Scale	Attitude
16	Army and MOS Fit (ALQ)	IU ALQ: Sum of two scales: the Needs-Supplies Army Fit Scale, and the General MOS Fit Scale	Attitude
17	Disciplinary Action Total Score (ALQ)	IU ALQ: Disciplinary action total score	Attrition
18	Physical Fitness (APFT)	IU ALQ: What was your last APFT score?	Attrition
19	Attrition cognitions (ALQ)	IU ALQ: Attrition cognition mean score	Attrition
20	Attrition at six months (TTAS)	TTAS: Attrition six months	Attrition

Note: JKT: Job Knowledge Test, RITMS: Resident Individual Training Management System, AW: Army-Wide, PRS: Performance Rating Scales, BCT: Basic Combat Training, AIT: Advanced Individual Training, OSUT: One-Station Unit Training, TTAS: Tier Two Attrition Screen, ATTRS: Army Training Requirements Reporting System; APFT: Army Physical Fitness Test.

Analyses

We conducted the following analyses:

- Preliminary Data Investigation
- Evaluation of Supplemental Predictor Composites
- Evaluation of Unit Weighted Substitutional Predictor Composites

We discuss our procedures for these analyses in more detail below.

Prior to performing any analyses, we applied the following data cleaning procedures:

- Identify, flag, and exclude duplicate cases from the analyses.
- Examine frequency distributions for each variable of interest, looking for unlabelled values, out-of-range values, theoretically impossible values, or undiscovered missing value designators, typos, inconsistent response formats, and changes in unit of measurement.
- Examine patterns of missing values in order to identify potential issues with question clarity, data entry or revision, and/or data file merging prior to flat file delivery, using frequency distributions and visual inspection of the data files.
- Examine variable pairs that are expected to have some relationship to one another to identify combinations of values that are not plausible, are extremely unlikely, or are contradictory, using crosstabs and/or scatterplots.

We worked closely with ARI researchers to identify and address any potential problems.

Preliminary Data Investigation

We conducted the following preliminary analysis:

- a preliminary investigation of the data in order to show the characteristics of the sample and to identify issues regarding the five subtest predictors.
- demographic statistics of the sample at each of three data collection points: (1) reception battalion predictor administration, (2) EOT criterion administration, and (3) IU criterion administration. We used gender, race, ethnicity, age, education level, component, and MOS to describe the demographics of the sample.
- descriptive statistics and intercorrelations among the five predictors.
- Cohen's d for each subtest to assist in describing subgroup differences for each protected group relative to its respective majority group.

The results of these analyses are reported in Chapter 4.

Evaluation of the Supplemental Predictor Composite

The purpose of this set of analyses was to determine whether adding the AO subtest to the AFQT would improve prediction of the criteria and reduce subgroup differences. We developed the Supplemental Predictor composite as a unit-weighted sum of each of five subtests, including the four subtests used to form the AFQT and the AO subtest. For the sake of comparison, we calculated the Baseline composite as the unit-weighted sum of each of the four AFQT subtests. We compared unit-weighted models here because we expect the current AFQT composite, and any changes to that composite, would likely employ unit weighting of the subtests.

We compared the Baseline and Supplemental composites in a series of analyses, including:

- Regression analyses, where each criterion was regressed on either the Baseline or Supplemental composite. Prior to conducting these regressions, we implemented multivariate range-restriction corrections by correcting the sample covariance matrix using the AFQT score variance from the 1997 Youth Population as the indicator of population variance. We also adjusted each multiple R result for shrinkage using Rozeboom's (1978) formula 8.
- Subgroup analyses, where we calculated Cohen's *d* to compare the composite scores for each protected group relative to their respective majority groups.
- Bias analyses (Cleary, 1968) for each subgroup comparison, where we regressed the criterion on a dummy coded variable representing subgroup membership and then the respective predictor composite. Using the Cleary model in this manner allowed us to assess whether there was differential prediction in terms of both intercept and slope differences. To further examine potential bias, we identified the equation and regression line for each subgroup. Then we plotted the regression lines and looked for differences in predicted performance along the regression lines. This allowed us to determine, for each subgroup, the type and degree of prediction error associated with using the total regression line (as opposed to the respective subgroup regression line).

To address issues with a potentially large experiment-wise error rate, we set the alpha level at .01 (less than the traditional .05) and, when possible, we provided effect size estimates to aid in interpreting the meaningfulness of the statistics.

Evaluation of Substitutional Predictor Composites

The purpose of this set of analyses was to determine whether replacing any of the four AFQT subtests with the AO subtest would improve prediction of criteria and reduce subgroup differences to a greater extent than simply adding AO to the existing AFQT model. We calculated four Substitutional predictor composites by substituting the AO subtest for one of the four subtests that are currently used to calculate the AFQT score.

We conducted a series of analyses to compare the four possible Substitutional predictor composites to the unit weighted composite used to form the AFQT, including:

- Regression analyses, where each criterion was regressed on the Substitutional composite. Prior to conducting these regressions, we implemented multivariate range-restriction corrections by correcting the sample covariance matrix according to the AFQT score variance from the 1997 Youth Population as the indicator of population variance. We also adjusted each multiple R result for shrinkage using Rozeboom's (1978) formula 8.
- Subgroup analyses, where we calculated Cohen's *d* to compare the composite scores for each protected group relative to their respective majority groups.
- Bias analyses (Cleary, 1968) for each subgroup comparison, where we regressed the criterion on a dummy coded variable representing subgroup membership and then the respective predictor composite. Using the Cleary model in this manner allowed us to assess whether there was differential prediction in terms of both intercept and slope differences. To further examine potential bias, we identified the equation and regression line for each subgroup. Then we plotted the regression lines and looked for differences in predicted performance along the regression lines. This allowed us to determine, for each subgroup, the type and degree of prediction error associated with using the total regression line (as opposed to the respective subgroup regression line). We conducted these analyses only in those situations where the percentage of variance predicted by the Substitutional predictor composite (indicated by R^2) was at least five percent, and greater than that for the AFQT composite.

CHAPTER 4: RESULTS AND DISCUSSION

Preliminary Data Investigation

Demographics

Demographic characteristics for the EOT, IU, and longitudinal validation population (LVP) samples are presented in Table 4.1. The most notable result is that the EOT sample differs from the IU and LVP samples in terms of the distribution of MOS. The EOT sample was composed almost entirely of Soldiers in the six target MOS, while participants in these MOS composed 29.1 percent of the IU sample and 21.0 percent of the LVP sample participants. Other differences among these samples are less pronounced and therefore may be incidental to the MOS differences.

Table 4.1. Participant Demographic Characteristics.

	LVP Sample		EOT Sample		IU Sample	
Subgroup	*n*	*%*	*n*	*%*	*n*	*%*
Gender						
Male	8,646	80.0	2,081	90.8	1,246	81.9
Female	2,113	19.5	207	9.0	270	17.7
Race						
White	8,431	78.0	1,975	86.2	1,187	78.0
Black	1,527	14.1	157	6.8	192	12.6
Other	818	7.6	153	6.7	137	9.0
Ethnicity						
White, non-Hispanic	7,541	69.7	1,775	77.4	1,049	68.9
Hispanic	1,527	14.1	323	14.1	236	15.5
Highest Education Level						
High School Degree or Greater	8,103	74.9	1,667	72.7	1,169	76.8
No High School Degree	2,682	24.8	623	27.2	353	23.2
Component						
Active Army	5,370	49.7	1,385	60.4	1,082	71.1
Army National Guard	3,793	35.1	694	30.3	269	17.7
Army Reserve	1,651	15.3	213	9.3	171	11.2
MOS						
11B: Infantryman	1,797	16.6	671	29.3	315	20.7
19K: Armor Crewman	579	5.4	469	20.5	95	6.2
31B: Military Police	1,482	13.7	715	31.2	199	13.1
91B: Light Wheel Vehicle Mechanic	475	4.4	218	9.5	59	3.9
68W: Health Care Specialist	304	2.8	136	5.9	36	2.4
88M: Motor Transport Operator	511	4.7	72	3.1	59	3.9
Other	5,666	52.4	11	0.5	759	49.9
Total	10,814		2,292		1,522	

Note. The figures reported do not total due to missing data. MOS: Military Occupational Specialty, LVP Sample: Collected in Reception Battalion, EOT Sample: collected at the end of Initial Entry Training (IET) including either one-station unit training (OSUT) or advanced individual training (AIT) with data collection beginning in the fall of 2007 and continuing through the summer of 2008, IU Sample: Targeted the same sample of Soldiers at about 18-20 months time in service (TIS).

Predictor Characteristics

Descriptive Statistics

Descriptive statistics for the AFQT and AO subtests are presented in Table 4.2. Mean scores were lowest on WK and highest on AO; however, in this sample AO displayed the most variability in scores relative to the norming sample.

Table 4.2. Descriptive Statistics for the AFQT and AO Subtests.

Subtest	N	Min.	Max.	M	SD
Word Knowledge (WK)	9,902	20	61	49.94	5.97
Paragraph Comprehension (PC)	9,467	19	60	51.47	5.09
Arithmetic Reasoning (AR)	10,498	26	66	51.82	6.29
Mathematics Knowledge (MK)	10,633	26	67	52.17	6.30
Assembling Objects (AO)	9,875	25	70	54.88	7.95

Note: All subtest scores are normed according to the 1997 Youth Population sample, where the mean of each subtests was 50 and the standard deviation was 10.

Intercorrelations

The AFQT and AO intercorrelation matrix is presented in Table 4.3. Correlations were highest between subtests in the same domain; specifically, the correlation between AR and MK was .56, and the correlation between WK and PC was .43. Correlations between subtests measuring different domains were lower; however, all correlations were significant, partly due to the large sample size. With regard to AO, this subtest correlated more strongly with the two math subtests, and AR in particular, when compared to the two verbal subtests.

Table 4.3. AFQT and AO Subtest Intercorrelation Matrix.

	Subtest				
Subtest	WK	PC	AR	MK	AO
Word Knowledge (WK)					
Paragraph Comprehension (PC)	0.43				
Arithmetic Reasoning (AR)	0.25	0.28			
Mathematics Knowledge (MK)	0.10	0.15	0.56		
Assembling Objects (AO)	0.16	0.19	0.39	0.32	

Note. n ranges from 8,708 to 10,437. All correlations are significant at $p < .01$.

Subgroup Differences

Subgroup differences were calculated using Cohen's d and t-tests. The results are presented in Table 4.4. AO displayed significant subgroup differences favoring the majority group in all three comparisons; however, in some instances, these effect sizes were smaller than those observed on the existing AFQT subtests. In particular, AR displayed a stronger difference between males and females in comparison to AO. In addition, AR, WK, and PC displayed stronger differences

between White and Hispanic participants in comparison to AO. AO, AR, and WK each showed medium to large differences favoring White over Black participants.

Table 4.4. Subgroup Differences Reported Using Cohen's d.

Subtest	Subgroup	*n*	*M*	*SD*	d_m
Word Knowledge (WK)	Female	1,976	49.11	5.90	-.17**
	Male	7,926	50.15	5.97	NA
	Black	1,352	47.42	5.88	**-.61****
	Hispanic	1,447	48.45	5.96	**-.43****
	White	6,502	50.88	5.61	NA
Paragraph Comprehension (PC)	Female	1,894	51.22	4.87	-.06*
	Male	7,573	51.54	5.14	NA
	Black	1,320	50.17	5.15	-.36**
	Hispanic	1,401	50.69	5.14	**-.25****
	White	6,160	51.96	4.95	NA
Arithmetic Reasoning (AR)	Female	2,088	50.24	6.08	**-.32****
	Male	8,410	52.21	6.28	NA
	Black	1,388	48.78	5.78	**-.63****
	Hispanic	1,494	50.83	6.08	**-.29****
	White	6,977	52.60	6.18	NA
Mathematics Knowledge (MK)	Female	2,095	52.87	5.92	.14**
	Male	8,538	52.00	6.38	NA
	Black	1,395	51.62	5.92	-.10**
	Hispanic	1,498	51.81	6.22	-.07*
	White	7,100	52.25	6.33	NA
Assembling Objects (AO)	Female	1,909	53.71	7.73	-.18**
	Male	7,966	55.16	7.98	NA
	Black	1,283	50.93	8.51	-.60**
	Hispanic	1,423	54.75	7.76	-.11**
	White	6,570	55.61	7.64	NA

Note. Negative values of *d* represent subgroup differences that favor the majority group. Statistically significant subgroup differences at the .05 level are indicated by a single asterisk; subgroup differences at the .01 level are indicated by a double asterisk. Effect sizes for AFQT subtests that are greater than the associated effect size for the AO subtest are presented in bold.

Criterion Characteristics

Descriptive Statistics

Descriptive statistics for the 20 criterion variables included in the preliminary analyses are presented in Table 4.5. One issue of note in these results is that there are large differences in the number of valid cases for each criterion. Specifically, the criteria relating to attrition had the largest sample sizes, while the criteria gathered at EOT had the second largest sample sizes, and the criteria pertaining to in-unit performance had the lowest sample size.

Table 4.5. Descriptive Statistics for Criterion Variables.

	Criterion	*n*	*M*	*SD*
End of Training				
1	MOS Job Knowledge (JKT)	2093	.00	1.00
2	Course Average (RITMS)	1474	.01	.97
3	Army-Wide Performance Ratings (PRS)	2277	30.13	4.99
4	MOS Performance Ratings (PRS)	2232	.00	1.00
5	Affective Commitment (ALQ)	2196	3.87	.68
6	Army and MOS Fit (ALQ)	2217	7.68	1.37
7	Number of disciplinary incidents during Army training (ALQ)	2217	.49	.91
8	Physical Fitness (APFT)	2208	245.16	31.71
9	Attrition cognitions (ALQ)	2193	1.65	.70
10	Total Number of Recycles during BCT and/or AIT/OSUT (ATTRS)	9681	.09	.32
In-Unit				
11	Army Wide Job Knowledge (JKT)	1374	18.53	3.56
12	MOS Job Knowledge (JKT)	628	.00	.99
13	Army-Wide Performance Ratings (PRS)	914	70.14	16.13
14	MOS Performance Ratings (PRS)	435	.00	.99
15	Affective Commitment (ALQ)	1409	3.58	.85
16	Army and MOS Fit (ALQ)	1409	7.16	1.47
17	Disciplinary Action Total Score (ALQ)	1409	.56	1.13
18	Physical Fitness (APFT)	1314	242.42	39.14
19	Attrition cognitions (ALQ)	1409	1.69	.79
20	Attrition at six months (TTAS)	5811	.09	.29

Note: JKT: Job Knowledge Test, RITMS: Resident Individual Training Management System, AW: Army-Wide, PRS: Performance Rating Scales, BCT: Basic Combat Training, AIT: Advanced Individual Training, OSUT: One-Station Unit Training, TTAS: Tier Two Attrition Screen, ATTRS: Army Training Requirements Reporting System; APFT: Army Physical Fitness Test. Criterion #1.#2. #4, #12, and #14 have mean values of 0 because they were created by standardizing by MOS within the sample.

Intercorrelations

Criterion variable intercorrelations are presented in Table 4.6. Significant correlations are provided in boldface type.

Table 4.6. Criterion Variable Intercorrelations.

ID	Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
End of Training																					
1	MOS Job Knowledge (JKT)																				
2	Course Average (RITMS)	na																			
3	Army-Wide Performance Ratings (PRS)	**.18**	na																		
4	MOS Performance Ratings (PRS)	**.15**	na	**.72**																	
5	Affective Commitment (ALQ)	**.07**	na	**.08**	**.11**																
6	Army and MOS Fit (ALQ)	**.10**	na	**.19**	**.15**	**.68**															
7	Number of disciplinary incidents during Army training(ALQ)	**-.13**	na	**-.29**	**-.22**	**-.14**	**-.19**														
8	Physical Fitness (APFT)	.00	na	**.29**	**.21**	.05	**.13**	**-.14**													
9	Attrition cognitions (ALQ)	**-.13**	na	**-.21**	**-.17**	**-.65**	**-.61**	**.21**	**-.11**												
10	Total Number of Recycles during BCT and/or AIT/OSUT (ATTRS)	**-.06**	**-.09**	**-.15**	**-.12**	-.05	**-.09**	**.11**	**-.07**	**.09**											
In-Unit																					
11	Army Wide Job Knowledge (JKT)	**.35**	**.34**	.10	.09	.01	.11	-.13	.08	-.08	-.04										
12	MOS Job Knowledge (JKT)	**.46**	na	.09	.13	.02	.08	-.1	.02	-.03	.06	**.57**									
13	Army-Wide Performance Ratings (PRS)	.05	.11	**.24**	**.16**	-.05	.05	**-.24**	**.16**	-.03	**-.10**	**.15**	**.13**								
14	MOS Performance Ratings (PRS)	.06	na	**.20**	.12	-.06	.04	-.12	.09	.02	**-.14**	.10	**.17**	**.80**							
15	Affective Commitment (ALQ)	-.02	.04	.03	.09	**.31**	**.27**	-.02	.02	**-.22**	.01	.07	.06	**.16**	.12						
16	Army and MOS Fit (ALQ)	.07	.10	.11	**.15**	**.22**	**.39**	-.03	.08	**-.22**	-.05	**.12**	**.13**	**.23**	**.20**	**.68**					
17	Disciplinary Action Total Score (ALQ)	-.07	-.15	**-.13**	-.12	-.04	-.09	**.25**	-.12	.08	**.08**	**-.08**	-.01	**-.44**	**-.29**	**-.20**	**-.23**				
18	Physical Fitness (APFT)	.05	.04	**.24**	**.17**	.04	.11	**-.24**	**.56**	-.02	-.06	**.09**	.01	**.13**	.03	.03	.03	-.05			
19	Attrition cognitions (ALQ)	-.11	-.07	-.11	-.10	**-.22**	**-.21**	.10	-.10	**.26**	.05	**-.19**	**-.15**	**-.32**	**-.25**	**-.55**	**-.52**	**.31**	**-.13**		
20	Attrition at six months (TTAS)	-.05	**-.16**	**-.16**	**-.15**	**-.11**	**-.10**	**.08**	-.03	**.19**	**.04**	-.01	(.a)	.03	.06	.04	.01	.04	.04	.04	

Note: Correlations in boldface are significant at the .01 level. na: Not applicable as correlations could not be computed as there were no observations with non-missing values on both variables. (a.): Could not be computed because at least one of the variables was constant.

Evaluation of Supplemental Predictor Composites

Table 4.7 provides R^2 statistics for overall predictor models, corrected for restriction of range and adjusted for shrinkage. For each criterion, we tested two models. The Baseline model included a unit-weighted composite of the four subtests that currently comprise the AFQT (WK, PC, AR, and MK). The Supplemental model included the AO subtest in addition to the four AFQT subtests. Both composites were unit-weighted because the Army currently unit-weights the four AFQT subtests (AFQT= AR + MK + (2 x VE) where VE = PC + WK). We compared R^2 statistics for the two models to determine the extent to which the AO subtest explained variance in each criterion over the AFQT.

Table 4.7. Prediction Equations

Criteria		Model Type	*n*	*R*	R_A	R_C	R_{CA}	R^2_{CA}	$R^2_{CA}\Delta$
End-of-Training									
1	MOS Job Knowledge (JKT)	B	1603	.357	.356	.561	.560	.314 *	
		S	1603	.379	.378	.573	.572	.327 *	.013 *
2	Course Average (RITMS)	B	1187	.262	.259	.422	.420	.177 *	
		S	1187	.254	.250	.417	.415	.173 *	-.004
3	Army-Wide Performance Ratings (PRS)	B	1758	.155	.151	.258	.256	.066 *	
		S	1758	.180	.176	.273	.271	.073 *	.008 *
4	MOS Performance Ratings (PRS)	B	1728	.140	.136	.237	.235	.055 *	
		S	1728	.161	.158	.250	.248	.061 *	.006 *
5	Affective Commitment (ALQ)	B	1688	.031	.000	.058	.046	.002	
		S	1688	.026	.000	.054	.042	.002	.000
6	Army and MOS Fit (ALQ)	B	1705	.015	.000	.021	.000	.000	
		S	1705	.018	.000	.023	.000	.000	.000
7	Number of disciplinary incidents during Army training (ALQ)	B	1705	.051	.038	.087	.080	.006	
		S	1705	.064	.054	.094	.088	.008	.001
8	Physical Fitness (APFT)	B	1696	.060	.049	.091	.084	.007	
		S	1696	.054	.042	.088	.081	.007	-.001
9	Attrition cognitions (ALQ)	B	1686	.044	.028	.070	.061	.004	
		S	1686	.044	.028	.071	.062	.004	.000

Table 4.7. Prediction Equations (Continued)

	Criteria	Model Type	*n*	*R*	R_A	R_C	R_{CA}	R^2_{CA}		$R^2_{CA}\Delta$
In-Unit										
10	Total Number of Recycles during BCT and/or AIT/OSUT (ATTRS)	B	7366	.042	.038	.068	.066	.004	*	
		S	7366	.046	.043	.070	.068	.005	*	.000
11	Army Wide Job Knowledge (JKT)	B	992	.400	.398	.591	.590	.348	*	
		S	992	.414	.412	.598	.597	.356	*	.009
12	MOS Job Knowledge (JKT)	B	474	.349	.344	.512	.509	.259	*	
		S	474	.381	.377	.530	.528	.278	*	.020
13	Army-Wide Performance Ratings (PRS)	B	702	.076	.055	.116	.103	.011		
		S	702	.117	.104	.142	.132	.017	*	.007
14	MOS Performance Ratings (PRS)	B	344	.060	.000	.094	.056	.003		
		S	344	.099	.064	.119	.091	.008		.005
15	Affective Commitment (ALQ)	B	1014	.066	.049	.108	.098	.010		
		S	1014	.056	.034	.101	.091	.008		-.001
16	Army and MOS Fit (ALQ)	B	1014	.006	.000	.013	.000	.000		
		S	1014	.004	.000	.007	.000	.000		.000
17	Disciplinary Action Total Score (ALQ)	B	1014	.006	.000	.013	.000	.000		
		S	1014	.014	.000	.000	.000	.000		.000
18	Physical Fitness (APFT)	B	941	.056	.032	.083	.070	.005		
		S	941	.063	.043	.088	.075	.006		.001
19	Attrition cognitions (ALQ)	B	1014	.042	.000	.063	.045	.002		
		S	1014	.071	.055	.082	.069	.005		.003
20	Attrition at six months (TTAS)	B	4529	.021	.000	.032	.024	.001	*	
		S	4529	.040	.034	.044	.039	.002	*	.001

Note: B: Baseline Model, S: Supplemental Model. R = uncorrected multiple R. R_A = uncorrected multiple R adjusted for shrinkage. R_C = multiple R corrected for range restriction. R_{CA} = corrected multiple R adjusted for shrinkage. R^2_{CA} = corrected R^2 adjusted for shrinkage. R^2_{CA} Change = change in R^2_{CA} from Baseline model to Supplemental model. * $p < .01$.

Results of the regression analyses on EOT criteria show AO significantly contributes to the prediction of performance-based outcomes. The Supplemental model explained the greatest amount of variance beyond the Baseline model in job knowledge test scores ($\Delta R^2 = .013$, $p < .01$). In addition, the AO subtest significantly improved the prediction of Army-wide performance dimensions ($\Delta R^2 = .008$, $p < .01$) and MOS-Specific performance dimensions ($\Delta R^2 = .006$, $p < .01$). An explanation for the relationships between AO subtest scores and performance ratings could be that the AO subtest predicts Soldiers' performance of the technical aspects of their jobs (e.g., assembling and disassembling weapons) through learning and job knowledge. The performance ratings at least partly reflect this technical performance, especially through the PRS dimensions referencing knowledge and skill of common tasks and MOS-Specific tasks, leading the AO subtest to positively predict performance ratings in general.

The AO subtest did not contribute to the prediction of criteria that are not as directly related to job performance. For example, the Supplemental model did not predict self-rated fit ($\Delta R^2 = .000$, $p > .01$), attrition cognitions ($\Delta R^2 = .000$, $p > .01$), or physical fitness test scores

(ΔR^2 = -.001, p > .01) over the Baseline model. Thus, based on findings from the EOT criteria, it seems that the AO subtest has the greatest promise as a predictor of the extent to which Soldiers understand and execute the technical aspects of their jobs. The AO subtest performed less well as a predictor of Soldiers' attitudes and motivation. It should be noted, however, as Table 4.7 shows, that the AFQT does not strongly predict motivation-based criteria either. This result is consistent with previous research that shows cognitive aptitude predicts more cognitively driven criteria better than it does less cognitively-driven criteria (Ones, Viswesvaran, & Dilchert, 2005).

Results of analyses on IU criteria present similar conclusions regarding the predictive potential of the AO subtest. More specifically, the Supplemental model significantly added to the prediction of Army-wide (ΔR^2 = .009, p < .01) and MOS-Specific (ΔR^2 = .020, p < .01) job knowledge test scores and ratings of Army-wide performance dimensions (ΔR^2 = .007, p < .01). Aside from attrition cognitions, which both the Baseline and Supplemental models significantly predicted (ΔR^2 = .003, p < .01, for the Supplemental model), the Supplemental model did not significantly predict attitudinal criteria, disciplinary incidents, or fitness test scores over the Baseline model.

We also considered 6-month attrition as a criterion. Results show the current configuration of the AFQT to predict attrition (R^2 = .002, p < .01) and the AO subtest to significantly contribute to its prediction (ΔR^2 = .001, p < .01). Although these statistics are small and likely significant due to a large sample size, these effects may be meaningful to the Army considering the cost of attrition.

Table 4.8 provides the subgroup analyses, where we calculated Cohen's *d* to compare the predictor composite scores for each protected group relative to their respective majority groups. This table also provides a comparison between the Baseline and Supplemental models in terms of subgroup differences.

The results show that adding the AO subtest to the AFQT resulted in a small increase in the subgroup differences for male-female comparisons, as well as black-white comparisons. This is consistent with previous research (Carey, 1994; Larson & Alderton, 1997; Russell & Peterson, 2001; Russell, Reynolds, & Campbell, 1994) suggesting that while these subgroups differ in AO subtest performance, the differences are smaller or comparable to those found on the AFQT. Carey's (1994) research also suggests that adding other mechanically-oriented ASVAB tests (as opposed to the AO subtest) would result in relatively larger subgroup differences.

Test score differences for the Hispanic-white comparison were slightly reduced when the AO subtest was added to the AFQT. This is consistent with previous findings showing that the AO subtest results in a smaller difference in test performance between Hispanics and Whites than that found with the AFQT (Alderton et al., 1997; Knapp & Heffner, 2009).

Table 4.8. Subgroup Differences for Predictor Composites.

	Baseline Model				*Supplemental Model*				
	n	*M*	*SD*	*d*	*n*	*M*	*SD*	*d*	*Difference in ds*
Female	1649	50.43	3.53	-.10	1649	50.97	3.58	-.14	-.04
Male	6599	50.82	3.85	-	6599	51.52	3.92	-	-
Black	1184	49.29	3.50	-.52	1184	49.55	3.64	-.63	-.11
Hispanic	1267	49.98	3.53	-.34	1267	50.85	3.56	-.29	.05
White	5293	51.24	3.78	-	5293	51.94	3.82	-	-

Note: * p < .01. A positive d indicates that the protected group scored higher than the majority group.

Table 4.9 summarizes the results of our examination of differential prediction across the nine criteria where the Baseline or Supplemental predictor composites predicted a significant portion of the variance in those criteria. This table shows that while differential prediction occurs often across both the Baseline and Supplemental models, this differential prediction usually favors the minority group. The Baseline and Supplemental predictor composites share three exceptions to this:

- Underprediction of the training course average of females when the male regression line is used
- Weak prediction of the Army-Wide Performance Ratings (PRS) at EOT of Hispanics relative to Whites

Table 4.9. Summary of Differential Prediction Findings across the Criteria Significantly Predicted by Baseline or Supplemental Predictor Composites.

	Baseline Model						*Supplemental Model*					
	Main Effect			*Interaction*			*Main Effect*			*Interaction*		
	ns	+	-	ns	+	-	ns	+	-	ns	+	-
Female	3	4	1	8	1	0	4	4	1	7	2	0
Black	9	5	0	8	0	0	4	5	0	9	0	0
Hispanic	7	2	0	8	1	0	7	2	0	8	0	1

Note: Cells indicate the number of Main Effects or Interactions that were significant, and the direction of those differences. "ns" = not significant, "+" = significant and favoring the minority group, and "-" = significant and favoring the majority group.

Figures 4.1 through 4.3 provide details on the differential prediction analyses for the criteria associated with these three comparisons, including:

- The intercept and slope values for the regression lines when the Baseline and Supplemental predictor composites are used to predict the criterion, where these regressions are reported for the following groups: Total, Male, Female, White, Black, and Hispanic.
- An indication of the direction and significance of the main effect, where a significant main effect indicates that adding subgroup membership improves prediction of the criterion. A "+" indicates that minority group membership tends to be associated with overestimates in criterion performance. A "-" indicates that minority group membership

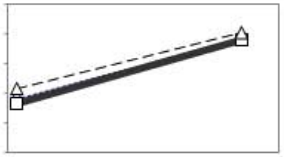
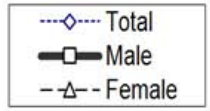
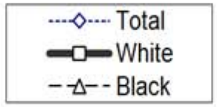
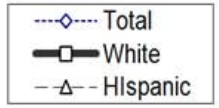
tends to be associated with underestimates in criterion performance. These differences were tested for female to male comparisons, black to white comparisons, and Hispanic to white comparisons.

- An indication of the direction and significance of the interaction effect (indicating a difference in the slope of the relationship between the predictor and criterion across subgroups). A "+" indicates overprediction of minority group performance on the criterion, where the relationship between the predictor and the criterion is greater for the minority group than it is for the majority group. A "-" indicates underprediction of minority group performance on the criterion, where the relationship between the predictor and the criterion is less for the minority group than it is for the majority group.
- A graphical representation of each regression line for each set of comparisons.
- Differences in predicted scores at different points along the regression line, where the subgroup regression lines were compared at several points along the line and the differences in the predicted scores are reported in terms of standard deviation units. Positive values indicate that the majority group regression line overestimates the criterion score for the minority group at the given point in question.
- Comparisons of differences in predicted scores when moving from the Baseline to the Supplemental model. Positive values indicate that the Supplemental model provides an estimate of criterion performance that is more favorable to the minority group than what would be provided with the Baseline model.

The appendix provides these details for the remainder of the criteria.

In general, these findings suggest that, where the criterion is effectively predicted by the Baseline or Supplemental model, 1) use of the Supplemental model would result in few cases where differential prediction is unfavorable to minorities, and 2) relative to the Baseline model, the Supplemental model has no adverse effect on the level of differential prediction with any of the subgroup comparisons, and in one case, actually results introducing differential prediction favorable to females.

Figure 4.1: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT RITMS: course average, standardized within MOS

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,187	-3.52	0.07				-3.56	0.07				
Male	855	-3.82	0.07				-3.94	0.08				
Female	332	-3.15	0.06	-	ns		-3.04	0.06	-	ns		
White	620	-3.16	0.06				-3.32	0.06				
Black	308	-3.85	0.08	ns	ns		-3.96	0.08	ns	ns		
Hispanic	190	-3.91	0.08	ns	ns		-4.03	0.08	ns	ns		

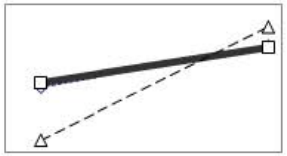
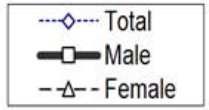
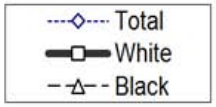
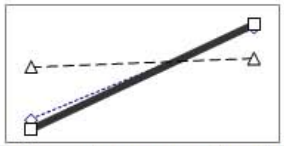
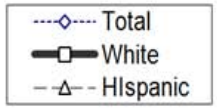
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.69	-.42	.14	-.54	-.64	-.62	-.08	-.07	-.72	-.42	.16	-.59	-.65	-.67	-.05	-.07	.02	.03	.00
-1 SD	-.41	-.18	.12	-.32	-.36	-.34	-.04	-.02	-.43	-.18	.14	-.34	-.35	-.36	.00	-.02	.01	.03	.00
Mean	-.14	.06	.10	-.09	-.08	-.05	.01	.03	-.14	.06	.11	-.10	-.05	-.06	.04	.03	.00	.03	.00
+1 SD	.13	.29	.09	.14	.20	.24	.05	.08	.15	.30	.08	.15	.25	.25	.09	.08	-.01	.04	.00
+2 SD	.41	.53	.07	.37	.48	.52	.10	.13	.44	.53	.05	.40	.55	.55	.14	.13	-.02	.04	.00
n	855	332		620	308	190			855	332		620	308	190					
M	-.12	.01		-.03	-.18	-.13			-.12	.01		-.03	-.18	-.13					
SD	1.02	.89		.89	.96	1.32			1.02	.89		.89	.96	1.32					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.2: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT PRS: Sum of AW Scores

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,758	-2.15	0.04				-2.39	0.05				
Male	1,608	-1.94	0.04				-2.14	0.04				
Female	150	-6.10	0.11	+	+		-6.92	0.13	+	+		
White	1,279	-2.45	0.05				-2.69	0.05				
Black	113	-3.26	0.06	ns	ns		-3.43	0.06	ns	ns		
Hispanic	273	-0.17	0.00	ns	-		-0.28	0.01	ns	-		

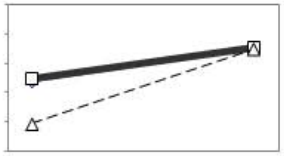
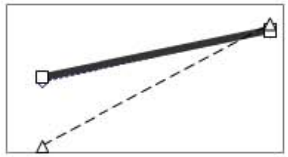
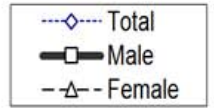
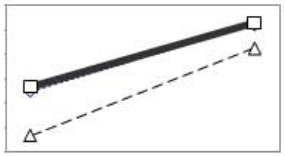
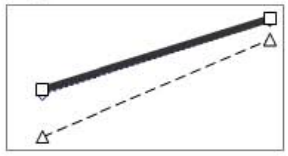
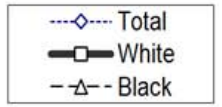
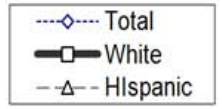
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.30	-1.19	-.26	-.41	-.57	.00	-.15	.37	-.33	-1.28	-.28	-.45	-.57	-.02	-.12	.39	-.02	.03	.02
-1 SD	-.16	-.77	-.18	-.24	-.35	.01	-.10	.23	-.17	-.81	-.18	-.26	-.33	.00	-.07	.24	-.01	.03	.01
Mean	-.02	-.36	-.10	-.07	-.12	.02	-.05	.08	-.02	-.33	-.09	-.07	-.09	.02	-.02	.08	.01	.03	.00
+1 SD	.12	.05	-.02	.10	.10	.04	.00	-.06	.13	.14	.00	.12	.15	.05	.03	-.07	.02	.02	-.01
+2 SD	.26	.47	.06	.27	.33	.05	.05	-.20	.28	.61	.10	.31	.39	.07	.07	-.22	.03	.02	-.02
n	1608	150		1279	113	273			1608	150		1279	113	273					
M	-.02	-.32		-.05	-.23	.02			-.02	-.32		-.05	-.23	.02					
SD	.99	1.13		1.01	1.06	.94			.99	1.13		1.01	1.06	.94					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.3: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT PRS: Sum of standardized MOS scores

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,728	-1.93	0.04				-2.13	0.04				
Male	1,579	-1.79	0.04				-1.96	0.04				
Female	149	-4.81	0.09	+	ns		-5.52	0.10	+	+		
White	1,256	-1.82	0.04				-2.03	0.04				
Black	111	-2.79	0.05	+	ns		-3.09	0.05	+	ns		
Hispanic	272	-0.90	0.02	ns	ns		-0.89	0.02	ns	ns		

Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.26	-1.03	-.23	-.26	-.67	-.17	-.39	.08	-.28	-1.11	-.25	-.29	-.68	-.18	-.38	.11	-.02	.01	.02
-1 SD	-.13	-.71	-.17	-.13	-.49	-.11	-.35	.02	-.14	-.74	-.18	-.15	-.48	-.12	-.32	.03	-.01	.02	.01
Mean	.00	-.39	-.12	.00	-.31	-.05	-.30	-.05	.00	-.37	-.11	.00	-.28	-.06	-.27	-.05	.01	.03	-.01
+1 SD	.13	-.07	-.06	.13	-.13	.01	-.25	-.11	.14	.00	-.04	.14	-.08	.00	-.21	-.13	.02	.04	-.02
+2 SD	.26	.25	-.01	.26	.05	.07	-.21	-.17	.29	.37	.02	.29	.13	.06	-.16	-.21	.03	.05	-.04
n	1579	149		1256	111	272			1579	149		1256	111	272					
M	.00	-.36		.01	-.40	-.07			.00	-.36		.01	-.40	-.07					
SD	.97	1.14		1.00	.97	.96			.97	1.14		1.00	.97	.96					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Evaluation of Substitutional Predictor Composites

We conducted a series of analyses to compare the four possible Substitutional predictor composites to the unit weighted composite used to form the AFQT.

Table 4.10 compares the regression results when substituting the AO subtest for each of the AFQT subtests (i.e., substituting AO for each subtest one at a time). The results show that substituting the AO subtest for any one of the subtests used to form the AFQT either improved or had a negligible impact on prediction of 19 out of 20 criterion scores examined. The exception to this was the RITMS course average where substitution with AO led to small decrements (R^2 was .009 to .020 lower) in the prediction of criterion scores.

Some of the strongest gains were in the prediction of job knowledge at EOT and IU when the AO subtest was substituted for either the AR subtest or the MK subtests. This is consistent with the research of Mayberry and Hiatt (1990) who found that the AO subtest is an effective predictor of job knowledge test scores across a variety of Army jobs. Substituting the AO subtest for the WK subtest or the PC subtest led to slight improvements in the prediction of performance ratings at EOT and IU (e.g., differences in R^2 = .009).

Table 4.11 compares the subgroup differences found when substituting the AO subtest for each of the AFQT subtests one at a time. These findings show that substituting the AO subtest had little to no impact on subgroup differences in most cases. The only case where there was a meaningful impact was the case where the AO subtest was substituted for the MK subtest: this resulted in a small increase in the subgroup differences between Whites and Blacks on the predictor composite. To reduce the subgroup differences for both Blacks and Females, it appears that the best strategy would be to substitute the AO subtest for the AR subtest, while the most effective strategy for reducing subgroup differences for Hispanics would be to substitute the AO subtest for the WK subtest.

Figures 4.4 through 4.10 provide the bias analyses for the six occasions where the percentage of variance predicted in the criterion was at least 1 percent greater for the Substitutional predictor composite relative to the Baseline predictor composite. These figures show that while differential prediction occurs often across both the Baseline and Substitutional models, this differential prediction usually favors the minority group. There is only one case, where the differential prediction finding favors the majority group. The result, in that case, is a weak prediction of the Army-Wide Performance Ratings (PRS) at EOT of Hispanics relative to Whites. This result occurs to the same degree with the Baseline model as it does with the Substitutional model. These findings suggest that, where the criterion is effectively incrementally predicted by substituting the AO subtest, there is no additional adverse effect on the level of differential prediction with any of the subgroup comparisons.

Table 4.10. Regression Results when Substituting Assembling Objects (AO) for Each of the AFQT Subtests.

Criterion		*n*	Predictor Composite	Subtest Dropped	*R*	R_A	R_C	R_{CA}	R^2_{CA}	*Diff* R^2_{CA}
ID	Name									
End-of-Training										
1	MOS Job Knowledge (JKT)	1603	AR+MK+WK+PC	na	.357	.356	.561	.560	.314	NA
			AO+MK+WK+PC	AR	.380	.378	.572	.571	.326	.012
			AR+AO+WK+PC	MK	.385	.383	.575	.574	.330	.016
			AR+MK+AO+PC	WK	.340	.338	.548	.548	.300	-.014
			AR+MK+WK+AO	PC	.355	.354	.559	.558	.311	-.003
2	Course Average (RITMS)	1187	AR+MK+WK+PC	na	.262	.259	.422	.420	.177	NA
			AO+MK+WK+PC	AR	.245	.242	.411	.410	.168	-.009
			AR+AO+WK+PC	MK	.222	.218	.397	.396	.157	-.020
			AR+MK+AO+PC	WK	.247	.244	.411	.409	.168	-.009
			AR+MK+WK+AO	PC	.236	.232	.406	.404	.163	-.013
3	Army-Wide Performance Ratings (PRS)	1758	AR+MK+WK+PC	na	.155	.151	.258	.256	.066	NA
			AO+MK+WK+PC	AR	.166	.163	.265	.263	.069	.003
			AR+AO+WK+PC	MK	.163	.160	.263	.261	.068	.002
			AR+MK+AO+PC	WK	.189	.186	.279	.277	.077	.011
			AR+MK+WK+AO	PC	.183	.180	.275	.273	.075	.009
4	MOS Performance Ratings (PRS)	1728	AR+MK+WK+PC	na	.140	.136	.237	.235	.055	NA
			AO+MK+WK+PC	AR	.150	.147	.243	.241	.058	.003
			AR+AO+WK+PC	MK	.152	.148	.244	.242	.059	.003
			AR+MK+AO+PC	WK	.161	.157	.250	.247	.061	.006
			AR+MK+WK+AO	PC	.167	.163	.253	.251	.063	.008
5	Affective Commitment (ALQ)	1705	AR+MK+WK+PC	na	.031	.000	.058	.046	.002	NA
			AO+MK+WK+PC	AR	.020	.000	.050	.036	.001	-.001
			AR+AO+WK+PC	MK	.027	.000	.055	.043	.002	.000
			AR+MK+AO+PC	WK	.015	.000	.047	.032	.001	-.001
			AR+MK+WK+AO	PC	.029	.000	.056	.045	.002	.000
6	Army and MOS Fit (ALQ)	1705	AR+MK+WK+PC	na	.015	.000	.021	.000	.000	NA
			AO+MK+WK+PC	AR	.020	.000	.024	.000	.000	.000
			AR+AO+WK+PC	MK	.002	.000	.013	.000	.000	.000
			AR+MK+AO+PC	WK	.033	.000	.033	.000	.000	.000
			AR+MK+WK+AO	PC	.016	.000	.021	.000	.000	.000
7	Number of disciplinary incidents during Army training (ALQ)	1705	AR+MK+WK+PC	na	.051	.038	.087	.080	.006	NA
			AO+MK+WK+PC	AR	.061	.050	.092	.086	.007	.001
			AR+AO+WK+PC	MK	.067	.058	.096	.090	.008	.002
			AR+MK+AO+PC	WK	.068	.058	.097	.091	.008	.002
			AR+MK+WK+AO	PC	.059	.048	.091	.085	.007	.001

Table 4.10. Regression Results when Substituting Assembling Objects (AO) for Each of the AFQT Subtests. (Continued)

Criterion		*n*	Predictor Composite	Subtest Dropped	*R*	R_A	R_C	R_{CA}	R^2_{CA}	*Diff* R^2_{CA}
ID	Name									
8	Physical Fitness (APFT)	1696	AR+MK+WK+PC	na	.060	.049	.091	.084	.007	NA
			AO+MK+WK+PC	AR	.033	.000	.075	.066	.004	-.003
			AR+AO+WK+PC	MK	.019	.000	.066	.056	.003	-.004
			AR+MK+AO+PC	WK	.087	.080	.109	.104	.011	.004
			AR+MK+WK+AO	PC	.060	.049	.092	.085	.007	.000
9	Attrition cognitions (ALQ)	1686	AR+MK+WK+PC	na	.044	.028	.070	.061	.004	NA
			AO+MK+WK+PC	AR	.041	.022	.068	.059	.004	.000
			AR+AO+WK+PC	MK	.042	.023	.069	.060	.004	.000
			AR+MK+AO+PC	WK	.053	.041	.076	.068	.005	.001
			AR+MK+WK+AO	PC	.033	.000	.063	.053	.003	-.001
10	Total Number of Recycles during BCT and/or AIT/OSUT (ATTRS)	7366	AR+MK+WK+PC	na	.042	.038	.068	.066	.004	NA
			AO+MK+WK+PC	AR	.047	.044	.071	.069	.005	.000
			AR+AO+WK+PC	MK	.039	.035	.066	.064	.004	.000
			AR+MK+AO+PC	WK	.050	.048	.074	.072	.005	.001
			AR+MK+WK+AO	PC	.040	.037	.067	.065	.004	.000
In-Unit										
11	Army Wide Job Knowledge (JKT)	992	AR+MK+WK+PC	na	.400	.398	.591	.590	.348	NA
			AO+MK+WK+PC	AR	.405	.403	.592	.591	.350	.002
			AR+AO+WK+PC	MK	.433	.431	.608	.607	.369	.021
			AR+MK+AO+PC	WK	.357	.354	.562	.561	.315	-.033
			AR+MK+WK+AO	PC	.390	.387	.583	.582	.339	-.009
12	MOS Job Knowledge (JKT)	474	AR+MK+WK+PC	na	.349	.344	.512	.509	.259	NA
			AO+MK+WK+PC	AR	.385	.380	.532	.530	.280	.022
			AR+AO+WK+PC	MK	.387	.382	.534	.531	.282	.023
			AR+MK+AO+PC	WK	.358	.353	.515	.512	.262	.004
			AR+MK+WK+AO	PC	.357	.351	.515	.512	.262	.004
13	Army-Wide Performance Ratings (PRS)	702	AR+MK+WK+PC	na	.076	.055	.116	.103	.011	NA
			AO+MK+WK+PC	AR	.118	.106	.144	.133	.018	.007
			AR+AO+WK+PC	MK	.107	.092	.136	.125	.016	.005
			AR+MK+AO+PC	WK	.137	.127	.157	.148	.022	.011
			AR+MK+WK+AO	PC	.118	.106	.144	.134	.018	.007
14	MOS Performance Ratings (PRS)	344	AR+MK+WK+PC	na	.060	.000	.094	.056	.003	NA
			AO+MK+WK+PC	AR	.081	.029	.108	.077	.006	.003
			AR+AO+WK+PC	MK	.096	.059	.118	.090	.008	.005
			AR+MK+AO+PC	WK	.117	.089	.133	.109	.012	.009
			AR+MK+WK+AO	PC	.115	.087	.130	.106	.011	.008

Table 4.10. Regression Results when Substituting Assembling Objects (AO) for Each of the AFQT Subtests. (Continued)

Criterion		*n*	Predictor Composite	Subtest Dropped	*R*	R_A	R_C	R_{CA}	R^2_{CA}	*Diff* R^2_{CA}
ID	Name									
15	Affective Commitment (ALQ)	1014	AR+MK+WK+PC	na	.066	.049	.108	.098	.010	NA
			AO+MK+WK+PC	AR	.050	.023	.097	.086	.007	-.002
			AR+AO+WK+PC	MK	.053	.028	.099	.088	.008	-.002
			AR+MK+AO+PC	WK	.050	.022	.096	.085	.007	-.002
			AR+MK+WK+AO	PC	.050	.023	.097	.086	.007	-.002
16	Army and MOS Fit (ALQ)	1014	AR+MK+WK+PC	na	.006	.000	.013	.000	.000	NA
			AO+MK+WK+PC	AR	.005	.000	.006	.000	.000	.000
			AR+AO+WK+PC	MK	.003	.000	.007	.000	.000	.000
			AR+MK+AO+PC	WK	.012	.000	.001	.000	.000	.000
			AR+MK+WK+AO	PC	.002	.000	.008	.000	.000	.000
17	Disciplinary Action Total Score (ALQ)	1014	AR+MK+WK+PC	na	.006	.000	.013	.000	.000	NA
			AO+MK+WK+PC	AR	.012	.000	.001	.000	.000	.000
			AR+AO+WK+PC	MK	.019	.000	.004	.000	.000	.000
			AR+MK+AO+PC	WK	.023	.000	.007	.000	.000	.000
			AR+MK+WK+AO	PC	.015	.000	.001	.000	.000	.000
18	Physical Fitness (APFT)	941	AR+MK+WK+PC	na	.056	.032	.083	.070	.005	NA
			AO+MK+WK+PC	AR	.052	.025	.081	.067	.005	.000
			AR+AO+WK+PC	MK	.055	.030	.083	.069	.005	.000
			AR+MK+AO+PC	WK	.084	.070	.102	.092	.008	.004
			AR+MK+WK+AO	PC	.053	.025	.082	.067	.005	.000
19	Attrition cognitions (ALQ)	1014	AR+MK+WK+PC	na	.042	.000	.063	.045	.002	NA
			AO+MK+WK+PC	AR	.067	.051	.080	.067	.004	.002
			AR+AO+WK+PC	MK	.074	.060	.085	.072	.005	.003
			AR+MK+AO+PC	WK	.084	.072	.092	.081	.007	.005
			AR+MK+WK+AO	PC	.069	.053	.081	.068	.005	.003
20	Attrition at six months (TTAS)	4529	AR+MK+WK+PC	na	.021	.000	.032	.024	.001	NA
			AO+MK+WK+PC	AR	.040	.034	.044	.039	.002	.001
			AR+AO+WK+PC	MK	.024	.011	.034	.027	.001	.000
			AR+MK+AO+PC	WK	.058	.054	.057	.053	.003	.002
			AR+MK+WK+AO	PC	.047	.042	.049	.044	.002	.001

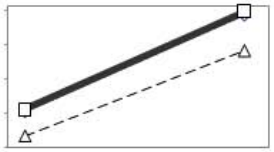
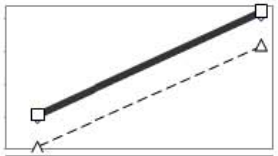
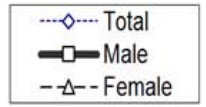
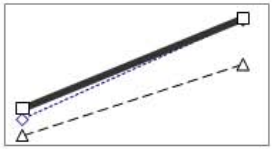
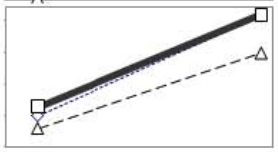
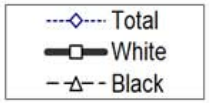
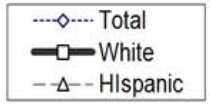
Note: R_A = uncorrected multiple *R* adjusted for shrinkage. R_C = corrected multiple R^2. R_{CA} = corrected multiple *R* adjusted for shrinkage. R^2_{CA} = corrected R^2 adjusted for shrinkage. *Diff* R^2_{CA} = difference between R^2_{CA}'s relative to the AFQT predictor composite (AR+MK+WK+PC).

Table 4.11. Subgroup Comparisons for Unit Weighted Substitutional Predictor Composites.

Predictor Composite	Subgroup	*n*	*M*	*SD*	*d*	*diff d*
AR+MK+WK+PC	Female	1530	201.71	14.07	-.10	NA
	Male	6214	203.24	15.32	-	NA
	Black	1184	197.15	13.98	-.52	NA
	Hispanic	1267	199.91	14.11	-.34	NA
	White	5293	204.96	15.11	-	NA
AO+MK+WK+PC	Female	1530	205.30	14.54	-.06	.04
	Male	6214	206.20	15.70	-	-
	Black	1184	199.30	15.08	-.57	-.05
	Hispanic	1267	204.07	14.62	-.26	.08
	White	5293	207.99	15.29	-	-
AR+AO+WK+PC	Female	1530	202.58	15.29	-.24	-.14
	Male	6214	206.55	16.48	-	-
	Black	1184	196.42	15.68	-.76	-.24
	Hispanic	1267	202.97	15.01	-.35	-.01
	White	5293	208.52	15.88	-	-
AR+MK+AO+PC	Female	1530	205.96	15.88	-.11	-.01
	Male	6214	207.80	17.10	-	-
	Black	1184	200.48	16.29	-.53	.00
	Hispanic	1267	206.22	16.00	-.18	.15
	White	5293	209.29	16.78	-	-
AR+MK+WK+AO	Female	1530	203.73	15.73	-.15	-.05
	Male	6214	206.22	17.11	-	-
	Black	1184	197.70	15.99	-.62	-.09
	Hispanic	1267	203.86	15.70	-.25	.09
	White	5293	207.97	16.74	-	-

Note: *d = Cohen's d* relative to the comparison group; *diff d* = the difference in the *d*'s for that subgroup comparison relative to the AFQT predictor composite (AR+MK+WK+PC)

Figure 4.4: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT JKT: Army Class score standardized within MOS

		Predictor: AR+MK+WK+PC					Predictor: AO+MK+WK+PC					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,603	-5.03	0.02				-5.68	0.03				
Male	1,468	-5.07	0.02				-5.70	0.03				
Female	135	-4.93	0.02	+	ns		-6.07	0.03	+	ns		
White	1,176	-4.51	0.02				-5.13	0.02				
Black	97	-4.10	0.02	+	ns		-4.72	0.02	+	ns		
Hispanic	244	-5.49	0.03	+	ns		-6.34	0.03	+	ns		

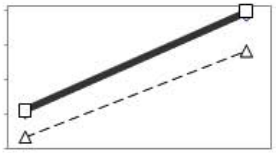
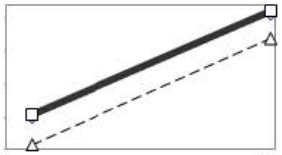
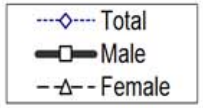
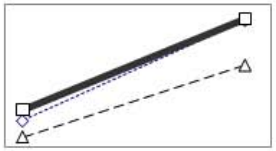
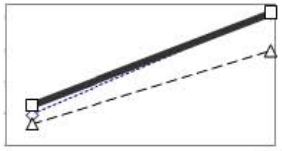
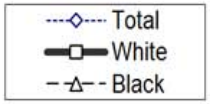
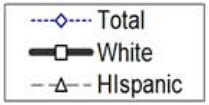
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The total regression line is often obscured by the fact that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs Across Models		
	Predictor: AR+MK+WK+PC								Predictor: AO+MK+WK+PC										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.79	-1.18	-.12	-.68	-1.06	-.99	-.38	-.30	-.87	-1.35	-.15	-.75	-1.10	-1.12	-.35	-.35	-.03	.04	-.06
-1 SD	-.43	-.87	-.13	-.36	-.80	-.61	-.45	-.24	-.47	-.97	-.15	-.39	-.80	-.70	-.41	-.29	-.02	.04	-.04
Mean	-.08	-.56	-.14	-.04	-.55	-.24	-.52	-.19	-.07	-.58	-.15	-.03	-.50	-.27	-.47	-.22	-.01	.05	-.03
+1 SD	.28	-.24	-.16	.28	-.29	.14	-.58	-.14	.32	-.19	-.15	.33	-.20	.16	-.53	-.16	.00	.05	-.02
+2 SD	.64	.07	-.17	.61	-.04	.52	-.65	-.08	.72	.20	-.16	.69	.09	.59	-.60	-.09	.01	.05	-.01
n	1468	135		1176	97	244			1468	135		1176	97	244					
M	-.08	-.53		.00	-.68	-.33			-.08	-.53		.00	-.68	-.33					
SD	.98	.84		.97	.94	.91			.98	.84		.97	.94	.91					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.5: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT JKT: Army Class score standardized within MOS

		Predictor: AR+MK+WK+PC					Predictor: AR+AO+WK+PC					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,603	-5.03	0.02				-5.38	0.03				
Male	1,468	-5.07	0.02				-5.35	0.03				
Female	135	-4.93	0.02	+	ns		-5.89	0.03	+	ns		
White	1,176	-4.51	0.02				-4.89	0.02				
Black	97	-4.10	0.02	+	ns		-4.26	0.02	+	ns		
Hispanic	244	-5.49	0.03	+	ns		-6.13	0.03	+	ns		

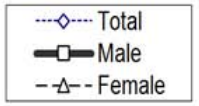
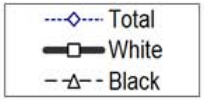
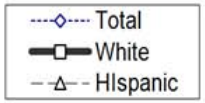
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The total regression line is often obscured by the fact that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total Group	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs Across Models		
	Predictor: AR+MK+WK+PC								Predictor: AR+AO+WK+PC									
	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.79	-1.18	-.12	-.68	-1.06	-.99	-.38	-.30	-.87	-1.32	-.14	-.77	-1.05	-1.13	-.28	-.34	-.02	.10	-.05
-1 SD	-.43	-.87	-.13	-.36	-.80	-.61	-.45	-.24	-.47	-.92	-.13	-.40	-.76	-.69	-.36	-.27	.00	.09	-.03
Mean	-.08	-.56	-.14	-.04	-.55	-.24	-.52	-.19	-.08	-.52	-.13	-.04	-.48	-.25	-.44	-.20	.01	.08	-.01
+1 SD	.28	-.24	-.16	.28	-.29	.14	-.58	-.14	.32	-.12	-.13	.32	-.20	.19	-.52	-.12	.03	.06	.01
+2 SD	.64	.07	-.17	.61	-.04	.52	-.65	-.08	.71	.29	-.13	.68	.09	.63	-.60	-.05	.04	.05	.03
n	1468	135		1176	97	244			1468	135		1176	97	244					
M	-.08	-.53		.00	-.68	-.33			-.08	-.53		.00	-.68	-.33					
SD	.98	.84		.97	.94	.91			.98	.84		.97	.94	.91					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.6: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT PRS: Sum of AW Scores

		Predictor: AR+MK+WK+PC					Predictor: AR+MK+AO+PC					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,758	-2.15	0.01				-2.41	0.01				
Male	1,608	-1.94	0.01				-2.17	0.01				
Female	150	-6.10	0.03	+	+		-6.95	0.03	+	+		
White	1,279	-2.45	0.01				-2.68	0.01				
Black	113	-3.26	0.02	ns	ns		-3.41	0.02	ns	ns		
Hispanic	273	-0.17	0.00	+	-		-0.34	0.00	+	-		

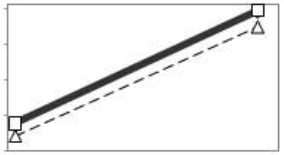
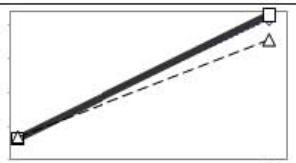
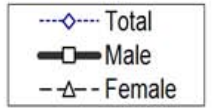
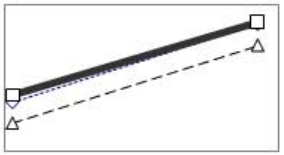
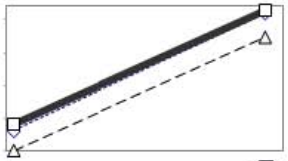
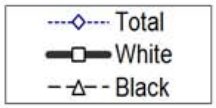
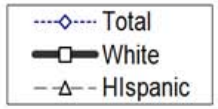
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The total regression line is often obscured by the fact that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs Across Models		
	Predictor: AR+MK+WK+PC								Predictor: AR+MK+AO+PC										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.30	-1.19	-.26	-.41	-.57	.00	-.15	.37	-.36	-1.36	-.29	-.47	-.62	-.03	-.14	.40	-.03	.01	.03
-1 SD	-.16	-.77	-.18	-.24	-.35	.01	-.10	.23	-.19	-.84	-.19	-.27	-.36	-.01	-.09	.24	-.01	.01	.01
Mean	-.02	-.36	-.10	-.07	-.12	.02	-.05	.08	-.02	-.33	-.09	-.06	-.10	.02	-.04	.08	.01	.01	-.01
+1 SD	.12	.05	-.02	.10	.10	.04	.00	-.06	.15	.19	.01	.14	.15	.05	.01	-.08	.03	.01	-.02
+2 SD	.26	.47	.06	.27	.33	.05	.05	-.20	.32	.71	.11	.35	.41	.08	.06	-.24	.05	.01	-.04
n	1608	150		1279	113	273			1608	150		1279	113	273					
M	-.02	-.32		-.05	-.23	.02			-.02	-.32		-.05	-.23	.02					
SD	.99	1.13		1.01	1.06	.94			.99	1.13		1.01	1.06	.94					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.7: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU JKT: AW Army Class total score

		Predictor: AR+MK+WK+PC					Predictor: AR+AO+WK+PC					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	992	-5.41	0.03				-5.55	0.03				
Male	825	-5.34	0.03				-5.69	0.03				
Female	167	-5.39	0.02	+	ns		-4.70	0.02	+	ns		
White	616	-5.02	0.02				-5.31	0.03				
Black	125	-5.82	0.03	-	ns		-5.67	0.02	-	ns		
Hispanic	174	-3.67	0.02	+	ns		-3.96	0.02	+	ns		

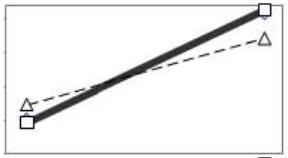
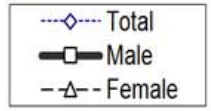
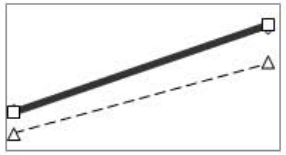
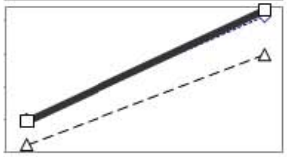
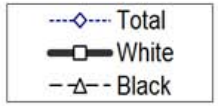
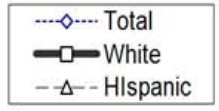
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The total regression line is often obscured by the fact that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs Across Models		
	Predictor: AR+MK+WK+PC								Predictor: AR+AO+WK+PC									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.96	-1.14	-.08	-.85	-1.42	-.75	-.52	.10	-1.09	-1.06	.01	-.98	-1.38	-.83	-.36	.14	.09	.16	.05
-1 SD	-.56	-.76	-.08	-.47	-1.02	-.48	-.50	-.01	-.63	-.70	-.03	-.55	-.95	-.52	-.36	.03	.05	.13	.04
Mean	-.16	-.37	-.09	-.09	-.62	-.22	-.48	-.12	-.17	-.34	-.07	-.12	-.53	-.21	-.37	-.09	.02	.11	.03
+1 SD	.24	.01	-.09	.28	-.22	.05	-.46	-.23	.29	.02	-.11	.31	-.10	.10	-.37	-.20	-.02	.09	.03
+2 SD	.63	.40	-.10	.66	.17	.31	-.44	-.34	.74	.38	-.15	.74	.33	.41	-.37	-.31	-.05	.07	.02
n	825	167		616	125	174			825	167		616	125	174					
M	-.15	-.42		-.02	-.78	-.30			-.15	-.42		-.02	-.78	-.30					
SD	.98	.97		.93	1.07	.88			.98	.97		.93	1.07	.88					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.8: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU JKT: Overall Army Class total score standardized within MOS

		Predictor: AR+MK+WK+PC					Predictor: AO+MK+WK+PC					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	474	-4.56	0.02				-5.18	0.02				
Male	421	-4.74	0.02				-5.45	0.03				
Female	53	-3.21	0.01	+	ns		-3.33	0.02	+	ns		
White	336	-4.81	0.02				-5.53	0.03				
Black	32	-4.42	0.02	+	ns		-5.04	0.02	+	ns		
Hispanic	75	-3.85	0.02	+	ns		-4.10	0.02	+	ns		

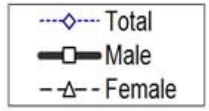
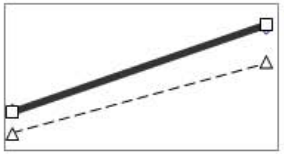
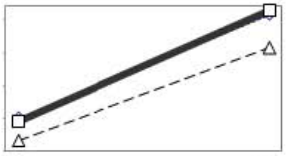
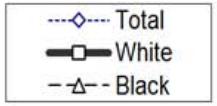
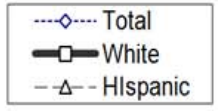
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The total regression line is often obscured by the fact that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs Across Models		
	Predictor: AR+MK+WK+PC								Predictor: AO+MK+WK+PC									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.84	-.64	.07	-.84	-1.21	-.64	-.33	.18	-.93	-.66	.09	-.93	-1.28	-.69	-.32	.22	.02	.01	.04
-1 SD	-.47	-.39	.03	-.47	-.90	-.34	-.39	.11	-.51	-.42	.03	-.50	-.94	-.37	-.39	.12	.01	.00	.01
Mean	-.10	-.15	-.02	-.09	-.60	-.04	-.46	.05	-.10	-.17	-.02	-.08	-.60	-.06	-.46	.02	-.01	.00	-.03
+1 SD	.27	.09	-.06	.29	-.29	.27	-.52	-.02	.32	.07	-.08	.34	-.25	.25	-.53	-.08	-.02	-.01	-.06
+2 SD	.64	.34	-.10	.66	.01	.57	-.59	-.08	.73	.32	-.14	.76	.09	.57	-.60	-.18	-.04	-.01	-.09
n	421	53		336	32	75			421	53		336	32	75					
M	-.10	-.18		-.03	-.70	-.15			-.10	-.18		-.03	-.70	-.15					
SD	1.03	.90		1.02	1.10	.86			1.03	.90		1.02	1.10	.86					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.9: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT RITMS: course average, standardized within MOS

	Predictor: AR+MK+WK+PC						Predictor: AR+AO+WK+PC					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	474	-4.56	0.02				-4.90	0.02				
Male	421	-4.74	0.02				-5.21	0.02				
Female	53	-3.21	0.01	+	ns		-3.03	0.01	+	ns		
White	336	-4.81	0.02				-5.23	0.02				
Black	32	-4.42	0.02	+	ns		-4.83	0.02	+	ns		
Hispanic	75	-3.85	0.02	+	ns		-4.14	0.02	+	ns		

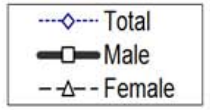
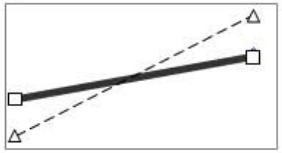
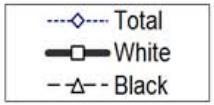
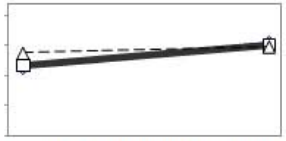
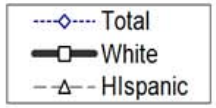
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The total regression line is often obscured by the fact that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total Group	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs Across Models		
	Predictor: AR+MK+WK+PC								Predictor: AR+AO+WK+PC									
	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.84	-.64	.07	-.84	-1.21	-.64	-.33	.18	-.95	-.60	.11	-.94	-1.23	-.71	-.26	.21	.05	.07	.03
-1 SD	-.47	-.39	.03	-.47	-.90	-.34	-.39	.11	-.53	-.36	.05	-.52	-.88	-.38	-.32	.13	.03	.07	.02
Mean	-.10	-.15	-.02	-.09	-.60	-.04	-.46	.05	-.11	-.12	.00	-.10	-.52	-.04	-.38	.05	.01	.08	.00
+1 SD	.27	.09	-.06	.29	-.29	.27	-.52	-.02	.31	.12	-.06	.32	-.17	.30	-.44	-.02	-.01	.08	-.01
+2 SD	.64	.34	-.10	.66	.01	.57	-.59	-.08	.73	.35	-.12	.74	.18	.63	-.50	-.10	-.02	.08	-.02
n	421	53		336	32	75			421	53		336	32	75					
M	-.10	-.18		-.03	-.70	-.15			-.10	-.18		-.03	-.70	-.15					
SD	1.03	.90		1.02	1.10	.86			1.03	.90		1.02	1.10	.86					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure 4.10: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU PRS: Sum of AW scores

	Predictor: AR+MK+WK+PC						Predictor: AR+MK+AO+PC					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	702	-0.97	0.00				-1.37	0.01				
Male	595	-0.80	0.00				-1.30	0.01				
Female	107	-1.89	0.01	+	ns		-1.72	0.01	+	ns		
White	426	-0.92	0.00				-1.13	0.01				
Black	99	-2.53	0.01	ns	ns		-2.60	0.01	ns	ns		
Hispanic	119	0.92	0.00	+	ns		-0.30	0.00	+	ns		

Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The total regression line is often obscured by the fact that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs Across Models		
	Predictor: AR+MK+WK+PC								Predictor: AR+MK+AO+PC									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.17	-.31	-.05	-.19	-.42	.15	-.21	.30	-.25	-.31	-.02	-.23	-.44	-.02	-.19	.19	.03	.02	-.11
-1 SD	-.11	-.17	-.02	-.12	-.23	.08	-.10	.18	-.15	-.17	-.01	-.14	-.23	.00	-.08	.13	.02	.02	-.05
Mean	-.05	-.03	.01	-.06	-.04	.01	.02	.06	-.05	-.03	.01	-.06	-.02	.03	.04	.08	.00	.02	.02
+1 SD	.00	.12	.04	.01	.15	-.06	.13	-.06	.05	.10	.02	.03	.19	.05	.15	.02	-.02	.02	.08
+2 SD	.06	.26	.08	.07	.33	-.13	.24	-.18	.15	.24	.04	.12	.40	.08	.26	-.03	-.04	.02	.15
n	595	107		426	99	119			595	107		426	99	119					
M	-.05	-.05		-.05	-.10	.03			-.05	-.05		-.05	-.10	.03					
SD	1.00	1.01		.94	.99	1.14			1.00	1.01		.94	.99	1.14					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

CHAPTER 5: CONCLUSIONS

There were several important findings in the current study. The foremost contribution of the present study was to show that adding the AO subtest to the AFQT has the ability to improve prediction of important criteria. These criteria include MOS-Specific job knowledge and performance on Army-wide and MOS-Specific performance dimensions. Adding the AO subtest to the AFQT improved prediction of these criteria both at EOT and after 18 to 20 months of TIS. There was also a slight improvement in the prediction of attrition at six months. This improvement in prediction occurred in a similar manner across subgroups, with no decrements in the strength or accuracy of prediction for Females, Blacks, or Hispanics relative to their majority comparison groups.

The Army may wish to consider including the AO subtest in the AFQT predictor composite. Adding the AO subtest to the AFQT composite would likely increase the prediction of performance and job knowledge. Although our analyses showed an increase in the differences among subgroups when the AO subtest was added to the AFQT composite, this finding should be interpreted with caution. The incumbent sample used in this investigation is affected by differential range restriction across subgroups. That is, the variance of the AFQT scores is reduced (or restricted) in the incumbent sample (relative to the applicant population) and the degree of that range restriction differs across subgroups. This makes it difficult to predict applicant subgroup differences based on incumbent data. Thus, we recommend that future investigations examine subgroup differences using an applicant sample. It is interesting to note, however, that even when we use this incumbent sample to examine the impact of any additional subgroup differences, we found that those differences were reflected in performance and job knowledge criteria, suggesting that the revised AFQT composite would be fair and unbiased to minority groups.

The findings of this investigation are limited by several factors. First, a limited set of MOSs were included in the investigation, which might affect the generalizability of the results. However, the six target MOSs are high density in the Army and are fairly representative of most other MOS in terms of job characteristics, according to Moriarty, Campbell, Heffner, and Knapp (2009). Therefore, we expect the results to have utility for understanding the prediction of performance across MOS. Second, subgroup membership was particularly low for several variables in the IU timeframe, and this may have led to type II error in detection of any interactions between the predictor and the subgroup variable in predicting the criterion. Third, significance test findings need to be interpreted with caution because there are two factors inherent in this investigation that increase experiment-wise error: the large number of significance tests conducted, and the large sample sizes. To address this issue, we set the alpha level at .01 (less than the traditional .05) and we provided effect size estimates to aid in interpreting the meaningfulness of the statistics.

Future research should examine how adding the AO subtest to the AFQT could affect prediction for performance of different types of jobs. Using the AO subtest as a supplement to the current AFQT composite, or as a replacement for one of the subtests would likely provide a greater improvement in the prediction of job performance for some jobs than it would for others. In particular, the jobs that are most likely to see improved prediction with the addition of the AO

subtest are jobs whose tasks require spatial aptitude (e.g., Light Wheel Vehicle Mechanic [91B]). To determine whether spatial aptitude requirements moderate the predictive power of the AFQT, we suggest conducting a series of moderated multiple regressions. The moderator in these regressions could be defined by coding each job for the amount of spatial aptitude required. This research might also examine how prediction of technical training is differentially affected in these types of jobs. Documentation of improved prediction in technical training restarts/attrition would provide great benefit to the Army in terms of improved assignment and reductions in Soldier attrition.

Future research should also systematically examine the financial utility of adding the AO subtest to the AFQT. Changes in criterion outcomes and subgroup differences could be subjected to economic modeling to help better understand the overall impact of adding the AO subtest. This research could also be used used to examine the impact of using different methods to combine or use subtest scores. For example, using the scores in a non-compensatory manner might serve to reduce subgroup differences and have relatively little impact on validity.

REFERENCES

Abrahams, N. M., Pass, J. J., Kusulas, J. W., Cole, D. R, & Kieckhaefer, W. F. (1993). *Incremental validity of experimental computerized tests for predicting training criteria in military technical schools.* San Diego, CA: Navy Personnel Research and Development Center.

Alderton, D., Wolfe, J., & Larson, G. (1997). The ECAT battery. *Military Psychology, 9*(1), 5-37.

Allen, M. T., Caramagno, J. P., Ingerick, M., & Heffner, T. S. (2010). *Army Class Quarterly Attrition Analysis: May 2010* (Final Report 10-13). Alexandria, VA: U.S. Army Institute for the Behavioral and Social Sciences.

Bertua, C., Anderson, N., & Salgado, J. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology, 78*(3), 387-409.

Busciglio, H, H. (1990). *The Incremental Validity of Spatial and Perceptual-Psychomotor Tests Relative to the Armed Services Vocational Aptitude Battery* (Technical Report 883). Alexandria, VA: U.S. Army Institute for the Behavioral and Social Sciences.

Busciglio, H, H. (1991). *The Usefulness of Project A Spatial Tests for Predicting Comprehensive Performance Measures* (Research Note 91-17). Alexandria, VA: U.S. Army Institute for the Behavioral and Social Sciences.

Busciglio, H. H., & Palmer, D. R. (1996). *An Empirical Assessment of Coaching and Practice Effects on Three Army Tests of Spatial Aptitude* (Research Note 96-70). Alexandria, VA: U.S. Army Institute for the Behavioral and Social Sciences.

Carey, N. B. (1994). Computer predictors of mechanical job performance: Marine Corps findings. *Military Psychology, 6*(1), 1-30.

Carretta, T. R., & King, R. E. (2008). Improved military air traffic controller selection methods as measured by subsequent training performance. *Aviation, Space, and Environmental Medicine, 79,* 36–43.

Cleary, T. A. (1968). Test bias: prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement, 5,* 115-124.

Deary, I., Irwing, P., Der, G., & Bates, T. (2007). Brother-sister differences in the g factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence, 35*(5), 451-456.

Defense Manpower Data Center (2006). *CAT-ASVAB forms 1 & 2* (Technical Bulletin No. 1). Seaside, CA.

Defense Manpower Data Center (2008). *CAT-ASVAB forms 5 - 9* (Technical Bulletin No. 3). Seaside, CA.

Defense Manpower Data Center (2009). *CAT-ASVAB forms 3 & 4* (Technical Bulletin No. 2). Seaside, CA.

Department of Defense. (1994). Technical manual for the ASVAB 18/19 Career Exploration Program. Washington, DC: US Government Printing Office.

Driskell, J., Hogan, J., Salas, E., & Hoskin, B. (1994). Cognitive and personality predictors of training performance. *Military Psychology, 6*(1), 31-46.

Earles, J., & Ree, M. (1992). The predictive validity of the ASVAB for training grades. Educational and Psychological Measurement, 52(3), 721-725.

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology, 77*, 337-353.

González, M., Campos, A., & Pérez, M. (1997). Mental imagery and creative thinking. *Journal of Psychology: Interdisciplinary and Applied, 131(4*), 357-364.

Halpern, D.F. (1997). Sex differences in intelligence. *American Psychologist, 52*, 1091-1102.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32*(2), 175-191.

Held, J., Alderton, D., Foley, P., & Segall, D. (1993). Arithmetic reasoning gender differences: Explanations found in the Armed Services Vocational Aptitude Battery (ASVAB). *Learning and Individual Differences, 5*(2), 171-186.

Ingerick, M., Diaz, T., & Putka, D. (2009). *Investigations into Army enlisted classification systems: Concurrent validation report* (Technical Report 1244). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Heffner, T. S. (2009). *Validating future force performance measures (Army Class): End of Training Longitudinal Validation* (Technical Report 1257). Arlington, VA: U.S. Army Institute for the Behavioral and Social Sciences.

Kyllonen, P., & Chaiken, S. (2003). Dynamic Spatial Ability and Psychomotor Performance. *International Journal of Testing, 3(3),* 233-249.

Larson, G., & Alderton, D. (1997). Test-retest results for the ECAT battery. *Military Psychology, 9*(1), 39-47.

Laurence, J.H., Naughton, J., & Harris, D.A. (1996). *Attrition revisited: Identifying the problem and its solutions* (ARI Research Note 96-20). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498.

Loehlin J.C., Lindzey G., Spuhler J.N. (1975). Race differences in intelligence. New York: Freeman.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Maier, M. H., & Hiatt, C. M. (1986a). *Evaluating the Appropriateness of the Numerical Operations and Math Knowledge Subtests in the AFQT (Armed Forces Qualification Test).* Accession Number: ADA180266 (from DTIC).

Maier, M. H., & Hiatt, C. M. (1986b). *Scaling the Proposed New Armed Forces Qualification Test Composed of Verbal, Arithmetic Reasoning, and Math Knowledge.* Accession Number: ADB118541 (from DTIC).

Maitland, S. B., Intrieri, R. C., Schaie, K. W., & Willis, S. L. (2000). Gender differences and changes in cognitive abilities across the adult life span. *Aging, Neuropsychology, and Cognition, 7* (1), 32-53.

Malinowski, J. (2001). Mental rotation and real-world wayfinding. *Perceptual and Motor Skills, 92*(1), 19-30.

Manrique, F. (1998). *The Effects of Display Type and Spatial Ability on Performance during a Virtual Reality Simulation (98-008D).* WPAFB, OH: Department of the Air Force.

Mayberry, P. W., & Hiatt, C. M. (1990). *Incremental Validity of New Tests in Prediction of Infantry Performance* (CRM 90-110). Alexandria, VA: Center for Naval Analyses.

Monzon, R. I., & P. P. & Foley, P. P. (1988). *Assessment of Alternative Armed Forces Qualification Test (AFQT) Composite Definitions.* Accession Number: ADA199691 (from DTIC)

Moriarty, K.O., Campbell, R.C., Heffner, T.S., & Knapp, D.J. (2009). *Investigations into Army enlisted classification systems (Army Class): Reclassification test and criterion development report* (FR 08-54). Alexandria, VA: Human Resources Research Organization.

Office of the Secretary of Defense (2010, September 21). *Official Site of the ASVAB.* Retrieved from: http://officialasvab.com/.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Cognitive ability in personnel selection decisions. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection.* Malden, MA: Blackwell Publishing (pp. 143-173).

Oppler, S. H., Peterson, N. G., Whetzel, D. L., Steele, D., Childs, R. A., Park, R. K., Rosse, R. L., Rehling, J. F., Brantner, T. M., & Kieckhaefer, W. F. (1992). *Selection and Classification Tests for Critical Military Occupational Specialties* (Research Note 92-74). Alexandria, VA: U.S. Army Institute for the Behavioral and Social Sciences.

Putka, D. J., & Bradley, K. M. (2008). *Relations between Select21 Predictor Measures and First-Term Attrition* (Research Note 2008-02). Arlington, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Ree, M., & Carretta, T. (1995). Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement*, *55*(2), 268-277.

Roberts, R., Goff, G., Anjoul, F., Kyllonen, P., Pallier, G., & Stankov, L. (2000). The Armed Services Vocational Aptitude Battery (ASVAB): Little more than acculturated learning (Gc)!?. *Learning and Individual Differences, 12*(1), 81-103.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. *Psychological Bulletin, 85,* 1348-1351.

Russell, T. L., & Peterson, N. G. (2001). The Experimental Battery: Basic attribute scores for predicting performance in a population of jobs. In J. P. Campbell, & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification.* Mahwah, NJ: Lawrence Erlbaum Associates.

Russell, T. L., Reynolds, D. H., & Campbell, J. P. (Eds.) (1994). *Building a joint-service classification research roadmap: Individual differences measurement* (AL/HR-TP-1994-0009). Brooks AFB, TX: Armstrong Laboratory.

Sager, C., Peterson, N., Oppler, S., & Rosse, R. (1997). An examination of five indexes of test battery performance: Analysis of the ECAT battery. *Military Psychology, 9*(1), 97-120.

Salgado, J., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International Validity Generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology, 56(3),* 573-605.

Stermer, N. S. (1988). *Meta-Analysis of Armed Service Vocational Aptitude Battery Composite Validity Data*. Wright-Patterson AFB, OH: Air Force Institute of Technology.

Suzuki, L.A. & Valencia, R.R. (1997). Race-ethnicity and measured intelligence. *American Psychologist, 52,* 1103-1114.

Thomas, J. (2010, March 16). *Cost of a new recruit* (Information paper). Fort Knox, KY: U.S. Army Training and Doctrine Command.

U.S. Army Recruiting Command (2010). *U.S. Army Recruiting Command goals.* Retrieved from http://www.usarec.army.mil/hq/apa/goals.htm

Uriell, Z. A. (2008). *Comparing Paper-and-Pencil and Web-based Test Results: An Equating Study for AFCT Form 19G* (NPRST-TR-08-4). Millington, TN: Navy Personnel Research, Studies, and Technology.

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

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive function. *Personality and Individual Differences, 35* (4), 863-875.

Welsh, J. R., Kucinkas, S. K., & Curran, L. T. (1990). *Armed Services Vocational Battery (ASVAB): Integrative review of validity studies* (Technical Report No. 90-22). Brooks Air Force Base, TX: Air Force Systems Command.

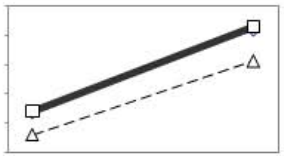
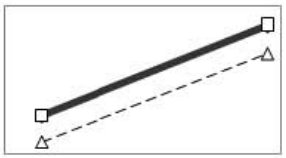
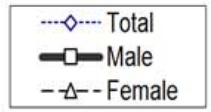
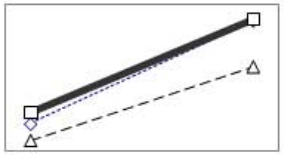
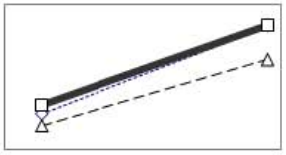
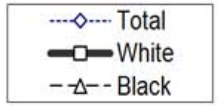
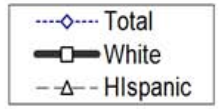
Wise, L., Welsh, J., Grafton, F., Foley, P., Earles, J., Sawin, L., & Divgi, D. R. (1992). *Sensitivity and fairness of the Armed Services Vocational Aptitude Battery (ASVAB) Technical Composites* (DMDC Technical Report 92-002). Washington, DC: Defense Manpower Data Center, Personnel Testing Division.

Wolfe, J. (1997). Stepwise selection of ECATs for maximum mean validity. *Military Psychology, 9*(1), 85-95.

Wolfe, J., Moreno, K., & Segall, D. (1997). Evaluating the predictive validity of CAT-ASVAB. *Computerized adaptive testing: From inquiry to operation (pp. 175-179).* Washington, DC US: American Psychological Association.

Appendix A

Figure A.1: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT JKT : Army Class score standardized within MOS

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,603	-5.03	0.10				-5.41	0.10				
Male	1,468	-5.07	0.10				-5.41	0.10				
Female	135	-4.93	0.09	+	ns		-5.73	0.10	+	ns		
White	1,176	-4.51	0.09				-4.86	0.09				
Black	97	-4.10	0.07	+	ns		-4.49	0.08	+	ns		
Hispanic	244	-5.49	0.10	+	ns		-6.11	0.11	+	ns		

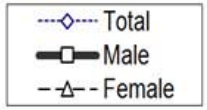
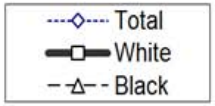
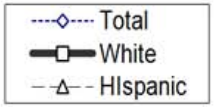
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.79	-1.18	-.12	-.68	-1.06	-.99	-.38	-.30	-.84	-1.28	-.13	-.72	-1.07	-1.10	-.35	-.36	-.02	.03	-.06
-1 SD	-.43	-.87	-.13	-.36	-.80	-.61	-.45	-.24	-.46	-.91	-.14	-.38	-.79	-.68	-.41	-.29	-.01	.04	-.05
Mean	-.08	-.56	-.14	-.04	-.55	-.24	-.52	-.19	-.08	-.54	-.14	-.03	-.50	-.27	-.47	-.22	.00	.04	-.03
+1 SD	.28	-.24	-.16	.28	-.29	.14	-.58	-.14	.30	-.17	-.14	.31	-.22	.15	-.53	-.15	.01	.05	-.01
+2 SD	.64	.07	-.17	.61	-.04	.52	-.65	-.08	.68	.20	-.14	.65	.07	.57	-.59	-.08	.03	.06	.00
n	1468	135		1176	97	244			1468	135		1176	97	244					
M	-.08	-.53		.00	-.68	-.33			-.08	-.53		.00	-.68	-.33					
SD	.98	.84		.97	.94	.91			.98	.84		.97	.94	.91					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.2: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT RITMS: course average, standardized within MOS

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,187	-3.52	0.07				-3.56	0.07				
Male	855	-3.82	0.07				-3.94	0.08				
Female	332	-3.15	0.06	-	ns		-3.04	0.06	-	ns		
White	620	-3.16	0.06				-3.32	0.06				
Black	308	-3.85	0.08	ns	ns		-3.96	0.08	ns	ns		
Hispanic	190	-3.91	0.08	ns	ns		-4.03	0.08	ns	ns		

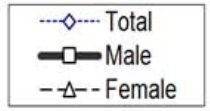
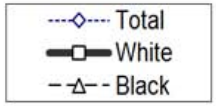
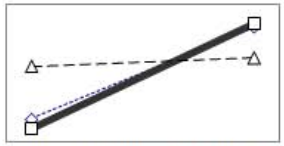
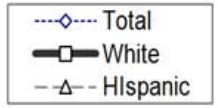
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.69	-.42	.14	-.54	-.64	-.62	-.08	-.07	-.72	-.42	.16	-.59	-.65	-.67	-.05	-.07	.02	.03	.00
-1 SD	-.41	-.18	.12	-.32	-.36	-.34	-.04	-.02	-.43	-.18	.14	-.34	-.35	-.36	.00	-.02	.01	.03	.00
Mean	-.14	.06	.10	-.09	-.08	-.05	.01	.03	-.14	.06	.11	-.10	-.05	-.06	.04	.03	.00	.03	.00
+1 SD	.13	.29	.09	.14	.20	.24	.05	.08	.15	.30	.08	.15	.25	.25	.09	.08	-.01	.04	.00
+2 SD	.41	.53	.07	.37	.48	.52	.10	.13	.44	.53	.05	.40	.55	.55	.14	.13	-.02	.04	.00
n	855	332		620	308	190			855	332		620	308	190					
M	-.12	.01		-.03	-.18	-.13			-.12	.01		-.03	-.18	-.13					
SD	1.02	.89		.89	.96	1.32			1.02	.89		.89	.96	1.32					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.3: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT PRS: Sum of AW Scores

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,758	-2.15	0.04				-2.39	0.05				
Male	1,608	-1.94	0.04				-2.14	0.04				
Female	150	-6.10	0.11	+	+		-6.92	0.13	+	+		
White	1,279	-2.45	0.05				-2.69	0.05				
Black	113	-3.26	0.06	ns	ns		-3.43	0.06	ns	ns		
Hispanic	273	-0.17	0.00	ns	-		-0.28	0.01	ns	-		

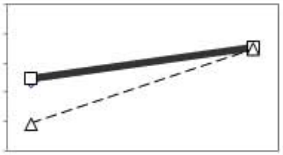
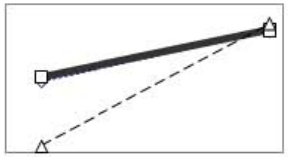
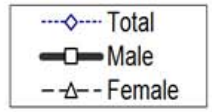
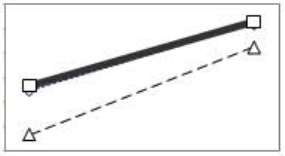
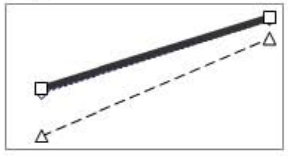
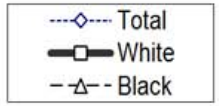
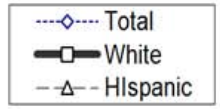
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.30	-1.19	-.26	-.41	-.57	.00	-.15	.37	-.33	-1.28	-.28	-.45	-.57	-.02	-.12	.39	-.02	.03	.02
-1 SD	-.16	-.77	-.18	-.24	-.35	.01	-.10	.23	-.17	-.81	-.18	-.26	-.33	.00	-.07	.24	-.01	.03	.01
Mean	-.02	-.36	-.10	-.07	-.12	.02	-.05	.08	-.02	-.33	-.09	-.07	-.09	.02	-.02	.08	.01	.03	.00
+1 SD	.12	.05	-.02	.10	.10	.04	.00	-.06	.13	.14	.00	.12	.15	.05	.03	-.07	.02	.02	-.01
+2 SD	.26	.47	.06	.27	.33	.05	.05	-.20	.28	.61	.10	.31	.39	.07	.07	-.22	.03	.02	-.02
n	1608	150		1279	113	273			1608	150		1279	113	273					
M	-.02	-.32		-.05	-.23	.02			-.02	-.32		-.05	-.23	.02					
SD	.99	1.13		1.01	1.06	.94			.99	1.13		1.01	1.06	.94					

Note: Effect Size calculated using pooled *SD*s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.4: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT PRS: Sum of standardized MOS scores

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,728	-1.93	0.04				-2.13	0.04				
Male	1,579	-1.79	0.04				-1.96	0.04				
Female	149	-4.81	0.09	+	ns		-5.52	0.10	+	+		
White	1,256	-1.82	0.04				-2.03	0.04				
Black	111	-2.79	0.05	+	ns		-3.09	0.05	+	ns		
Hispanic	272	-0.90	0.02	ns	ns		-0.89	0.02	ns	ns		

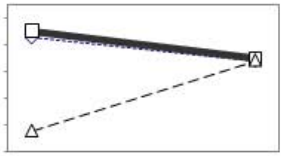
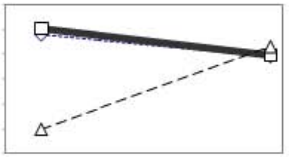
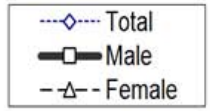
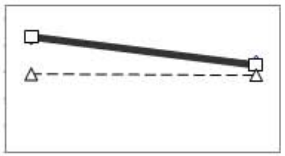
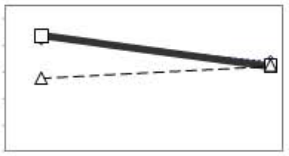
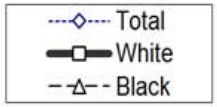
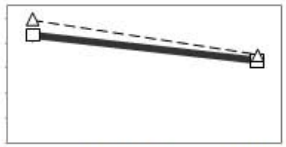
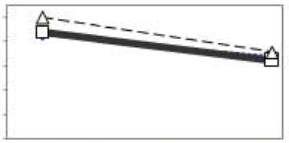
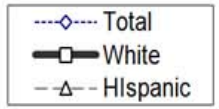
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.26	-1.03	-.23	-.26	-.67	-.17	-.39	.08	-.28	-1.11	-.25	-.29	-.68	-.18	-.38	.11	-.02	.01	.02
-1 SD	-.13	-.71	-.17	-.13	-.49	-.11	-.35	.02	-.14	-.74	-.18	-.15	-.48	-.12	-.32	.03	-.01	.02	.01
Mean	.00	-.39	-.12	.00	-.31	-.05	-.30	-.05	.00	-.37	-.11	.00	-.28	-.06	-.27	-.05	.01	.03	-.01
+1 SD	.13	-.07	-.06	.13	-.13	.01	-.25	-.11	.14	.00	-.04	.14	-.08	.00	-.21	-.13	.02	.04	-.02
+2 SD	.26	.25	-.01	.26	.05	.07	-.21	-.17	.29	.37	.02	.29	.13	.06	-.16	-.21	.03	.05	-.04
n	1579	149		1256	111	272			1579	149		1256	111	272					
M	.00	-.36		.01	-.40	-.07			.00	-.36		.01	-.40	-.07					
SD	.97	1.14		1.00	.97	.96			.97	1.14		1.00	.97	.96					

Note: Effect Size calculated using pooled *SD*s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.5: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT ALQ: Affective Commitment Scale

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,688	4.20	-0.01				4.19	-0.01				
Male	1,545	4.26	-0.01				4.28	-0.01				
Female	143	2.81	0.02	-	ns		2.55	0.02	-	ns		
White	1,239	4.25	-0.01				4.27	-0.01				
Black	102	3.81	0.00	ns	ns		3.63	0.00	ns	ns		
Hispanic	258	4.42	-0.01	ns	ns		4.42	-0.01	ns	ns		

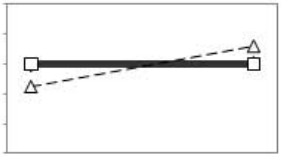
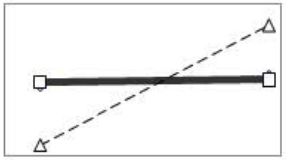
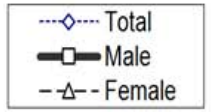
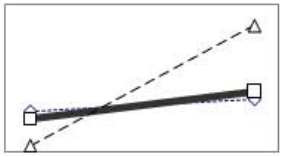
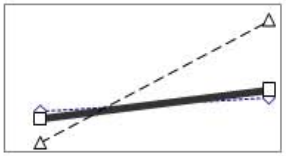
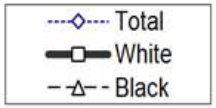
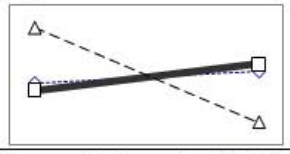
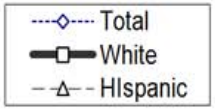
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	3.95	3.58	-.16	3.93	3.80	4.00	-.18	.08	3.95	3.55	-.17	3.94	3.78	4.00	-.21	.08	-.02	-.03	.00
-1 SD	3.93	3.65	-.12	3.91	3.79	3.96	-.15	.07	3.93	3.63	-.13	3.91	3.79	3.96	-.16	.07	-.01	-.01	.00
Mean	3.90	3.71	-.08	3.88	3.79	3.92	-.12	.06	3.90	3.71	-.08	3.88	3.80	3.93	-.11	.06	.00	.01	.00
+1 SD	3.87	3.78	-.04	3.86	3.79	3.89	-.08	.04	3.87	3.80	-.03	3.85	3.81	3.89	-.05	.05	.01	.03	.01
+2 SD	3.85	3.84	.00	3.83	3.79	3.85	-.05	.03	3.85	3.88	.02	3.82	3.82	3.86	.00	.04	.02	.05	.01
n	1545	143		1239	102	258			1545	143		1239	102	258					
M	3.90	3.72		3.88	3.79	3.93			3.90	3.72		3.88	3.79	3.93					
SD	.67	.77		.68	.76	.66			.67	.77		.68	.76	.66					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.6: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT ALQ: Sum of two scales: the Needs-Supplies Army Fit Scale, and the General MOS Fit Scale

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,705	-0.18	0.00				-0.20	0.00				
Male	1,561	-0.04	0.00				-0.05	0.00				
Female	144	-2.44	0.05	ns	ns		-2.79	0.05	ns	ns		
White	1,251	-0.39	0.01				-0.42	0.01				
Black	104	-1.61	0.03	ns	ns		-1.65	0.03	ns	ns		
Hispanic	261	1.41	-0.03	ns	ns		1.46	-0.03	ns	ns		

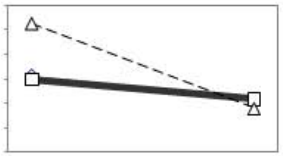
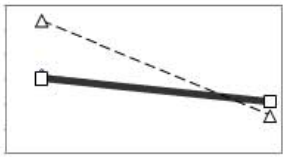
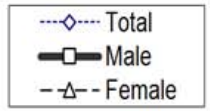
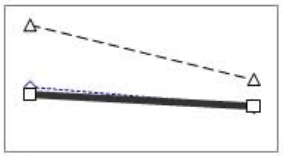
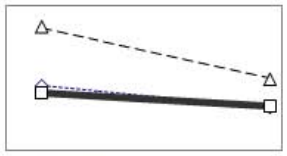
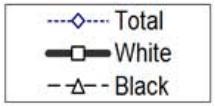
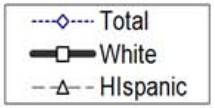
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns' indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total Group	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.01	-.39	-.11	-.06	-.17	.20	-.10	.24	-.01	-.43	-.12	-.07	-.17	.22	-.09	.25	-.01	.01	.01
-1 SD	-.01	-.22	-.06	-.04	-.05	.10	-.01	.12	-.01	-.23	-.06	-.04	-.04	.11	.00	.13	.00	.01	.01
Mean	-.01	-.04	-.01	-.01	.07	.00	.07	.01	-.01	-.03	-.01	-.01	.08	.01	.08	.01	.00	.01	.01
+1 SD	.00	.13	.04	.02	.19	-.10	.16	-.11	.00	.17	.05	.02	.21	-.10	.17	-.11	.01	.01	.00
+2 SD	.00	.30	.09	.05	.31	-.20	.25	-.22	.00	.36	.10	.05	.33	-.20	.26	-.22	.02	.02	.00
n	1561	144		1251	104	261			1561	144		1251	104	261					
M	-.01	-.03		-.01	.01	.03			-.01	-.03		-.01	.01	.03					
SD	1.02	1.04		1.03	1.04	.97			1.02	1.04		1.03	1.04	.97					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.7: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT ALQ: Number of disciplinary incidents during Army training

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,705	1.18	-0.01				1.27	-0.01				
Male	1,561	1.08	-0.01				1.16	-0.01				
Female	144	3.10	-0.05	+	ns		3.36	-0.05	+	ns		
White	1,251	0.88	-0.01				0.95	-0.01				
Black	104	2.79	-0.04	+	ns		2.67	-0.03	+	ns		
Hispanic	261	0.74	-0.01	ns	ns		0.83	-0.01	ns	ns		

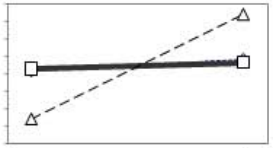
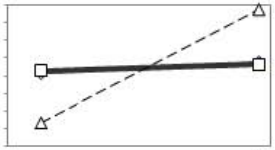
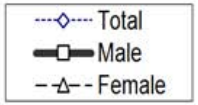
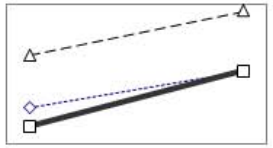
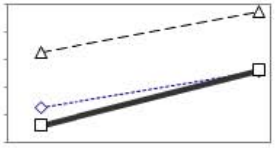
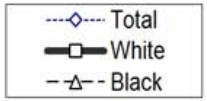
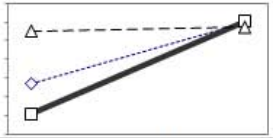
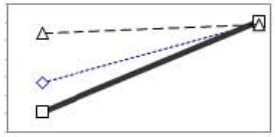
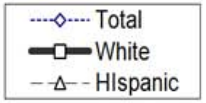
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	.60	1.05	.14	.56	1.22	.52	.56	-.04	.61	1.08	.14	.57	1.20	.53	.54	-.04	.00	-.03	.00
-1 SD	.56	.88	.10	.53	1.09	.50	.47	-.03	.56	.89	.10	.53	1.07	.50	.46	-.03	.00	-.02	.00
Mean	.51	.70	.06	.50	.96	.48	.39	-.02	.51	.69	.06	.50	.95	.48	.38	-.02	.00	-.01	.00
+1 SD	.47	.53	.02	.47	.82	.46	.30	-.01	.47	.50	.01	.47	.82	.45	.30	-.02	-.01	.00	.00
+2 SD	.43	.36	-.02	.45	.69	.44	.21	-.01	.42	.31	-.03	.44	.70	.43	.22	-.01	-.01	.02	.00
n	1561	144		1251	104	261			1561	144		1251	104	261					
M	.52	.69		.50	1.02	.48			.52	.69		.50	1.02	.48					
SD	.93	1.03		.92	1.35	.81			.93	1.03		.92	1.35	.81					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.8: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT ALQ: What was your last APFT score?

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,696	222.13	0.44				221.72	0.44				
Male	1,558	232.65	0.23				232.46	0.23				
Female	138	38.06	4.04	ns	+		23.08	4.29	ns	+		
White	1,244	208.66	0.68				208.58	0.67				
Black	103	227.41	0.54	+	ns		229.05	0.50	+	ns		
Hispanic	260	245.87	0.03	+	ns		243.96	0.06	+	ns		

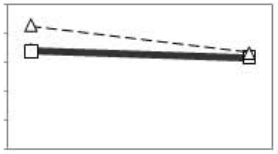
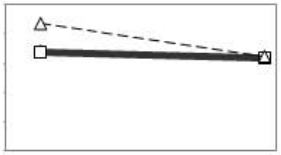
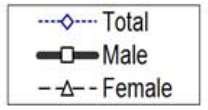
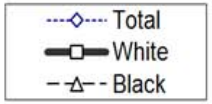
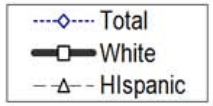
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total Group	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	242.85	214.36	-.29	238.14	250.93	247.04	.40	.27	242.82	213.05	-.30	238.16	251.27	246.78	.41	.26	-.01	.01	-.01
-1 SD	243.71	229.21	-.15	240.62	252.91	247.14	.38	.19	243.70	229.03	-.15	240.65	253.13	247.02	.39	.19	.00	.01	.00
Mean	244.57	244.05	-.01	243.11	254.89	247.24	.37	.12	244.57	245.01	.00	243.14	255.00	247.25	.37	.12	.01	.00	.00
+1 SD	245.43	258.90	.14	245.59	256.87	247.34	.35	.05	245.44	260.98	.16	245.62	256.87	247.49	.35	.06	.02	.00	.00
+2 SD	246.29	273.75	.28	248.07	258.85	247.43	.34	-.02	246.31	276.96	.31	248.11	258.74	247.73	.33	-.01	.03	.00	.01
n	1558	138		1244	103	260			1558	138		1244	103	260					
M	244.56	245.62		243.35	253.96	247.21			244.56	245.62		243.35	253.96	247.21					
SD	30.46	37.55		30.83	31.36	31.16			30.46	37.55		30.83	31.36	31.16					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.9: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT ALQ: Attrition cognitions scale

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,686	2.05	-0.01				2.07	-0.01				
Male	1,546	1.99	-0.01				1.99	-0.01				
Female	140	3.54	-0.03	+	ns		3.87	-0.04	+	+		
White	1,238	2.11	-0.01				2.14	-0.01				
Black	101	3.50	-0.03	ns	ns		3.54	-0.04	ns	ns		
Hispanic	261	1.69	0.00	ns	ns		1.64	0.00	ns	ns		

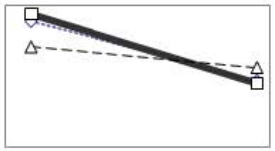
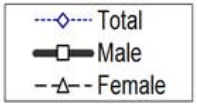
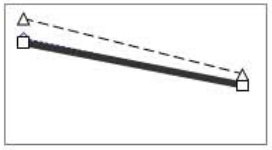
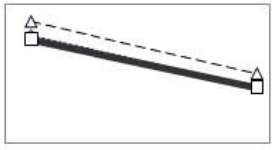
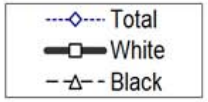
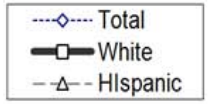
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total Group	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	1.69	2.14	.18	1.73	1.98	1.60	.34	-.17	1.69	2.18	.20	1.74	1.97	1.59	.32	-.19	.02	-.01	-.01
-1 SD	1.66	2.02	.15	1.70	1.85	1.59	.20	-.14	1.66	2.04	.15	1.70	1.84	1.59	.19	-.15	.01	-.02	-.01
Mean	1.64	1.90	.11	1.67	1.72	1.58	.07	-.11	1.64	1.89	.10	1.67	1.71	1.58	.05	-.11	.00	-.02	.00
+1 SD	1.61	1.78	.07	1.64	1.59	1.58	-.06	-.08	1.61	1.75	.06	1.63	1.57	1.58	-.08	-.07	-.01	-.02	.01
+2 SD	1.59	1.66	.03	1.60	1.46	1.57	-.20	-.05	1.59	1.61	.01	1.60	1.44	1.58	-.22	-.03	-.02	-.02	.02
n	1546	140		1238	101	261			1546	140		1238	101	261					
M	1.64	1.89		1.67	1.78	1.59			1.64	1.89		1.67	1.78	1.59					
SD	.68	.91		.71	.68	.65			.68	.91		.71	.68	.65					

Note: Effect Size calculated using pooled *SD*s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.10: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = EOT ATRRS Training Data: Total number of restarts during IMT

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	7,366	0.28	0.00				0.29	0.00				
Male	5,807	0.31	0.00				0.33	0.00				
Female	1,559	0.14	0.00	ns	ns		0.16	0.00	ns	ns		
White	4,665	0.26	0.00				0.28	0.00				
Black	1,108	0.33	0.00	ns	ns		0.31	0.00	ns	ns		
Hispanic	1,138	0.27	0.00	ns	ns		0.26	0.00	ns	ns		

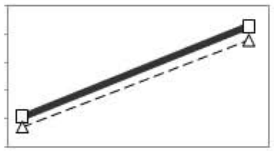
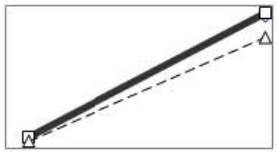
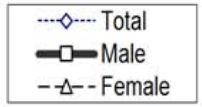
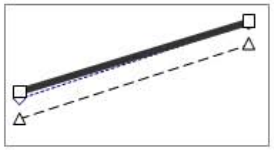
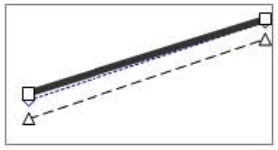
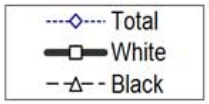
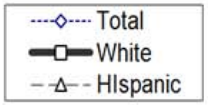
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	.13	.10	-.04	.12	.14	.12	.07	.00	.13	.10	-.04	.12	.14	.12	.05	-.01	.00	-.01	-.01
-1 SD	.11	.09	-.03	.10	.13	.10	.06	-.01	.11	.09	-.03	.11	.13	.10	.05	-.01	.00	-.01	.00
Mean	.10	.09	-.01	.09	.11	.09	.05	-.01	.10	.09	-.01	.09	.11	.09	.04	-.01	.00	.00	.00
+1 SD	.08	.09	.01	.08	.10	.08	.04	-.01	.08	.08	.01	.08	.10	.08	.04	-.01	.00	.00	.01
+2 SD	.07	.08	.02	.07	.08	.06	.03	-.02	.06	.08	.02	.07	.08	.06	.04	.00	.00	.01	.01
n	5807	1559		4665	1108	1138			5807	1559		4665	1108	1138					
M	.10	.09		.09	.12	.09			.10	.09		.09	.12	.09					
SD	.34	.34		.33	.40	.33			.34	.34		.33	.40	.33					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.11: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU JKT: AW Army Class total score

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	992	-5.41	0.10				-5.65	0.11				
Male	825	-5.34	0.10				-5.73	0.11				
Female	167	-5.39	0.10	+	ns		-4.99	0.09	+	ns		
White	616	-5.02	0.10				-5.26	0.10				
Black	125	-5.82	0.10	+	ns		-6.07	0.11	+	ns		
Hispanic	174	-3.67	0.07	ns	ns		-3.97	0.07	ns	ns		

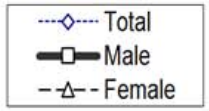
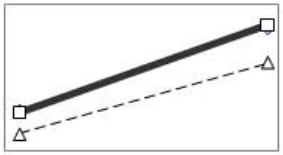
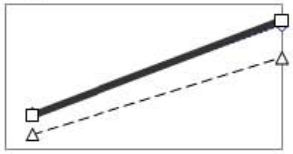
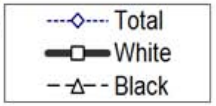
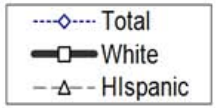
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.96	-1.14	-.08	-.85	-1.42	-.75	-.52	.10	-1.03	-1.08	-.02	-.90	-1.42	-.81	-.47	.09	.05	.04	-.01
-1 SD	-.56	-.76	-.08	-.47	-1.02	-.48	-.50	-.01	-.60	-.72	-.05	-.50	-.99	-.52	-.45	-.02	.03	.05	-.01
Mean	-.16	-.37	-.09	-.09	-.62	-.22	-.48	-.12	-.16	-.36	-.08	-.10	-.57	-.23	-.42	-.13	.00	.05	-.01
+1 SD	.24	.01	-.09	.28	-.22	.05	-.46	-.23	.27	-.01	-.11	.30	-.14	.06	-.40	-.23	-.02	.06	.00
+2 SD	.63	.40	-.10	.66	.17	.31	-.44	-.34	.70	.35	-.14	.70	.29	.35	-.38	-.34	-.05	.06	.00
n	825	167		616	125	174			825	167		616	125	174					
M	-.15	-.42		-.02	-.78	-.30			-.15	-.42		-.02	-.78	-.30					
SD	.98	.97		.93	1.07	.88			.98	.97		.93	1.07	.88					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.12: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU JKT: Overall Army Class total score standardized within MOS

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	474	-4.56	0.09				-4.95	0.09				
Male	421	-4.74	0.09				-5.19	0.10				
Female	53	-3.21	0.06	ns	ns		-3.35	0.06	ns	ns		
White	336	-4.81	0.09				-5.27	0.10				
Black	32	-4.42	0.07	+	ns		-4.79	0.08	+	ns		
Hispanic	75	-3.85	0.07	ns	ns		-4.20	0.08	ns	ns		

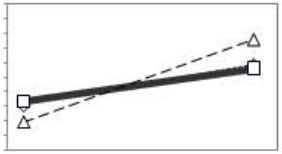
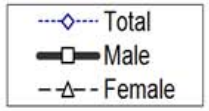
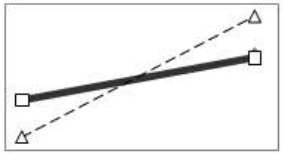
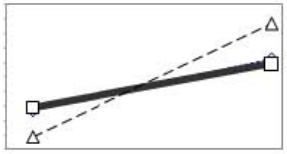
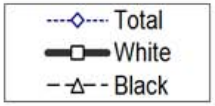
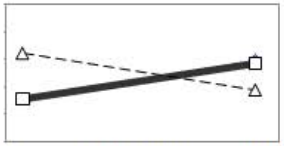
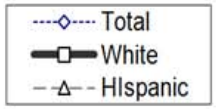
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.84	-.64	.07	-.84	-1.21	-.64	-.33	.18	-.90	-.65	.08	-.90	-1.23	-.70	-.30	.18	.02	.03	.01
-1 SD	-.47	-.39	.03	-.47	-.90	-.34	-.39	.11	-.50	-.40	.03	-.50	-.90	-.37	-.37	.11	.01	.03	.00
Mean	-.10	-.15	-.02	-.09	-.60	-.04	-.46	.05	-.10	-.15	-.01	-.09	-.57	-.05	-.43	.04	.00	.02	-.01
+1 SD	.27	.09	-.06	.29	-.29	.27	-.52	-.02	.30	.11	-.06	.32	-.24	.28	-.50	-.04	-.01	.02	-.02
+2 SD	.64	.34	-.10	.66	.01	.57	-.59	-.08	.69	.36	-.11	.72	.09	.60	-.57	-.11	-.01	.02	-.03
n	421	53		336	32	75			421	53		336	32	75					
M	-.10	-.18		-.03	-.70	-.15			-.10	-.18		-.03	-.70	-.15					
SD	1.03	.90		1.02	1.10	.86			1.03	.90		1.02	1.10	.86					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.13: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU PRS: Sum of AW scores

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	702	-0.97	0.02				-1.23	0.02				
Male	595	-0.80	0.01				-1.12	0.02				
Female	107	-1.89	0.04	ns	ns		-1.75	0.03	ns	ns		
White	426	-0.92	0.02				-1.07	0.02				
Black	99	-2.53	0.05	ns	ns		-2.60	0.05	ns	ns		
Hispanic	119	0.92	-0.02	ns	ns		0.23	0.00	ns	ns		

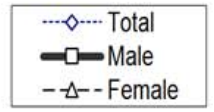
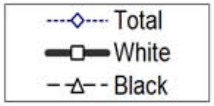
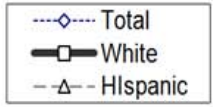
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.17	-.31	-.05	-.19	-.42	.15	-.21	.30	-.22	-.29	-.03	-.21	-.41	.05	-.18	.24	.03	.03	-.06
-1 SD	-.11	-.17	-.02	-.12	-.23	.08	-.10	.18	-.13	-.16	-.01	-.14	-.21	.04	-.07	.16	.01	.03	-.02
Mean	-.05	-.03	.01	-.06	-.04	.01	.02	.06	-.05	-.03	.01	-.06	-.02	.02	.04	.07	.00	.02	.01
+1 SD	.00	.12	.04	.01	.15	-.06	.13	-.06	.03	.10	.03	.02	.18	.01	.15	-.01	-.02	.02	.05
+2 SD	.06	.26	.08	.07	.33	-.13	.24	-.18	.11	.23	.05	.09	.37	-.01	.26	-.09	-.03	.02	.09
n	595	107		426	99	119			595	107		426	99	119					
M	-.05	-.05		-.05	-.10	.03			-.05	-.05		-.05	-.10	.03					
SD	1.00	1.01		.94	.99	1.14			1.00	1.01		.94	.99	1.14					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.14: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU PRS: Sum of Standardized MOS Scores

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	344	-0.74	0.01				-1.00	0.02				
Male	311	-0.54	0.01				-0.85	0.02				
Female	33	-1.91	0.04	ns	ns		-1.68	0.03	ns	ns		
White	247	-1.01	0.02				-1.23	0.02				
Black	22	-2.14	0.04	ns	ns		-1.91	0.04	ns	ns		
Hispanic	53	0.84	-0.02	ns	ns		0.27	0.00	ns	ns		

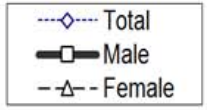
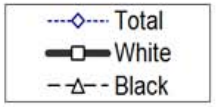
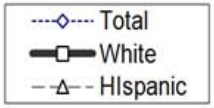
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	-.07	-.36	-.09	-.15	-.40	.13	-.26	.26	-.12	-.32	-.06	-.18	-.34	.06	-.17	.22	.03	.09	-.04
-1 SD	-.03	-.22	-.06	-.07	-.24	.07	-.18	.13	-.05	-.20	-.05	-.09	-.21	.04	-.13	.12	.01	.05	-.02
Mean	.01	-.08	-.03	.00	-.09	.01	-.10	.01	.01	-.09	-.03	.00	-.08	.02	-.08	.02	.00	.02	.01
+1 SD	.05	.05	.00	.08	.06	-.05	-.02	-.12	.07	.03	-.01	.09	.06	.00	-.04	-.08	-.01	-.02	.04
+2 SD	.09	.19	.03	.15	.22	-.12	.06	-.25	.13	.15	.00	.18	.19	-.02	.01	-.18	-.03	-.05	.07
n	311	33		247	22	53			311	33		247	22	53					
M	.01	-.11		.01	-.12	.03			.01	-.11		.01	-.12	.03					
SD	.97	1.12		.99	.85	.87			.97	1.12		.99	.85	.87					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.15: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU ALQ: Affective Commitment Scale

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,014	4.30	-0.01				4.28	-0.01				
Male	826	4.31	-0.01				4.32	-0.01				
Female	188	4.11	-0.01	-	ns		4.00	-0.01	ns	ns		
White	624	4.14	-0.01				4.09	-0.01				
Black	132	3.46	0.00	ns	ns		3.52	0.00	ns	ns		
Hispanic	179	5.94	-0.04	ns	ns		5.99	-0.04	ns	ns		

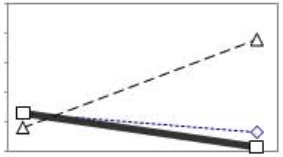
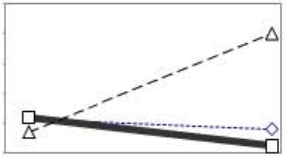
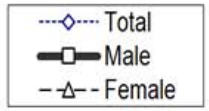
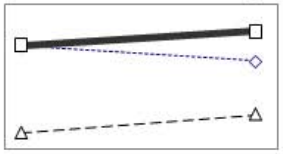
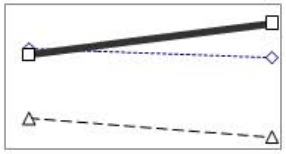
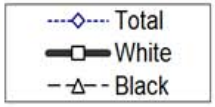
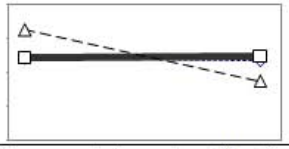
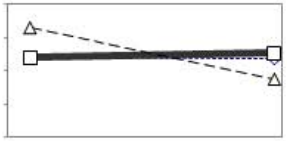
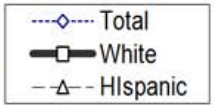
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	3.69	3.71	.01	3.67	3.49	4.02	-.20	.37	3.69	3.69	.00	3.66	3.49	4.04	-.18	.40	-.01	.02	.03
-1 SD	3.63	3.67	.02	3.62	3.49	3.85	-.15	.23	3.63	3.66	.02	3.62	3.49	3.86	-.14	.25	.00	.01	.02
Mean	3.58	3.64	.03	3.58	3.49	3.67	-.10	.10	3.58	3.64	.03	3.58	3.49	3.68	-.10	.11	.00	.00	.01
+1 SD	3.52	3.60	.04	3.54	3.49	3.50	-.05	-.04	3.52	3.61	.05	3.54	3.49	3.50	-.06	-.04	.01	-.01	.00
+2 SD	3.46	3.56	.05	3.49	3.50	3.33	.00	-.17	3.46	3.58	.06	3.50	3.49	3.33	-.02	-.18	.01	-.02	-.01
n	826	188		624	132	179			826	188		624	132	179					
M	3.57	3.64		3.57	3.49	3.73			3.57	3.64		3.57	3.49	3.73					
SD	.86	.75		.86	.82	.81			.86	.75		.86	.82	.81					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.16: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU ALQ: Sum of two scales: the Needs-Supplies Army Fit Scale, and the General MOS Fit Scale

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,014	0.11	0.00				0.06	0.00				
Male	826	0.18	0.00				0.14	0.00				
Female	188	-0.43	0.01	ns	ns		-0.47	0.01	ns	ns		
White	624	-0.06	0.00				-0.19	0.00				
Black	132	-0.27	0.00	ns	ns		-0.03	0.00	ns	ns		
Hispanic	179	2.52	-0.05	ns	-		2.59	-0.05	ns	-		

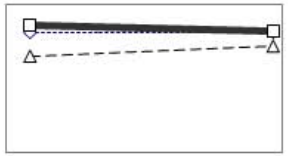
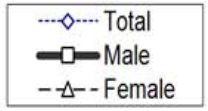
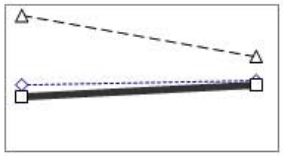
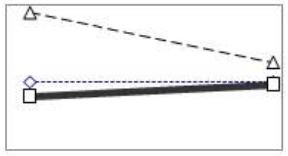
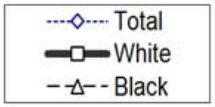
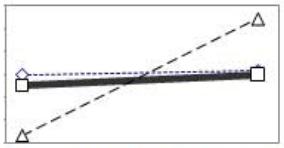
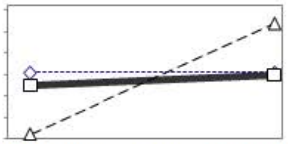
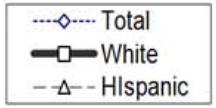
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	.02	-.01	-.01	.02	-.17	.42	-.17	.36	.01	-.02	-.01	-.01	-.14	.45	-.12	.40	.00	.05	.04
-1 SD	.00	.03	.01	.02	-.16	.24	-.17	.19	.00	.03	.01	.01	-.15	.25	-.14	.21	.00	.02	.03
Mean	-.01	.07	.03	.03	-.15	.05	-.16	.02	-.01	.07	.03	.03	-.16	.06	-.17	.03	.00	-.01	.01
+1 SD	-.03	.10	.06	.04	-.14	-.14	-.16	-.15	-.03	.11	.06	.04	-.16	-.14	-.19	-.16	.00	-.03	.00
+2 SD	-.04	.14	.08	.04	-.13	-.33	-.16	-.33	-.04	.15	.08	.06	-.17	-.33	-.22	-.35	.00	-.06	-.02
n	826	188		624	132	179			826	188		624	132	179					
M	-.02	.06		.03	-.15	.11			-.02	.06		.03	-.15	.11					
SD	1.02	.96		1.02	.96	.96			1.02	.96		1.02	.96	.96					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.17: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU ALQ: Disciplinary action total score

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,014	0.46	0.00				0.56	0.00				
Male	826	0.53	0.00				0.69	0.00				
Female	188	0.38	0.00	ns	ns		0.33	0.00	ns	ns		
White	624	0.17	0.01				0.18	0.01				
Black	132	2.05	-0.02	+	ns		2.27	-0.03	+	ns		
Hispanic	179	-2.82	0.07	ns	+		-2.83	0.06	ns	+		

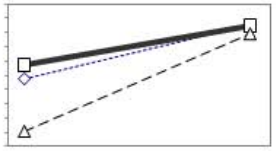
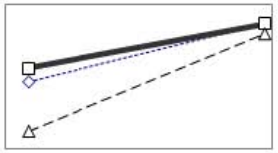
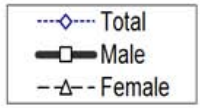
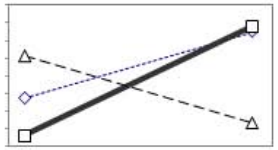
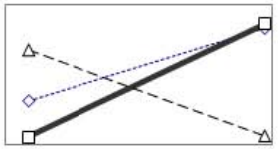
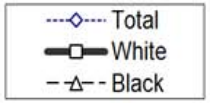
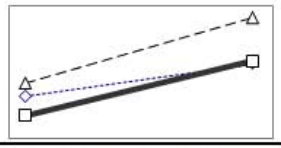
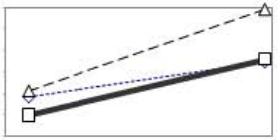
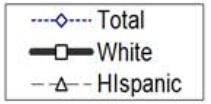
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	.58	.46	-.04	.45	1.12	.02	.50	-.38	.60	.46	-.06	.45	1.14	.00	.52	-.40	-.01	.01	-.02
-1 SD	.58	.47	-.04	.47	1.04	.27	.42	-.17	.60	.47	-.05	.48	1.04	.26	.42	-.19	-.01	.00	-.02
Mean	.59	.48	-.04	.50	.95	.53	.34	.03	.59	.48	-.04	.50	.93	.51	.32	.01	.00	-.01	-.01
+1 SD	.59	.49	-.04	.52	.87	.78	.26	.23	.58	.49	-.03	.52	.83	.77	.23	.22	.01	-.03	-.01
+2 SD	.60	.50	-.04	.55	.78	1.04	.18	.43	.57	.51	-.03	.55	.73	1.03	.13	.42	.01	-.04	-.01
n	826	188		624	132	179			826	188		624	132	179					
M	.59	.48		.50	.98	.45			.59	.48		.50	.98	.45					
SD	1.11	1.01		1.02	1.36	.98			1.11	1.01		1.02	1.36	.98					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.18: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU ALQ: What was your last APFT score?

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	941	218.81	0.48				216.62	0.52				
Male	781	226.10	0.36				223.94	0.39				
Female	160	194.48	0.87	ns	ns		195.26	0.85	ns	ns		
White	588	200.83	0.80				199.58	0.81				
Black	115	265.54	-0.49	ns	ns		271.84	-0.62	ns	ns		
Hispanic	166	200.66	0.97	-	ns		188.48	1.19	-	ns		

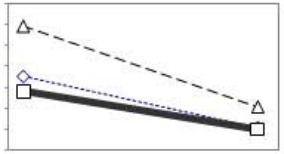
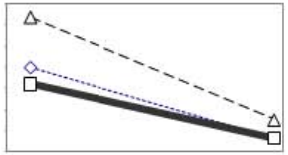
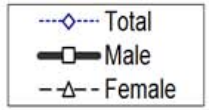
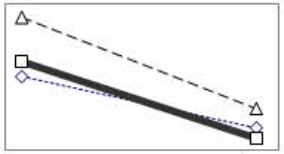
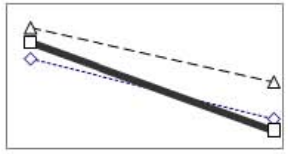
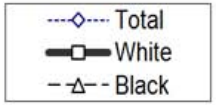
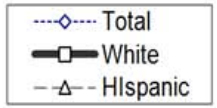
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	241.44	232.19	-.11	235.23	244.31	242.45	.22	.18	241.06	232.30	-.11	234.95	244.90	240.54	.24	.14	.01	.02	-.04
-1 SD	242.82	235.58	-.09	238.32	242.40	246.21	.10	.20	242.62	235.68	-.09	238.18	242.44	245.30	.10	.18	.00	.00	-.02
Mean	244.20	238.98	-.06	241.42	240.49	249.97	-.02	.21	244.19	239.07	-.06	241.41	239.98	250.06	-.03	.22	.00	-.01	.00
+1 SD	245.58	242.37	-.04	244.52	238.58	253.73	-.14	.23	245.75	242.45	-.04	244.64	237.52	254.81	-.17	.25	.00	-.03	.02
+2 SD	246.96	245.77	-.01	247.62	236.66	257.50	-.26	.25	247.31	245.84	-.02	247.87	235.05	259.57	-.31	.29	.00	-.04	.05
n	781	160		588	115	166			781	160		588	115	166					
M	244.25	238.39		242.05	241.24	248.66			244.25	238.39		242.05	241.24	248.66					
SD	36.22	40.54		36.42	41.55	36.51			36.22	40.54		36.42	41.55	36.51					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.19: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = IU ALQ: Attrition cognition mean score

	Baseline						Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	1,014	2.12	-0.01				2.25	-0.01				
Male	826	1.99	-0.01				2.11	-0.01				
Female	188	2.43	-0.01	ns	ns		2.57	-0.02	ns	ns		
White	624	2.33	-0.01				2.51	-0.02				
Black	132	2.53	-0.01	ns	ns		2.30	-0.01	ns	ns		
Hispanic	179	-0.02	0.03	ns	+		0.12	0.03	ns	+		

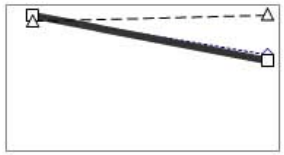
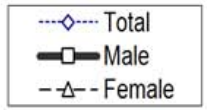
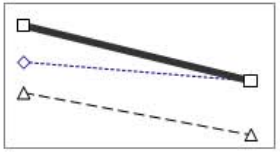
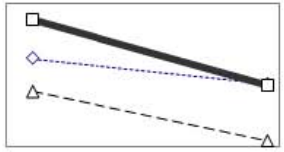
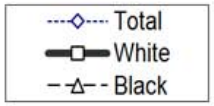
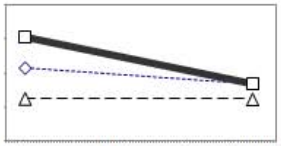
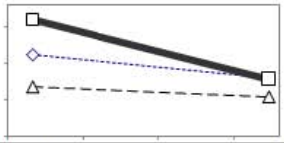
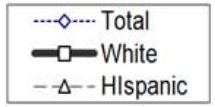
Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)															Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental									
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	1.74	1.90	.08	1.81	1.92	1.40	.12	-.45	1.76	1.92	.08	1.85	1.89	1.41	.05	-.48	.00	-.07	-.02
-1 SD	1.72	1.85	.07	1.77	1.86	1.52	.11	-.26	1.73	1.86	.07	1.78	1.85	1.52	.07	-.28	.00	-.04	-.02
Mean	1.70	1.80	.06	1.72	1.81	1.65	.10	-.07	1.70	1.80	.05	1.72	1.81	1.64	.10	-.09	.00	.00	-.01
+1 SD	1.67	1.75	.04	1.67	1.75	1.78	.09	.11	1.66	1.74	.04	1.66	1.77	1.76	.13	.10	.00	.04	-.01
+2 SD	1.65	1.70	.03	1.63	1.70	1.91	.08	.30	1.63	1.68	.02	1.60	1.74	1.88	.15	.30	-.01	.07	-.01
n	826	188		624	132	179			826	188		624	132	179					
M	1.69	1.81		1.71	1.83	1.61			1.69	1.81		1.71	1.83	1.61					
SD	.79	.86		.84	.77	.68			.79	.86		.84	.77	.68					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.

Figure A.20: Cleary Bias Analyses, Slope and Intercept Differences, Criterion = TTAS: Attrition six months

		Baseline					Supplemental					
Group	*n*	*a*	*b*	*ME*	*IE*	Graph	*a*	*b*	*ME*	*IE*	Graph	Legend
Total	4,529	0.17	0.00				0.20	0.00				
Male	3,667	0.18	0.00				0.21	0.00				
Female	862	0.06	0.00	ns	ns		0.09	0.00	ns	ns		
White	2,822	0.35	0.00				0.39	-0.01				
Black	674	0.22	0.00	+	ns		0.25	0.00	+	ns		
Hispanic	746	0.06	0.00	+	ns		0.11	0.00	+	ns		

Note: ME indicates the direction of any significant differences in the main effect for subgroup membership; IE indicates the direction of any significant differences in the interaction effect for subgroup membership; the values "+", "-", and 'ns" indicate positive, negative, and non significant differences respectively. Graphs show point values for the minimum and max values for the predictor composite and the criterion. The Total regression line is often obscured given that it coincides with that of the majority group.

Differences In Predicted Scores at Different Points Along the Regression Line

Mean Predictor Score for Total	Predicted Scores and Effect Sizes Showing Differences Between Subgroups (Minority - Majority)																Effect Size Diffs (Supp-Baseline)		
	Baseline								Supplemental										
Group	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	Male	Female	d_{MF}	White	Black	Hispanic	d_{WB}	d_{WH}	*MF*	*WB*	*WH*
-2 SD	.11	.10	-.01	.15	.07	.06	-.30	-.27	.11	.11	-.01	.16	.07	.07	-.32	-.27	.00	-.02	.00
-1 SD	.10	.11	.01	.14	.06	.06	-.29	-.22	.10	.11	.01	.14	.06	.06	-.30	-.22	.00	-.01	.00
Mean	.09	.11	.02	.12	.04	.06	-.27	-.17	.09	.11	.02	.12	.04	.06	-.28	-.17	.00	-.01	.00
+1 SD	.09	.11	.04	.10	.03	.06	-.26	-.12	.08	.11	.04	.10	.02	.06	-.26	-.12	.00	-.01	.00
+2 SD	.08	.12	.06	.08	.02	.06	-.24	-.07	.07	.11	.06	.08	.01	.05	-.25	-.07	.00	.00	.00
n	3667	862		2822	674	746			3667	862		2822	674	746					
M	.09	.11		.12	.05	.06			.09	.11		.12	.05	.06					
SD	.29	.31		.32	.21	.24			.29	.31		.32	.21	.24					

Note: Effect Size calculated using pooled *SD* s based on the predicted criterion score. n, *M*, and *SD* statistics reported for the predicted criterion score.