AFHRL-TR-90-22

AIR FORCE

H U M A N

R E S O U R C E S

ARMED SERVICES VOCATIONAL BATTERY (ASVAB): INTEGRATIVE REVIEW OF VALIDITY STUDIES

John R. Welsh, Jr.
Susan K. Kucinkas

Operational Technologies Corporation
5825 Callaghan Road, Suite 225
San Antonio, Texas 78228-1110

Linda T. Curran

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

July 1990

Final Report for Period September 1987 - February 1990

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

This report has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

HAROLD G. JENSEN, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1990	Final - September 1987 - February 1990

4. TITLE AND SUBTITLE

Armed Services Vocational Aptitude Battery (ASVAB): Integrative Review of Validity Studies

5. FUNDING NUMBERS

C - F41689-87-D-0012
PE - 63227F
PR - 2922
TA - 02
WU - 02

6. AUTHOR(S)

John R. Welsh, Jr.
Susan K. Kucinkas
Linda T. Curran

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Operational Technologies Corporation
5825 Callaghan Road, Suite 225
San Antonio, Texas 78228-1110

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFHRL-TR-90-22

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The purpose of this review was to integrate validity evidence relevant to the primary use of the Armed Services Vocational Aptitude Battery (ASVAB) as a selection and classification tool for military manpower, personnel, and training systems. The review covers the period from the first use of ASVAB Form 1 in 1966 in the DOD Student Testing Program to the latest reports of the validity for ASVAB Forms 11, 12, 13, and 14. The review presents the evidence for the construct, content, and criterion-related validity of the ASVAB. One hundred seventy-two studies from the military and civilian sectors and from the professional literature were reviewed and summarized to show averaged validity for military occupations. Reviewed studies established the validity of the ASVAB as a predictor of success in military technical training schools, and its validity for other criteria such as first-term attrition and job performance. Implications of the review for the military selection and classification systems are discussed.

14. SUBJECT TERMS

aggregated validity
construct validity
content validity
criterion-related validity
enlistment testing
equity

15. NUMBER OF PAGES

176

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Item 14 (Concluded):

first-term attrition
selection and classification
sole performance
time-to-completion
training success

Summary

The purpose of the present review is to integrate the validity studies of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 1 through 14. The structure and content of the different generations of the ASVAB are examined, and changes to the multiple-aptitude battery are documented. The review covers the use of the ASVAB by the U.S. Armed Forces for determining the enlistment aptitudes of applicants. Validity data and information related to the ASVAB's prediction of success in entry-level technical training courses are aggregated using meta-analytic techniques. Relevant research pertaining to other aspects of the validity of the ASVAB, such as content-related validity studies and construct validity studies, is also discussed.

This review discusses the validity of the ASVAB for a number of different types of criteria. Among them are final technical school training grade, time-to-completion for self–paced technical training courses, attrition from technical training, first-term attrition, and experimental job performance measures.

The primary conclusion from the review of the literature is that the ASVAB aptitude composites and Armed Forces Qualification Test (AFQT) are valid predictors of final school grades, self-paced technical school completion times, first–term attrition, and job performance measures. The consistent finding from empirical, criterion-related studies shows that the five composites examined in this review (Mechanical-M, Administrative-A, General-G, Electronics-E and the AFQT) all predict final technical school grades with an order of magnitude between .55 and .60 (corrected for restriction in range). The validity coefficients of these five ASVAB composites against other criteria are lower, but still appreciable.

The construct and content validity of the ASVAB were established through a number of studies comparing the ASVAB subtests and composites to other well-known multiple-aptitude batteries.

Studies that reported results relevant to the subgroup equity of the ASVAB are also reviewed and discussed. The bulk of the empirical evidence shows the ASVAB to be equitable for racial subgroups. The ASVAB is also equitable for males and females, with certain job-specific exceptions where female performance is over- or underpredicted.

The generalization of ASVAB validities from military occupational training success to civilian occupations is discussed in light of several validity generalization studies of the ASVAB and many studies of the relationship of the ASVAB to commercially available civilian test batteries. The ASVAB is a valid tool for counseling in its use as a predictor of civilian occupational training success. However, future research needs to address the issue of differential validity in the ASVAB and other multiple-aptitude batteries and to better determine the role of specific and general cognitive abilities in prediction of job performance and training success.

Preface

This research and development effort was conducted under Contract No. F41689-87-D-0012/5002, Task 02, Development of ASVAB Validity Compendium and Plan for a Military Testing Archive.

The authors wish to express their appreciation to Mr. Ray Nakasone, whose computer programming support and skill made it possible to bring all the data together. Thanks also go to Ms. Lynn Trent, whose patience and intelligent cataloging of all the validity studies aided immeasurably in keeping track of the project and provided an audit trail for the results.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Concluded)

LIST OF TABLES

LIST OF TABLES (Continued)

Table Page

LIST OF TABLES (Concluded)

ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB): INTEGRATIVE REVIEW OF VALIDITY STUDIES

I. INTRODUCTION

Description of the ASVAB

The Armed Services Vocational Aptitude Battery (ASVAB) is a group-administered, multiple-aptitude battery used since 1976 to determine the enlistment aptitudes of applicants for the U.S. Armed Forces. Since 1966, the ASVAB has also been administered in the nation's high schools for use in career exploration. The content, administrative oversight, administration conditions, normative score-scale, battery development method, calibration method, and type and quality of supporting validity research have all changed to varying extents over the years. The ASVAB's purpose has not changed, however; its fundamental use was and is to select and classify applicants for enlistment into the U.S. Armed Forces.

ASVAB subtest content areas have remained unchanged since the 1980 implementation of ASVAB Forms 8, 9, and 10. However, specific items have changed between forms of the battery. Validities of Forms 8, 9, and 10, as well as Form 14 (a high school or Department of Defense Student Testing version), have been summarized in several published documents (American Association of Counseling and Development [AACD], 1984; Department of Defense [DoD], 1984a, 1984b).

Table 1 describes the subtest content of these three generations of ASVAB. The present report will examine subtest content changes, and discuss the validity of each subtest; however, its primary emphasis is on the validity of the Military Services' classification composites (vice individual subtests). Focus on the composites is appropriate because only composite scores are used to select and classify individuals (Waters, Lawrence, & Camara, 1987).

Subtest Content of the ASVAB

The evolution of the content of the ASVAB from Form 1 through the present forms can be categorized on the basis of subtest content into three basic generations: Forms 1 through 4; Forms 5, 6, and 7; and Forms 8 through 17. Changes in the subtest content of the ASVAB between generations reflect the perceived changes in demands on new recruits.

These changes in aptitude demands or requirements are based, at least in part, on validity results of the recent operational battery and experience with current recruits in technical training.

Table 1. ASVAB Subtest Content by Form

Subtest	First Generation 1968-1975 Forms 1-4	Second Generation 1976-1980 Forms 5-7	Third Generation 1980-present Forms 8-17
Word Knowledge (WK)	25	30	35
Arithmetic Reasoning (AR)	25	20	30
Mechanical Comprehension (MC)	25	20	25
Electronics Information (EI)	25	30	20
Space Perception (SP)	25	20	
Coding Speed (CS)	100		84
Shop Information (SI)	25	20	
Automotive Information (AI)	25	20	
Auto & Shop Information (AS)			25
Tool Knowledge (TK)	25		
Numerical Operations (NO)		50	50
Mathematics Knowledge (MK)		20	25
General Science (GS)		20	25
Classification Inventory		87	
Attention to Detail (AD)		30	
Paragraph Comprehension (PC)			15
General Information (GI)		15	
Total number of items	300	382	334

Note. From Armed Services Vocational Aptitude Battery (ASVAB) Test Manual (p. 94) Department of Defense, 1984a, North Chicago, IL: Military Entrance Processing Command.

Since 1980, the ASVAB has contained 10 subtests; eight of them are power tests and two are speeded. For ASVAB Forms 8 through 17, the content, the number of items, time limits and a brief description for each subtest are provided in Table 2. The content of ASVAB Forms 8 through 17 is different from that of Forms 5, 6, and 7. Four particular subtests have been included in all forms of the ASVAB: Word Knowledge (WK), Arithmetic Reasoning (AR), Mechanical Comprehension (MC), and Electronics Information (EI). Although these four subtests have been in all three ASVAB generations, the lengths of WK and AR have increased, and the length of EI has decreased (Bayroff & Fuchs, 1970).

Table 2. Subtest Content of ASVAB Forms 8 through 17

Subtest (ASVAB Order)	Description	Number of items	Test time (mins)
General Science (GS)	Knowledge of the physical and biological sciences	25	11
Arithmetic Reasoning (AR)	Word problems emphasizing mathematical reasoning rather than mathematical knowledge	30	36
Word Knowledge (WK)	Understanding the meaning of words; i.e. vocabulary	35	11
Paragraph Comprehension (PC)	Presentation of short paragraphs followed by one or more multiple-choice items	15	13
Numerical Operations (NO)[a]	A speeded test of four arithmetic operations; i.e. addition, subtraction, multiplication and division	50	3
Coding Speed (CS)[a]	A speeded test of matching words and four-digit numbers	84	7
Auto and Shop Information (AS)	Knowledge of auto mechanics, shop practices and tool functions in verbal and pictorial items	25	11
Mathematics Knowledge (MK)	Knowledge of algebra, geometry, and fractions	25	24

Table 2. (Concluded)

Subtest (ASVAB Order)	Description	Number of items	Test time (mins)
Mechanical Comprehension (MC)	Understanding mechanical principles such as gears, levers, pulleys and hydraulics in verbal and pictorial items	25	19
Electronics Information (EI)	Knowledge of electronics and radio principles in verbal and pictorial items	20	9
Total		334	144

[a]Speeded subtest.

The reader is referred to the ASVAB Test Manual (DoD, 1984a) and the Counselor's Manual (AACD, 1984) for a more complete description of changes in the battery over time. For examples of the item types in each subtest and a detailed ASVAB item taxonomy for Forms 8 through 17, see Appendix A of the Technical Supplement to the High School Counselor's Manual (DoD, 1984b).

ASVAB Composites

The U. S. Military Services do not use individual ASVAB subtest scores for selection and classification but rather, composites composed of several subtests. Each Service defines its own set of selection and classification composites and periodically revises these composites as needs and validity data dictate. The Service composites used for ASVAB Forms 8 through 14, and their respective subtest composition, are presented in Table 3.

Table 3. Subtest Definitions of ASVAB Forms 8 through 14 Selector Composites Used by Military Services

Service	Composite[a]	Definition[b]
All	New AFQT[c]	2VE + AR + MK
	Old AFQT[c]	WK + PC + AR + NO/2
Army	GT	VE + AR
	GM	MK + EI + AS + GS
	EL	AR + MK + EI + GS
	CL	AR + MK + VE
	MM	NO + AS + MC + EI
	SC	AR + AS + MC + VE
	CO	CS + AR + MC + AS
	FA	AR + CS + MC + MK
	OF	NO + AS + MC + VE
	ST	VE + MK + MC + GS
Navy	EL	AR + MK + EI +GS
	E	AR + GS + 2MK
	CL	NO + CS + VE
	GT	VE + AR
	ME	VE + MC + AS
	EG	MK + AS
	CT	VE + AR + NO + CS
	HM	VE + MK + GS
	ST	VE + AR + MC
	MR	AR + MC + AS
Air Force	M	MC + GS + 2AS
	A	NO + CS + VE
	G	VE + AR
	E	AR + MK + EI + GS
Marine Corps	MM	AR + EI + MC + AS
	CL	VE + MK + CS
	GT	VE + AR + MC
	EL	AR + MK + EI +GS

Note. Navy scores are the sum of subtest standard scores. The other Services use composite standard scores which are linear transformations of the sum of subtest standard scores. The Air Force converts these composite standard scores to percentiles.

[a]Definitions of composite abbreviations are found in Table A-5.

[b]Definitions of subtest abbreviations are found in Table A-2.

[c]Old Armed Forces Qualification Test (AFQT) scores are the sum of raw subtest scores; new AFQT scores are the sum of subtest standard scores. New AFQT in effect after 1 January 1989.

Because the Services train almost all recruits, validity studies are conducted and reported about every 4 years to ensure the ASVAB forms currently in use are predictive of training success. These validations allow adjustments to selection and classification composites based on recruit performance in military training schools.

All Services use the Armed Forces Qualification Test (AFQT), a composite formed from ASVAB subtests, to report the overall aptitude level of enlisted accessions. The composition of the AFQT has changed only twice in the last 10 years: in October 1980, with the implementation of ASVAB Forms 8, 9, and 10; and in January 1989, with the implementation of ASVAB Forms 15, 16, and 17. Changes to the AFQT generally are made sparingly because of statutory provisions governing selection into the Military Services. These provisions are defined in terms of the AFQT percentile score-scale. For purposes of reporting the ranges of abilities of new recruits to the Congress, the AFQT score-scale is divided into the categories described in Table 4.

Table 4. Definition of AFQT Categories by Percentile Range

AFQT Category	Percentile Range
I	93 - 99
II	65 - 92
IIIa	50 - 64
IIIb	31 - 49
IV	10 - 30
V	01 - 09

Categories I and II are the highest ability categories and include scores at or above the 65th percentile. Category III is often subdivided into IIIa and IIIb at the 50th percentile to facilitate decisions about above-average and below-average recruits. The Services usually limit the number of Category IV recruits for a given year, and Federal statutes prohibit the enlistment of Category V applicants, whose scores fall below the 10th percentile.

Dual Role of the ASVAB

In addition to the enlistment testing program, the ASVAB is offered at no charge to the nation's secondary schools. The ASVAB in the DoD Student Testing Program (or high school testing program) is used for career exploration by students and counselors. The benefit to the DoD is that ASVAB (currently Form 14) scores from the high school testing program can be used for up to 2 years for purposes of enlistment. Because of the ASVAB's dual role, studies reporting the ASVAB validity for career exploration and for prediction of success in civilian occupations are also reviewed in this report.

Current ASVAB composites used with Form 14 for career counseling in the high school testing program are shown in Table 5. The three Academic Composites (Academic Ability, Verbal, and Math), which measure a student's potential for further formal education, are based on the results of factor analytic research on the ASVAB. The four Occupational Composites (Mechanical & Crafts; Business & Clerical; Electronics & Electrical; and Health, Social, & Technology) predict performance in four broad career areas (AACD, 1984, pp. 4–5). These composites were constructed from the results of studies into the validity of the ASVAB for predicting success in entry-level military occupations.

Table 5. High School ASVAB (Form 14) Composites

Composites	Subtests	Purpose
	Academic Composites	
Academic Ability (AA)	AR + (WK + PC)	Measures potential for further formal education
Verbal (VBL)	GS + WK + PC	Measures capacity for verbal activities
Math (MTH)	AR + MK	Measures capacity for mathematical activities

Table 5. (Concluded)

Composites	Subtests	Purpose
	Occupational Composites	
Mechanical and Crafts (MC)	AR + AS + MC + EI	
Business and Clerical (BC)	(WK + PC) + CS + MK	
Electronics and Electrical (EE)	GS + AR + MK + MC	
Health, Social, and Technology (HST)	AR + (WK + PC) + MC	

Note. From Counselors Manual for the Armed Services Vocational Aptitude Battery Form 14 (p. 4-5) 1984, North Chicago, IL: United States Military Entrance Processing Command.

ASVAB Administration

Production Testing

The ASVAB Forms used to test applicants for the Military Services are commonly referred to as "operational forms." Their use in the Military's selection and classification system is termed "production testing." Production testing for enlistment takes place at approximately 70 Military Entrance Processing Stations (MEPS) and at their satellite testing stations, Office of Personnel Management (OPM), and Mobile Examining Test Sites (METS). There are over 1,000 OPM and METS sites nationwide and overseas. The production ASVAB forms are administered to approximately 1 million applicants for enlistment each year.

DoD Student Testing Program

Each year the ASVAB is administered in approximately 14,000 of the nation's high schools to over 1.2 million secondary and post-secondary students.

The ASVAB is administered using directions contained in the ASVAB Administration Manual (DoD, 1983c). However, the standard production testing environment differs from that of the high school testing program. Generally there is a more standardized controlled testing situation at the MEPS, OPM, and METS. Also, the student testing environment varies from high school to high school, from crowded cafeterias to individually monitored classrooms with test instructions read over an intercom. The reality is that there is considerably more variability in the testing conditions of the DoD Student Testing Program than at the MEPS, OPM, and the METS. In spite of their environmental differences, the validity of the ASVAB for both the high school testing program and in military selection and classification testing will be discussed.

The present review integrates the body of validity evidence for the ASVAB, beginning with the ASVAB Form 1 used in 1966 in the DoD Student Testing Program. It covers published and unpublished research on all forms of the ASVAB. This review does not address the validity of ASVAB Forms 15, 16, and 17, as no reports yet exist.

Other predictors of criteria of interest to Military manpower planners are discussed in terms of the influence on, or relationship to, the ASVAB. These discussions do not dwell on the validity of the other types of predictors per se, but simply present results of studies that explicate empirical relationships with the ASVAB in the context of construct validity. For example, demographic influences on the validity of the ASVAB are among the construct validity research results discussed.

Validity

Definition

The term "validity," as used in this report, is derived from the Principles for the Validation and Use of Personnel Selection Procedures (Society for Industrial and Organizational Psychology, Inc. [SIOP], 1987) and the Standards for Educational and Psychological Testing (American Psychological Association [APA], 1985, hereafter referred to as the Standards). According to SIOP, there are not different types of validity so much as there are different aspects of validity:

> Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. The inferences regarding specific uses of a test are validated, not the test itself. (SIOP, 1987, p. 4)

The most recent edition of the Standards uses much the same wording, stressing the importance of what is validated, specifically the inferences drawn from test scores for purposes of selection and classification, rather than the tests or the procedures themselves. Thus, the present review considers and focuses on studies and reports that set the context for ASVAB use, and provides a quantitative review of empirical studies of the criterion-related validity of the battery. Various aspects of validity are discussed in the context of test use, and range from construct, content and criterion-related validity to face validity.

The Standards state that categorization of validity is generally a matter of convenience; i.e., strict classification of validity is nearly impossible. The conventional labels reflecting types of validity are used only to facilitate discussion.

Types of Validity

Construct Validity. Construct validity pertains to evidence that a test score is a measure of the psychological characteristic of interest (APA, 1985). The construct--in this case, ability--derives its meaning from a conceptual framework; and the pattern of relationships with variables and other constructs in that framework provides the meaning or validity of the construct. Evidence that supports parts of the framework adds to the validity of the constructs in the framework. The focus of construct validity is the pattern of the relationships in the nomological net that constitutes the framework of relationships. As Wainer and Braun (1988) noted, other categories of validity, such as criterion-related and content validity, are subsumed under construct validity. Construct validity covers all aspects of a test, from item development to the inferences drawn from its scores. Its broad definition accurately reflects the broad nature of construct validity. In this report, factor analytic studies, and studies providing empirical evidence of the relationship of the ASVAB to other multiple-aptitude batteries that purport to measure the same abilities, are taken as direct indicators of the ASVAB's construct validity.

Content Validity. The Standards (APA, 1985) state that content validity evidence "demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content" (p. 10). The usual procedure is to link the domain to the intended use of the instrument. In practice, this is difficult to achieve. Content validity of the ASVAB is established, in part, by the methods and processes used to develop the forms and to define the content of the items and subtests

(Bayroff & Fuchs, 1970; Jensen, Massey, & Valentine, 1976; Prestwood, Vale, Massey, & Welsh, 1985; Vitola & Alley, 1968). The content validity is heavily dependent on Service experience with a given ability measure and depends on the results of validity studies of previous batteries to identify criterion space that may not be measured by existing subtests. The need for new subtests may be signaled in a number of ways, but primarily through recruit performance on primary criteria of interest to manpower planners. Success in technical training is one such criterion. Increased attrition from certain types of training courses may give rise to hypotheses about currently untapped cognitive abilities needed to master certain technical content. Thus, new or experimental subtests are tried out, and if found to be predictive of criteria of interest, are eventually implemented.

Criterion-Related Validity. Criterion-related validity refers to the systematic relationship of test scores to one or more criteria (APA, 1985, p. 11). The bulk of ASVAB validity evidence falls into this category, and the primary criterion used in most ASVAB validity studies is performance in training. However, there is much evidence related to other criteria of interest such as first-term enlisted attrition, second-term attrition, supervisor ratings, and disciplinary actions. In addition, some evidence pertains to the validity of the ASVAB high school composites for prediction of success in civilian occupations.

Organization of Review

General

An historical perspective on the evolution of the ASVAB provides a proper setting for understanding the validity results. Section II provides this perspective and some context for the use of the test battery and for subsequent discussion of construct validity. Most validity evidence reported for the ASVAB is criterion-related and is covered in Section III. The large amount of empirical, criterion-related evidence necessitates use of quantitative methods to integrate the validity evidence. This condition has been precipitated by several unique features of validity research in the Military; in particular, the very large sample sizes and large numbers of occupations. Section IV describes the studies relative to content validity of the ASVAB. Because content validity is not generally demonstrated in a strictly empirical manner, it does not lend itself to the same sort of quantitative summarization as does the criterion-related validity research. Studies included in Section IV cover a wide variety of areas, from test development reports to analytically sophisticated item-level factor analyses. Section V reviews evidence of the construct validity of the ASVAB, including ASVAB subtest-level factor analytic studies and studies which compare the ASVAB to other multiple-aptitude batteries. Also covered in Section V are studies on the validity generalization of the ASVAB. Section VI reviews the validity and the equity of the ASVAB for population subgroups. Section VII contains the Summary and Conclusions, as well as some recommendations for future research that are based on the findings from this review.

Data Analyses and Summaries

Validity research in the Military enjoys some advantages not found in the civilian sector. One such advantage is the availability of large sample sizes (Ns) for analysis of predictor-criterion relationships. This is beneficial because large Ns lead to stable estimates of the relationships and provide high statistical power.

The empirical criterion-related validity studies reviewed in Section III are summarized using meta-analytic techniques where the published data are sufficiently detailed (i.e., include correlations, standard deviations and sample sizes). The meta-analytic summary provides average effect sizes (averaged validities after Fisher's r to Z transformations) across all studies that contain the required information. Hunter, Schmidt, and Jackson (1982) claimed that the r to Z transformation should not be made, because the sampling variance of r would be overestimated. James, Demaree, and Mulaik (1986) argued the opposite position, however. For convenience and because the literature indicates that any overestimation attributable to the Fisher's r to Z transformation is small (Schmidt, Hunter, & Raju, 1988), the decision was made to use the Fisher's r to Z transformation.

In addition, averaged validities or effect sizes, as well as standard deviations of the reported coefficients weighted by their respective sample sizes, were calculated for each of the broad classes of criteria summarized from the literature. Other summary information was also calculated in order to integrate the empirical findings.

The estimation of effect size is important for at least two reasons. First, given the extremely large numbers of people selected and classified on ASVAB composites, very small observed effects are likely to be both statistically significant and practically important. Even small increments in predictive validity yield large rewards in reduced attrition from technical training and in cost-avoidance (Automated Sciences Group & CACI-INC-Federal, 1988; Schmidt, Hunter, & Dunn, 1987; Vitola, Guinn, & Wilbourn, 1977). Second, the Services have typically made decisions on changing the composition of selection and classification composites based on small increments in predictive validity (Maier, 1982). For example, Schmidt, Hunter, and Dunn (1987) indicated that adding a measure of perceptual aptitude to the current battery could add small increments to its validity in prediction of training success (R^2 increase = .02). They estimated that this increment could amount to over $80 million in cost-avoidance from reduced training attrition.

Quantitatively summarizing this body of research has been complicated by the fact that before 1982, some studies did not report validity coefficients corrected for range restriction. After 1982, most studies reported both corrected and uncorrected correlations. To standardize reporting for this review yet include as much data as possible, uncorrected validities were combined across studies where a mixture of corrected and uncorrected correlations was found. This report does not attempt to correct averaged uncorrected validity coefficients for range restriction, sampling error, or unreliability in either the criterion or predictors, in the manner advocated by Schmidt and Hunter (1977). Instead, this study reports simple aggregated validity results of studies classified by type of criterion and predictor composites Several validity generalization studies were accomplished using Services validity data (Foley, 1986; Jones, 1988; Stermer, 1988) for Military testing validity data on ASVAB Forms 8, 9, and 10.

Included in this section on criterion-related validity are results from the DoD Student Testing Program. They, too, are quantitatively summarized where appropriate, and discussed in the context of the use of ASVAB high school composite scores as tools in student career exploration.

Types of Studies Included and Excluded from Review

Unpublished studies included in the present review are those which have been subjected to some type of review and approved by their respective Services. In most but not all cases, this amounts to being cleared for release after passing a review at the Services' personnel laboratories (as is the case for the Navy validity "letters"); in other cases, they may be approved but unpublished master's theses (e.g., Jones, 1988; Stermer, 1988) or studies commissioned by various Service agencies but never published as technical reports or other Service publications (e.g., Friedman, Crosson, Streicher, & Messersmith, 1986; Hunter, Crosson, & Friedman, 1985).

The majority of both published and unpublished validity studies reviewed in this report are criterion-related predictive validity studies that use success in training as the criterion.

Sources of Studies

The majority of the studies reviewed here are reports published by the Services' personnel research laboratories. However, a significant number of ASVAB validity studies were published by the Department of Defense (DoD), generally through the Office of the Assistant Secretary of Defense for Manpower and Personnel. These DoD studies generally address ASVAB validity for the prediction of first-term (and later) attrition, job performance measures, and performance in technical training schools.

Most studies were obtained from the publication source or the Defense Technical Information Center (DTIC). A search of the professional literature identified a small number of additional studies. Personal contact with researchers yielded a number of unpublished studies such as the six contained in official letters written to Navy personnel decision makers concerning ASVAB validity information for specific Navy occupations. Other unpublished reports reviewed were master's theses from accredited universities or from the Services' postgraduate schools. Still other reports covered contractual personnel research efforts provided to Government agencies in the form of contractor reports and were accepted by the Government agency monitoring the study.

II. HISTORICAL PERSPECTIVE

The first ASVAB evolved at a time when all four Services screened applicants for general trainability using the AFQT (as required by statute) and classified recruits using tests of more specific aptitudes to suit each individual Service's needs. Because many of the Service-specific cognitive tests were similar in content, the question was raised in the Department of Defense as to why there might not be a single test for all of the Services, rather than three different batteries which appeared to measure many of the same things (Bayroff & Fuchs, 1970, p. 1). The first ASVAB was developed in response to this concern.

ASVAB Battery, Subtest, and Composite Changes

Bayroff and Fuchs (1970) described the rationale for the development of ASVAB Form 1. The similarity of the different cognitive test batteries used for classification by the various Services provided the impetus for the first ASVAB and defined its content. Those subtests included in Form 1 because they were "interchangeable" among the separate Service classification tests were: Word Knowledge, Arithmetic Reasoning, Mechanical Comprehension, Space Perception, Shop Information, Automotive Information, and Electronics Information.

A number of reports document the development of various forms of the ASVAB over the years (Bayroff & Fuchs, 1970 and Vitola & Alley, 1968, ASVAB Form 1; Frankfelt, 1970, Forms 2 and 5; Weeks, Mullins, & Vitola, 1975, USAF Classification Batteries; Fruchter & Ree, 1977, candidate forms to replace Forms 5, 6, and 7; Andberg, Stillwell, Prestwood, & Welsh, 1988; Prestwood et al., 1985; and Ree, Mullins, Mathews, & Massey, 1982, Forms 11, 12, and 13). These studies taken collectively, document the changes in ASVAB content since Form 1 was first used in the high school testing program in 1966. The battery composition of ASVAB forms has remained unchanged since 1980, when Forms 8, 9, and 10 were implemented. The subtests that have appeared in all ASVAB forms are the same ones that formed the nucleus of ASVAB Form 1 (Bayroff & Fuchs, 1970; Jensen, Massey, & Valentine, 1976; Vitola & Alley, 1968; Vitola, Mullins, & Croll, 1973).

Both the subtests and the content of the battery have undergone evolutionary change. Many of the individual subtests increased in length and consequently, reliability. Some early ASVAB batteries used in the late 1960's prior to the Joint-Services ASVAB took as long as 5 hours to administer; today the ASVAB administration time is less than 3 hours (Weeks, Mullins, & Vitola, 1975). As the ASVAB evolved, some subtests were combined; e.g., Auto Information and Shop Information (combined to form the Auto/Shop Information subtest), and Word Knowledge and Paragraph Comprehension (combined at the composite level to form the VE composite). Table 1 summarizes content changes over time.

In all cases, the direction of ASVAB subtest change from 1966 to 1980 has been toward longer subtests to provide more reliable measurement. This same rationale has steered the Services away from use of individual subtests as predictors, dictating the use of composites for selection and classification and for career exploration in the high school testing program.

Discussions of the specific content changes over time are found in the ASVAB Test Manual (DoD, 1984a) and the ASVAB Counselor's Manual (AACD, 1984). Weeks et al. (1975) reviewed, evaluated, and compared all the Air Force's aptitude batteries used from 1948 to 1975, including ASVAB Forms 1-3 (Form 4 was developed during this time period but was never used). Their report gives a good perspective on the fluctuation and consistencies in the content of one Service's classification battery over that time period. Notable changes in the ASVAB subtest content from Form 1 to the present include the deletion of Space Perception, Attention to Detail, and Tool Knowledge, and the addition of Numerical Operations, Mathematics Knowledge, and Paragraph Comprehension.

ASVAB Forms 1 and 2 were used in the high school testing program and Form 3 was used by the Air Force for selection and classification from 1973 to 1975. Appendix A provides a more detailed summary of information on ASVAB Forms 1-3.

Score-Scale Changes Over Time

As stated in the Introduction, each of the four Services relies on a unique set of composites to select and classify applicants; each also uses a unique metric for reporting its specific composite scores. In addition, however, all subtests are transformed to a common standard score metric based on a representative sample of 1980 American Youth (see DoD, 1980, 1982b; Maier & Sims, 1986; Wegner & Ree, 1985).

It is important for manpower planners to have the capability to select and classify applicants on the basis of a score-scale that is meaningfully referenced to the manpower pool from which the Services are expected to draw their recruits. For this reason, the score–scale was redefined in October 1984. The previous score-scale (often referred to as the "1944 score-scale" or the "World War II scale") was referenced to the men under arms in December 1944. In October 1984 the score-scale was changed to one based on 1980 American youth (males and females, ages 18 to 23 years) to provide manpower planners with a more up-to-date reference population.

Changes in Military Standards and Policy

The nature of the military occupations, and consequently the classification structure of the Services' occupational specialties, changes over time. Lawrence, Waters, and Perelman (1983) provided a general overview of the types of issues that influence and place demands on the Military selection and classification systems. The aptitude screening measures operate in conjunction with other standards such as moral, educational, medical, and physical standards.

Results of Selection Policy

Aptitude scores, most notably the AFQT composite score, have, along with educational status, become an index of "quality." Manpower planning judgments about the efficacy of the selection and classification system in the Military are made in terms of the average level of aptitudes expressed in the percentile categories on the AFQT score-scale. In addition to aptitude level, the educational credentials of new enlistees provide another commonly used index of quality. Thus, though valid for prediction of training success (or even job performance), ASVAB composite scores may not be valid for other, less obvious uses of aptitude scores. Inferences about quality based solely on the ASVAB aptitude measures may be invalid in that important aspects of recruit quality, such as motivation, are not considered. Fluctuations in validity coefficients over time have to be interpreted in the light of the total system and the Service's personnel selection and classification policies, general economic and labor market conditions, national politics, and even national-level social policy as reflected in guidance provided by Congress.

The history of the changes in the aptitude and educational screening variables used by the Military is well documented in Eitelberg, Lawrence, Waters, and Perelman (1984). Their thorough and detailed summary presents a thorough overview of the types of issues surrounding the use of tests, test scores, and educational standards in the Military selection and classification system. The study offered two fundamental conclusions: (a) that the Military's screening system functions well when judged in light of the criterion of training success, but that the overall character of recruits has fluctuated due to factors which are not entirely the result of conscious personnel management decisions by the Military; and (b) that the Services should explore prediction of other criteria such as job performance criteria to aid the effort to produce an effective, all-volunteer military force. The implication of the Eitelberg et al. (1984) conclusions is that although DoD manpower planners' ability to predict success in training has remained consistently high, their ability to predict criteria such as job performance or first-term attrition has not matched the success attained with prediction of training success. Many factors outside the control of the managers of the selection and classification testing system influence the character and pattern of cognitive abilities desired in any group of recruits. As Eitelberg et al. (1984) pointed out, the apparent result is that "recruiting outcomes (except at the lower levels of 'quality') bear little relationship to the modifications in selection criteria" (p. 124).

Training and First-Term Attrition Criteria

The ASVAB has been used to predict a number of different categories of criteria: training, first-term attrition, second- and later-term attrition, and job performance or job proficiency. Among these, the most important and often used criterion in military validity studies is training success.

Military training is important because of the way the U.S. Armed Services develop their career enlisted personnel. The Services do not hire experienced and trained Noncommissioned Officers (NCOs) "off the street" in the manner that most civilian companies hire personnel (Waters, Lawrence, & Camara, 1987). In the civilian sector, if people with a particular talent or skill are needed but are not available within the organization, the company advertises and searches the general labor force for individuals with that talent or experience, and hires those that qualify through a tailored selection and/or classification system. The Military must "grow" its own experienced people and develop talent through formal and informal training. The entry-level enlisted training is expensive and represents a sizable investment by the American taxpayer in the skills and future development of its Armed Forces enlisted personnel.

Entry-level occupational training of most recruits begins, following Basic Military Training, with an initial assignment to a formal technical training school. Here another sizable investment in the individual is made. The degree of success attained in that training setting represents an important criterion. Obtained increments in the accuracy of prediction of success in first-term enlistee training are rewarded by substantial cost-avoidance. Cost-avoidance is also realized through avoidance of subsequent recruiting costs in replacing failures, as well as avoidance of lowered morale, force instability, and loss of individual self-esteem from failure (Lawrence, 1984).

Several published reports document the financial cost associated with failure in initial technical training. As mentioned earlier, a report by the Automated Sciences Group and CACI (1988) estimated cost savings associated with increments in validity of .02 (r^2) to be in excess of $80 million per year across all Services. Although costs of obtaining new recruits to replace losses due to training failure varied from Service to Service, from a low of $1,800 per individual recruit to a high of $4,300 per recruit in Fiscal Year 1986 (based on an average recruiting cost of over $3,800), these documentable costs capture only part of the expenses associated with training and training losses. As the Automated Sciences Group and CACI (1988) report states, "Considering only recruiting cost savings ($3800 per accession), a 5% improvement translates to an annual saving of $60 million per year at FY 88/89 accession rates (330,000 Enlistees)." It is axiomatic, given the current structure and functioning of the U. S. Military, that no matter what other criteria are judged to also be important for prediction in the selection and classification system, training success will always be relevant.

Many studies reviewed as part of the present effort document the importance of first–term attrition as another relevant criterion. First-term attrition is a broader criterion than attrition from training for it includes performance on first and subsequent military job assignments -- generally up to 48 months of service. The relevance of first-term attrition is established in much the same way as training success; that is, the cost-avoidance associated with recruiting and training replacements.

Classification

Discussion of classification begins with the Services' unique systems and their particular groupings of related military jobs, occupations, and career ladders. The basis of the grouping or clustering of entry-level enlisted jobs differs considerably from Service to Service; but in all cases, the jobs are clustered by the ASVAB aptitude composite used to select for entry into the job or occupational area.

The Air Force, comprising about 20% of the total Military Force, has over 200 occupational specialties (Air Force Specialty Codes - AFSCs); the Army, with slightly less than half of the annual DoD accessions, has over 350 Military Occupational Specialties (MOSs); the Navy, representing 22% of accessions, has over 200 Ratings; and the Marine Corps, at about 11% of accessions, has over 35 major MOSs. Given the large classification problem facing the Military Services, it is surprising that few studies have focused on the efficiency with which the classification composites classify recruits. Some notable exceptions to this are discussed below.

Exceptional research attention given to classification efficiency is exemplified in studies by Albert (1980); Alley, Treat, and Black (1988); Harris (1976); Maier (1982); Maier and Fuchs (1969, 1972, & 1978); and Maier and Truss (1983, 1985). These studies all clustered jobs or job families. Most of these reports followed the theory of differential classification as originally proposed by Brogden (1955).

According to the theory of differential classification, if each aptitude composite's validity is maximized in terms of its absolute validity, then there will be a maximization of the predicted performance of individuals within a cluster of specialties using the given composite. The maximized predicted performance of jobs will in turn lead to maximized differences between job clusters in predicted performance, thus maximizing the differences in validities between clusters of jobs with differing composites (differential validity).

The reliance of the Services on the theory of differential classification has important ramifications. It assumes that specific abilities can be measured and assessed for prediction of situationally specific criteria. The implications of this assumption will be discussed in Section V: Construct Validity Studies.

The Services also use differing standards and differing levels of the same standard (e.g., physical and moral standards) to select and classify individuals for entry into different occupations, as well as differing aptitude standards for the same type of job or occupational specialty. Eitelberg et al. (1984) discussed these differences in detail.

Complicating the problem of classification are differences between stated operational standards and the informal standards that are used by Service recruiters to select applicants. These informal standards operate to adjust the flow of applicants to the Services and are generally hidden from public view (Waters et al., 1987). These practices result in restricting variance, thus leading to poor estimates of validity.

Opportunities Created by Major Policy Changes and "Mistakes"

Mentioned earlier was the fact that Military manpower policy changes affect the observed relationship among aptitude predictors and criteria. Major manpower policy changes in the late 1960's and early 1970's effected under the rubric of "Project 100,000" adjusted the enlisted aptitude standards to accept individuals in the lower aptitude ranges who would not have previously qualified for entry into the Armed Forces. This policy change provided an opportunity to examine the performance of otherwise unqualified recruits against criteria which included training success, first-term attrition, and job performance. Studies examining the performance of these individuals are reviewed here.

Ramsberger and Means (1987) summarized the overall findings from studies of Project 100,000 new mental standards (NMS) men as follows:

> The NMS men did not perform as well as the overall control group in a number of significant ways. NMS men were more likely than control group members to recycle through basic training (Navy, Marine Corps, Air Force), and to need remedial training (Army, Navy, Air Force). They were less likely to complete skill training (Marine Corps and Air Force), and to be eligible for re-enlistment. (p. vi)
>
> In general, these differences remained even when the comparison group was limited to those in the lowest aptitude-qualified category. (p. vi)

Ramsberger and Means went on to report the results of within-military-job comparisons of NMS men. According to their analysis, there were fewer differences between the NMS men within job, but there were significant differences between performance of NMS men in high-skill and medium-skill jobs (more cognitively complex) than their performance in low-skill jobs.

A second event--discovery of an error in the ASVAB Forms 5, 6, and 7 AFQT score-scale in 1979--presented another opportunity to study performance of otherwise ineligible men. Military personnel researchers capitalized on the opportunity provided by both of these events over the years with a large number of studies of what Greenberg (1980) termed "Potential Ineligibles (PIs)." Results of studies capitalizing on both of these events (DoD, 1969; Greenberg, 1980; Grunzke, Guinn, & Stauffer, 1970; Plag & Goffman, 1967; Plag, Goffman, & Phelan, 1967; Plag, Wilkins, & Phelan, 1968; Ramsberger & Means, 1987; Shields & Grafton, 1983; Vineberg, Sticht, Taylor, & Caylor, 1971; and Vineberg & Taylor, 1972) indicated the training performance of Category IV enlistees nearly equaled that of control groups, with 95% of the new standards men completing basic training as against 98% for a control cohort across all Services. The attrition rates from entry-level technical training schools told a somewhat different story, with 10% of the new standards enlistees leaving entry-level training, versus about 4% in the control group.

The findings regarding PIs admitted during the period of improper ASVAB-AFQT scaling were summarized by Ramsberger and Means (1987) as follows:

> There was little variance in the performance of the PIs and the control groups on any of the four variables [attrition, promotion, re-enlistment eligibility, re–enlistment propensity]. This would indicate that minor adjustments to selection standards are unlikely to have a major impact on the Services, at least in the dimensions included in the present study.

> Large differences were found between high school graduates and non-graduates. Graduates were less likely to leave service prematurely, somewhat more likely to reach grade E-4 or above within three years, and more likely to be eligible for re-enlistment. Generally, graduates and non-graduates were just as likely to re-enlist when eligible to do so. Although the attrition, promotion, and re-enlistment eligibility/propensity rates varied widely by Service, the graduate, non-graduate differences were found across Services.
>
> In regard to job complexity, the magnitude of the PI/control group differences was similar across complexity levels. However, contrary to expectations, performance as indicated by the four suitability variables was actually better in medium- and high-complexity occupations than it was in low-complexity occupations. (p. viii)

III. CRITERION-RELATED VALIDITY STUDIES

Aptitude and the possession of a high school diploma are the two predictors used to gauge the quality of the Services' new recruits (Lawrence, 1984). For purposes of the present review, discussions of studies of the predictive validity of ASVAB-derived aptitude measures will be grouped according to the criterion employed (i.e., prediction of attrition from training, prediction of first-term attrition, and prediction of other job performance criteria). Validity for civilian occupations and synthetic validity will also be discussed. First, however, this section will address criterion problems in general, as well as the reliability of predictor and criterion measures.

Criterion Problems and Sources of Contamination

Wagner, Dirmeyer, Means, and Davidson (1982) provided an overview of the types of criterion problems experienced with military studies. They discussed the problems associated with various types of alternate criteria. Use of training attrition as a criterion, for example, has some unknown amount of contamination. Categorization of training attrition by the Military personnel systems into medical, administrative, academic, or motivational (disciplinary actions) is to some unknown extent inaccurate. In this respect, validity studies in the Military Services are like other validity studies in which criterion contamination presents problems of interpretation.

A criterion problem unique to military validity studies is the issue of managed attrition rates in training. That is, the rates are held to Service-specific limits by "washing back" new recruits (recycling recruits through specific blocks of a training program) rather than failing them when the Service requirement for force manning is extremely high or the recruiting market is extremely poor. This factor also contributes to the contamination of training success criteria (Wagner et al., 1982).

The Services have long been involved in a search for other relevant criteria as well as predictors. This has been true not because training success is irrelevant or ASVAB composites do not predict well, but because there are other criteria and associated predictors relevant to different types of personnel policy decisions. As an example, educational status (possession of a high school diploma) is well established in the literature (DoD, 1981a; Flyer & Elster, 1983; Guinn, 1977; Hiatt & Sims, 1980; Kantor & Guinn, 1975; Lawrence, 1984; Martin, 1977) as highly related to first-term attrition as well as to training attrition (pass/fail). In fact, failure to complete the first term of enlistment and failure or success in training are the criteria most often used in the Military.

Reliability of Criterion Measures

Only two studies were found that make assumptions about the reliability of criteria measures used routinely by the Military: Foley (1986) and Lee and Foley (1986), who assume a final technical school grade has a reliability of .90. This assumption is not particularly unreasonable, but there is no way to determine if it is correct. Recent efforts to develop job-performance-related measures have included reliability estimates of these measures or the raters in these methods.

Reliability of ASVAB Subtests and Composites

All the major forms of reliability estimates have been studied and reported in the literature for the ASVAB subtests and composites. The most frequently used forms of reliability estimates--Coefficient Alpha (in the KR-20 form for dichotomously scored items) as an internal consistency measure, and the alternate forms (and parallel forms) estimates of the reliability of subtest and composite scores--provide important information about the precision of measurement obtained with ASVAB subtests and composites. The reliability of a measure sets a limit on the validity of that measure. As shown in the formula below, the correlation between a predictor and a criterion, the validity coefficient, r_{yx}, is limited in order of magnitude by the product of the square roots of the reliability coefficients for the predictor (r_{xx}) and the criterion (r_{yy}):

$$r_{tyx} = \frac{r_{xy}}{\sqrt{r_{xx}}\ \sqrt{r_{yy}}}$$

KR-20 estimates of the reliability set the upper limits of the reliability coefficients for the measurement model most frequently used in the military personnel research community--the domain sampling model (Nunnally, 1978). The alternate forms reliability estimates provide an indication as to whether the tests are measuring relatively time-stable attributes of individuals. The close agreement between ASVAB internal consistency estimates of reliability and other reliability estimates reported in the literature are reassuring from this perspective.

To interpret validity information properly, reliability estimates for as many forms of the ASVAB as possible were garnered from the published literature; these are presented in Appendix B. Reliabilities for ASVAB Forms 1-3 are presented in Appendix B for historical comparison, as is the same reliability information for Forms 5, 6, and 7. Appendix B also contains alternate forms reliability estimates for ASVAB Forms 8, 9, and 10 with ASVAB Form 11a from the follow-on set of ASVAB Forms 11, 12, and 13 implemented in 1984.

Subtest reliability estimates, generally in the form of internal consistency coefficients, are provided for the sake of completeness (Ree et al., 1982). However, the parallel forms and alternate forms estimates of reliability, which are more relevant for selection and classification composites (parallel forms estimates), are presented in Tables 6 and 7. These estimates correlating Form 8a with Forms 9a, 9b, 10a, and 10b are taken from Palmer, Hartke, Ree, Welsh, and Valentine (1988). Internal consistency estimates from Ree, Welsh, Earles, and Curran (in press) for ASVAB Forms 15, 16, and 17 are included in Appendix B, and show values comparable to those of the subtests in Forms 11, 12, and 13. No studies have yet been published which estimate the parallel or alternate forms reliability or test-retest reliability of Forms 15, 16, and 17.

The parallel forms reliability estimates indicated in Tables 6 and 7 show the lowest reliability for the two speeded subtests in the ASVAB (NO and CS, .69 and .72, respectively), and the lowest power subtest reliability for PC (.75), which is also the shortest subtest at 15 items. As expected, the composites result in higher reliabilities because they are longer.

Table 6. Parallel Forms Reliability Coefficients[a] (r) of Subtests and Composites of ASVAB Form 8a with Forms 9a and 9b

Subtests[b]	r(9a)	r(9b)	Composites[c]	r(9a)	r(9b)
GS[d]	.79	.80	M	.91	.90
AR	.87	.87	A	.88	.88
WK	.88	.87	G	.93	.91
PC	.67	.67	E	.93	.92
NO	.70	.72	AFQT (old)[b]	.93	.92
CS	.75	.77			
AS	.84	.82			
MK	.84	.84			
MC	.78	.77			
EI	.72	.71			

Note. From Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (p.10) by P. Palmer, D. D. Hartke, M. J. Ree, J. R. Welsh, and L. D. Valentine, Jr., 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

[a]The estimates of the reliability coefficients are correlations with Ns ranging from 690 to 3,860 in Form 9a and from 680 to 3,959 in Form 9b.

[b]Raw scores used to estimate r.

[c]Standard scores used to estimate r.

[d]See Table A-2 for subtest abbreviations and Table A-5 for composite abbreviations.

Table 7. Parallel Forms Reliability Coefficients[a] (r) of Subtests and Composites of ASVAB Form 8a with Forms 10a and 10b.

Subtests[b]	r(10a)	r(10b)	Composites[c]	r(10a)	r(10b)
GS[d]	.80	.80	M	.92	.91
AR	.86	.86	A	.87	.87
WK	.87	.87	G	.92	.92
PC	.69	.69	E	.92	.92
NO	.72	.72	AFQT[b]	.92	.92
CS	.75	.75			
AS	.83	.83			
MK	.84	.84			
MC	.78	.70			
EI	.70	.70			

Note. From Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (p.10) by P. Palmer, D. D. Hartke, M. J. Ree, J. R. Welsh, and L. D. Valentine, Jr., 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

[a]The estimates of the reliability coefficients are correlations with Ns ranging from 1,056 to 6,473 in Form 10a and from 1,047 to 6,538 in Form 10b.

[b]Raw scores used to estimate r.

[c]Standard scores used to estimate r.

[d]For subtest abbreviations see Table A-2; for composite abbreviations see Table A-5.

Test-retest reliabilities of ASVAB Forms 8, 9, and 10 were examined by Friedman, Streicher, Wing, Grafton, and Mitchell (1983) for a sample of approximately 30,000 Army applicants in 1981. The authors indicated that the test-retest scores of applicants were relatively stable, but that the speeded tests showed the effects of practice. This may be arguable, as their results may be more simply explained as regression to the mean. Their findings also indicated that the AFQT was the most stable of all Service composites.

Because the ASVAB is used as a counseling tool in the DoD Student Testing Program, the reliability estimates for the high school composites for Forms 8, 9, and 10 are presented in Table 8 to aid in the interpretation of the High School ASVAB summary validity information presented at the end of this section. The alternate-forms reliabilities for ASVAB Forms 8, 9, and 10 are uniformly high, and of about the same order of magnitude as the Mechanical, Administrative, General, and Electronics (MAGE) composites used by the Services in the operational or production selection and classification testing system. Casual inspection of the reliabilities do not indicate any systematic gender or school-grade-related differences in the high school composite reliabilities. The reliability of the Mechanical and Crafts composite is somewhat lower for females than for males, but all reliability estimates for the composites are of the same general order of magnitude.

Table 8. Alternate-Forms Reliability Coefficients for High School Composites--ASVAB Forms 8, 9, and 10 by Gender and Grade Level

Composite	Grade 11	Grade 12	Two-Year college	Youth population
		Men		
Academic Ability	.94	.93	.88	
Verbal	.94	.93	.89	
Math	.93	.93	.92	
Mechanical and Crafts	.92	.92	.91	
Business and Clerical	.94	.93	.90	
Electronics and Electrical	.94	.93	.92	
Health, Social, and Technology	.95	.94	.92	
		Women		
Academic Ability	.92	.93	.88	
Verbal	.93	.93	.89	
Math	.91	.91	.90	
Mechanical and Crafts	.84	.86	.88	
Business and Clerical	.93	.92	.90	
Electronics and Electrical	.91	.92	.90	
Health, Social, and Technology	.92	.92	.90	
		Combined		
Academic Ability	.93	.93	.88	.94
Verbal	.93	.93	.89	.94
Math	.92	.92	.92	.94
Mechanical and Crafts	.89	.90	.92	.93
Business and Clerical	.94	.93	.90	.94
Electronics and Electrical	.93	.93	.92	.94
Health, Social, and Technology	.94.	.93	.93	.95

Note. From Counselors Manual for the Armed Services Vocational Aptitude Battery Form 14 (p. 85) 1984, North Chicago, IL: United States Military Entrance Processing Command.

Data Analysis

Despite the multitude of complications arising from the nature of the military selection and classification systems, sufficient uniformity and large enough sample sizes exist to allow meaningful examination of the validity of ASVAB composites. Though the Services use many different classification composites (as indicated in Table 4 and Appendix A), some composites have historically been defined in identical or similar fashion: the General, (G or GT) composite, the Administrative or Clerical (A or CL) composite, the Electronics (E or EL) composite, and the Mechanical (M or GM) composite. These aptitude composites are used to select and classify new recruits into occupational areas. MAGE composites have been used by the Air Force to cluster entry-level jobs since the early 1950's (Alley et al. 1988). They are used here to summarize training and job performance validity results across the Services, along with validity information on the AFQT and other specific Selector Aptitude Indices (SAIs) that are neither MAGE nor AFQT composites.

All studies containing criterion-related validity information were examined to determine if they contained sufficient information to allow validity coefficients from a given study to be aggregated with other studies' validity data relevant to a particular type of criterion. This meant that if a study's reported validity coefficients were to be aggregated with those of other studies, certain information had to be supplied: the sample size relevant to each validity coefficient, the ASVAB form involved, a clear indication of the type of criterion used, and whether or not the validity coefficients were corrected for restriction in range.

The aggregation of validity information proceeded as follows. The validity coefficients were averaged across jobs within a given study at either the subtest level (if the study report contained Ns for each subtest for each job) or composite level, for each type of criterion reported. Each study that provided sufficient information was summarized in the manner depicted in Table 9. The validity coefficients across studies were then aggregated at either the subtest or composite level, for each major type of criterion. The meta-analyses used in this study did not correct for various types of error as recommended by Hunter et al. (1982). Instead, the authors used simple aggregation of the validity information as recommended by Mullins and Rosenthal (1985).

Table 9 provides study summary information for the validity study of 100 Navy jobs by Booth-Kewley, Foley, and Swanson (1984) for ASVAB Forms 8, 9, and 10. Each of 25 other studies having sufficient information to permit the aggregation of their validity data is summarized in a separate table in Appendix C, with the same type of information as indicated in Table 9. Specifically, this information consists of a weighted average correlation or validity coefficient for each subtest or composite across all jobs in the specific study, the weighted standard deviation of the correlation coefficients for a specified subtest or composite, the total N for each of the averaged correlation coefficients, the number of validity coefficients used in the averaged value, and the Binomial Effect Size Display which is discussed below.

For Table 9, the averaged validity (uncorrected) of a composite or subtest, across military jobs examined within the study, is listed under the column "Mean r." The next column gives the standard deviation of the corresponding average validity coefficients, weighted by the frequency or sample size applicable to each specific job. All averages of correlations were done after Fisher's r to Z transformations. The authors realize there is some controversy surrounding the use of the Fisher's Z transformation (James et al. 1986; Schmidt et al. 1988) but, as discussed previously, elected to use the transformation for convenience. The next two columns in Table 9 show the total number (N) of subjects on which the averaged correlation was based, and then the number of jobs (number r's) comprising the within-study averaged validity coefficient. The last two columns give the Binomial Effect Size Display (BESD) as presented by Rosenthal and Rubin (1982). Rosenthal and Rubin have proposed the BESD as a way of directly interpreting the effect size as a proportional change in 1 percentage point of one variable related to change in another. It is defined as: BESD = $(.50 \pm r/2)$. In other words, the BESD is the proportional increase in success rate for a given observed correlation. Thus, a validity coefficient of $r = .30$ would result in a BESD of a change from a 35% (BESD = .50 - .30/2) success rate to a 65% (BESD = .50 + .30/2) success rate. The range of correlation coefficients indicated in the last two columns of Table 9 represents another way of viewing the effect size, or validity. In the case of analyses of effects sizes in validity studies, the larger the r, the greater the effect size.

Table 9. Study Validity for ASVAB Forms 8, 9, and 10, Composites and Subtests, for Navy Schools (by Course Selector Composite)

Composite	Mean[a] r	Standard[b] deviation	Total N	Number r's	BESD range
M	.326	.025	8,035	9	.342 .657
A	.246	.065	8,035	9	.377 .623
G	.354	.068	8,035	9	.322 .677
E	.391	.049	8,035	9	.304 .696
SAI	.347	.077	8,035	9	.327 .673
Type A Schools (General Technical Composite) Final School Grade (FSG) by Subtest					
(GS)	.397	.128	4,098	11	.301 .301
(AR)	.466	.156	4,098	11	.267 .267
(WK)	.385	.085	4,098	11	.308 .308
(PC)	.352	.085	4,098	11	.324 .324
(NO)	.208	.110	4,098	11	.396 .396
(CS)	.254	.112	4,098	11	.373 .373
(AS)	.309	.130	4,098	11	.346 .346
(MK)	.464	.123	4,098	11	.268 .268
(MC)	.374	.138	4,098	11	.313 .313
(EI)	.341	.093	4,098	11	.329 .329
(VE)	.412	.085	4,098	11	.294 .294
Type A Schools (Mechanical Composite) FSG by Subtest					
(GS)	.400	.218	1,464	9	.300 .300
(AR)	.458	.162	1,464	9	.271 .271
(WK)	.389	.235	1,464	9	.306 .306
(PC)	.399	.206	1,464	9	.300 .300
(NO)	.085	.098	1,464	9	.457 .457
(CS)	.187	.127	1,464	9	.406 .406
(AS)	.427	.130	1,464	9	.287 .287
(MK)	.441	.151	1,464	9	.280 .280
(MC)	.470	.209	1,464	9	.265 .265
(EI)	.390	.191	1,464	9	.305 .305
(VE)	.417	.246	1,464	9	.291 .291
Type A Schools (Electronics Composite) FSG by Subtest					
(GS)	.521	.125	973	6	.239 .761
(AR)	.578	.107	973	6	.211 .789
(WK)	.482	.188	973	6	.259 .741
(PC)	.465	.085	973	6	.267 .733
(NO)	.165	.129	973	6	.418 .582
(CS)	.246	.112	973	6	.377 .623
(AS)	.409	.066	973	6	.295 .705
(MK)	.639	.158	973	6	.181 .819
(MC)	.499	.068	973	6	.251 .749
(EI)	.503	.161	973	6	.249 .751
(VE)	.510	.176	973	6	.245 .755

Table 9. (Continued)

Composite	Mean[a] r	Standard[b] deviation	Total N	Number r's	BESD range
		Type A Schools (Administrative Composite) FSG by Subtest			
(GS)	.369	.055	916	4	.316 .684
(AR)	.410	.077	916	4	.295 .705
(WK)	.323	.080	916	4	.339 .661
(PC)	.284	.056	916	4	.358 .642
(NO)	.180	.048	916	4	.410 .590
(CS)	.194	.072	916	4	.403 .597
(AS)	.316	.076	916	4	.342 .658
(MK)	.361	.038	916	4	.320 .680
(MC)	.385	.063	916	4	.308 .692
(EI)	.355	.052	916	4	.323 .677
(VE)	.340	.074	916	4	.330 .670
		Type A Schools (Electronics Composite) Time to Completion (TTC) by Subtest			
(GS)	-.295	.049	4,243	7	.647 .647
(AR)	-.401	.046	4,243	7	.701 .701
(WK)	-.306	.042	4,243	7	.653 .653
(PC)	-.275	.068	4,243	7	.637 .637
(NO)	-.246	.051	4,243	7	.623 .623
(CS)	-.280	.072	4,243	7	.640 .640
(AS)	-.294	.053	4,243	7	.647 .647
(MK)	-.373	.038	4,243	7	.686 .686
(MC)	-.317	.076	4,243	7	.659 .659
(EI)	-.307	.048	4,243	7	.654 .346
(VE)	-.312	.054	4,243	7	.656 .344
		Type B Schools (Electronics Composite) TTC by Subtest			
(GS)	-.282	.025	5,941	3	.641 .359
(AR)	-.333	.032	5,941	3	.667 .333
(WK)	-.280	.052	5,941	3	.640 .360
(PC)	-.266	.050	5,941	3	.633 .367
(NO)	-.207	.035	5,941	3	.603 .397
(CS)	-.216	.065	5,941	3	.608 .392
(AS)	-.268	.064	5,941	3	.634 .366
(MK)	-.320	.055	5,941	3	.660 .340
(MC)	-.293	.046	5,941	3	.646 .354
(EI)	-.291	.042	5,941	3	.646 .354
(VE)	-.304	.051	5,941	3	.652 .348

Table 9. (Continued)

Composite	Mean[a] r	Standard[b] deviation	Total N	Number r's	BESD range
		Type BE Schools TTC by Subtest			
(GS)	-.464	.069	10,433	25	.732 .268
(AR)	-.590	.120	10,433	25	.795 .205
(WK)	-.393	.066	10,433	25	.697 .303
(PC)	-.377	.092	10,433	25	.689 .311
(NO)	-.312	.092	10,433	25	.656 .344
(CS)	-.329	.095	10,433	25	.664 .336
(AS)	-.370	.078	10,433	25	.685 .315
(MK)	-.629	.127	10,433	25	.814 .186
(MC)	-.485	.075	10,433	25	.742 .258
(EI)	-.404	.141	10,433	25	.702 .298
(VE)	-.419	.066	10,433	25	.710 .290
		Type BE/E Schools (AR+2MK+GS Composite) TTC by Subtest			
(GS)	-.349	.084	4,164	12	.675 .325
(AR)	-.520	.082	4,164	12	.760 .240
(WK)	-.359	.073	4,164	12	.679 .321
(PC)	-.338	.051	4,164	12	.669 .331
(NO)	-.284	.069	4,164	12	.642 .358
(CS)	-.286	.073	4,164	12	.643 .357
(AS)	-.328	.082	4,164	12	.664 .336
(MK)	-.540	.085	4,164	12	.770 .230
(MC)	-.409	.080	4,164	12	.704 .296
(EI)	-.387	.082	4,164	12	.694 .306
(VE)	-.371	.072	4,164	12	.686 .314
		Type BE/E Schools (Multiple Composite) TTC by Subtest			
(GS)	-.393	.149	4,535	15	.696 .304
(AR)	-.486	.167	4,535	15	.743 .257
(WK)	-.387	.121	4,535	15	.693 .307
(PC)	-.382	.117	4,535	15	.691 .309
(NO)	-.243	.080	4,535	15	.621 .379
(CS)	-.296	.083	4,535	15	.648 .352
(AS)	-.376	.097	4,535	15	.688 .312
(MK)	-.459	.152	4,535	15	.729 .271
(MC)	-.420	.098	4,535	15	.710 .290
(EI)	-.406	.154	4,535	15	.703 .297
(VE)	-.406	.116	4,478	14	.703 .297

Table 9. (Concluded)

Composite	Mean[a] r	Standard[b] Deviation	Total N	Number r's	BESD range
		FSG against AFQT			
AFQT (Whites)	.408	.141	3,346	8	.296 .704
AFQT (Blacks)	.204	.118	715	8	.398 .602
AFQT (Males)	.367	.202	2,816	8	.316 .684
AFQT (Females)	.423	.182	633	8	.288 .712

Note. The data are from Predictive validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, 10, against 100 Navy schools (NPRDC-TR-85-15) by S. Booth-Kewley, P. P. Foley, and L. Swanson, 1984, San Diego, CA: Navy Personnel Research and Development Center.

[a]Uncorrected.

[b]Weighted by study sample size.

The study-by-study validity information was then aggregated at the appropriate level of the predictors--subtest or composite--and summarized in the manner shown in Table 10 for ASVAB Forms 8, 9, and 10 subtests. Table 10 is discussed in greater detail in the section on subtest-level validity.

Table 10. ASVAB Forms 8, 9, and 10 Subtests Effect Sizes (Validities[a]) Against Final School Grade

Subtests	Mean[a] r's	SD[b]	Total N	Number r's	BESD[c] range
GS[d]	.64	.35	52,215	13	.18 .82
AR	.64	.25	52,215	13	.18 .82
WK	.63	.29	52,272	13	.19 .84
PC	.64	.40	52,215	13	.18 .82
NO	.49	.26	52,215	13	.26 .74
CS	.44	.17	52,215	13	.28 .72
AS	.49	.19	52,215	13	.25 .75
MK	.63	.25	52,215	13	.19 .81
MC	.58	.25	52,215	13	.21 .79
EI	.60	.38	52,215	13	.20 .80

[a]All subtest validities individually corrected for restriction in range, based on Fisher's r to Z transformations for mean validities.

[b]Weighted by sample size.

[c]Binomial Effect Size Display.

[d]Definitions of subtest abbreviations are found in Table A-2.

Prediction of Training Success

Though individual subtests of the ASVAB are never used in isolation to make personnel selection and classification decisions, validity information on subtests is presented and discussed because it contributes to the overall validity of the battery. All summary criterion–related validity evidence presented in this section is divided into three major types of training criteria: technical training final school grade (FSG), self–paced technical training time-to-completion (TTC, usually measured in days), and training attrition (measured as pass/fail or graduated/not graduated). Training attrition will be discussed in the job performance section under attrition-related studies.

Final School Grade (FSG)

Subtest Validity. Table 10 displays mean validity correlation coefficients, sample sizes, and standard deviations (weighted by study sample sizes) of subtest validity correlations of ASVAB Forms 8, 9, and 10 against FSG. The correlations in Table 10 represent averaged validities corrected for restriction in range, from Weltin and Popelka (1983) for Army occupations; from Booth-Kewley, (1984b), Booth-Kewley et al. (1984), and Curtis, Booth-Kewley, and Swanson (1984) for Navy occupations; from Maier and Truss (1983) for Marine Corps specialties; and from Jones (1988) for a sample of Air Force occupations. Fisher's r to Z transformation was performed before averaging and then the averaged Z value was back–transformed. Table 11 provides an "author" table to inform the reader as to the sources of the data and the extent of the contribution of each individual study to the averaged validity values and to the number of military jobs from each individual study.

There are no surprises in the data in Table 10, with the possible exception of the noticeably lower averaged validity for the AS subtest. The lower validities for the two speeded subtests (NO and CS) are consistent with their relatively lower reliability. These results are also consistent with the results of validity generalization studies (Jones, 1988; Rossmeissl & Stern, 1983; see discussion of validity generalization in Section IV).

The most striking feature of Table 10 is the order of magnitude of the corrected subtest validities compared to the uncorrected composite validities. Subtest validities should be interpreted in the context of the observed variability of the subtest coefficients--all have

Table 11. Author Table for ASVAB Forms 8, 9, and 10 for Subtests Against Final School Grade (Table 10)

Subtests	No. of Entries	Authors
All (except VE)	1	Booth-Kewley, S. (1984b)
	4	Booth-Kewley, S., Foley P. P., and Swanson, L. (1984)
	2	Curtis, J. S., Booth-Kewley, S., and Swanson, L. (1984)
	1	Maier, M. H., and Truss, A. R. (1983)
	4	Jones, G. E. (1988)
	1	Weltin, M. M. and Popelka, B. A. (1983)
VE	4	Booth-Kewley, S., Foley, P. P., and Swanson, L. (1984)
	2	Curtis, J. S., Booth-Kewley, S., and Swanson, L. (1984)

higher standard deviations when compared to the standard deviations of the validity coefficients of the composites discussed in the next section. This is expected because the composites have higher reliabilities than individual subtests, as well as larger average sample sizes. Minimum and maximum subtest validity coefficients in all these studies were corrected for restriction in range and represent validities against FSG for specific occupational specialties.

There is Service-by-Service variation in the subtest validity coefficients that is not apparent from the aggregated coefficients displayed in Table 10. In general, the magnitude of the Army coefficients ranged from lows of .36 for AS to highs of .55 for AR in Weltin and Popelka (1983); for the Air Force study (Jones, 1988) the range was from .40 for AS to .84 for AR ; for the Navy (Booth-Kewley et al., 1984), the range was from .36 for NO to .85 for MK.

The Army validities were uniform and in the .40s to mid .50s (corrected for restriction in range); the Air Force validities were higher and showed more variability between subtest validities across Air Force jobs in each of the four Air Force occupational clusters--Mechanical (M), Administrative (A), General (G), and Electronics (E). By comparing averaged validities across jobs and across Services, these differences are observed. For this reason, study by-study information is provided in Appendix C.

Subtest and composite validity information for the first generation of the ASVAB (Forms 1, 2, 3) and for the second generation of the ASVAB (Forms 5, 6, 7) is presented for the three major types of training criteria (FSG, TTC, and Job Performance) in Appendices D, E, and F, respectively. These appendices also contain author tables similar to Table 11.

Composite Validity. The Services' selection and classification composite data from studies with criterion-related validity information were analyzed according to common groupings across the four Services. Four of the classification composites (M, A, G, and E) have used common subtest definitions over the years and consequently provided a convenient way of summarizing composite validity across Services. Table 12 summarizes for ASVAB Forms 8, 9, and 10 the composite effect sizes (validity r's) for the MAGE composites, for the AFQT, and for averaged coefficients across the Selector Aptitude Indexes (SAIs) (for uncorrected composite validities). Table 13 contains the author table corresponding to Table 12. For purposes of comparison, ASVAB Forms 8, 9, and 10 composite validities, corrected for restriction in range, are presented in Table 14 (from one large study--Booth-Kewley et al., 1984).

Table 12. ASVAB Forms 8, 9, and 10 Composites Effect Sizes (Validities[a]) Against Final School Grade

Composites (uncorrected)	Mean r's	SD[b]	Total N	Number r's	BESD[c] range	
AFQT	.44	.09	224,048	19	.28	.72
M[d]	.47	.06	216,011	16	.27	.73
A	.46	.11	151,665	14	.27	.73
G	.54	.12	35,111	12	.23	.77
E	.48	.08	174,816	15	.26	.74
SAI	.47	.08	419,790	13	.26	.74
VE	.44	.07	8,389	6	.28	.72

[a]All validities individually corrected for restriction in range.
[b]Weighted by sample size.
[c]Binomial Effect Size Display.
[d]Composite abbreviations are found in Table A-5.

Table 13. Author Table for ASVAB Forms 8, 9, and 10 for Composites Against Final School Grade (Table 12)

Composites	No. of entries	Authors
M	1	Booth-Kewley, S., Foley, P. P., and Swanson, L. (1984)
	1	Maier, M. H., and Truss, A. R. (1983)
	4	McLaughlin, D. H., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., and Wang, M. (1984)
A	1	Maier, M. H., and Truss, A. R. (1983)
	3	McLaughlin, D. H., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., and Wang, M. (1984)
G	1	Booth-Kewley, S., Foley, P. P., and Swanson, L. (1984)
	1	Maier, M. H., and Truss, A. R. (1983)
E	1	Booth-Kewley, S., Foley, P. P., and Swanson, L. (1984)
	1	Maier, M. H., and Truss, A. R. (1983)
	3	McLaughlin, D. H., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., and Wang, M. (1984)
SAI	4	Booth-Kewley, S., Foley, P. P., and Swanson, L. (1984)
	2	Maier, M. H., and Truss, A. R. (1983)
	10	McLaughlin, D. H., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., and Wang, M. (1984)

Table 14. ASVAB Forms 8, 9, and 10 Composites Against Final School Grade Corrected for Restriction in Range

Composite[a]	Mean r's	SD[a]	Total N	Number r's	BESD range	
AFQT	.50	.07	8,035	9	.25	.50
M	.47	.09	8,035	9	.26	.74
A	.37	.06	8,035	9	.32	.68
G	.52	.08	8,035	9	.24	.76
E	.54	.11	8,035	9	.23	.77

Note. The data are from Predictive validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 against 100 Navy schools (NPRDC-TR-85-15) by S. Booth-Kewley, P. P. Foley, and L. Swanson, 1984, San Diego, CA: Navy Personnel Research and Development Center.

[a]Definitions for abbreviations are found in Table A-5.

[b]Weighted by sample size.

Table 15 shows the average validity for the M, A, G and E composites for ASVAB Forms 11, 12, and 13 estimated from validities of ASVAB Forms 8, 9, and 10 from one study by Maier and Truss (1985). The corrected MAGE composite validities shown in Table 14 for ASVAB Forms 8, 9, 10 are slightly lower in magnitude than the corrected composite validities for ASVAB Forms 11, 12, and 13 indicated in Table 15. There is no obvious explanation for this, but it is important to point out that the ASVAB Forms 11, 12, and 13 composite validities are estimated, and that both sets of aggregated composite validities are corrected for restriction in range.

Table 15. ASVAB Forms 11, 12, and 13 Composites Against Final School Grade Corrected for Restriction in Range

Composites[a]	Mean r	SD[b]	Total N	Number r's	BESD range	
M	.57	.10	16,478	6	.22	.78
A	.59	.11	16,478	6	.21	.79
G	.61	.08	16,478	6	.20	.80
E	.61	.10	16,478	6	.19	.81

Note. The data are from Validity of the Armed Services Vocational Aptitude Battery Forms 8, 9, and 10 with applications to Forms 11, 12, 13 and 14 (CNR-102) by M. H. Maier and A. R. Truss, 1985, Alexandria, VA: Center for Naval Analyses.
[a]Weighted by sample size.
[b]Definitions for abbreviations are found in Table A-5.

The uncorrected composite validities for final school grade are lower in magnitude for ASVAB Forms 6 and 7 (summarized validity information included in Appendix D) than for ASVAB Forms 8, 9, and 10 indicating some improvement in prediction from the previous generation.

The generally lower validity of the A composite for predicting FSG is a consistent finding across generations of the ASVAB. Some portion of this lower validity can be attributed to the A composite's consistently lower reliability. The lower reliability and validity of the A composite is probably because over the years, the Services have consistently constructed the A composite with as many speeded subtests as are available in the battery, in the belief that speeded tests should be good predictors of success in administrative and clerical jobs. These data clearly indicate a consistent trend across forms of the ASVAB that show not only less reliability, but consistently less predictive validity for the A composite. The finding of less predictive validity for clerical or administrative composites is consistent with findings in the literature that speeded tests are more sensitive to administration conditions (McLaughlin, Rossmeissl, Wise, Brandt, & Wang, 1984; Sims & Hiatt, 1981; Wegner & Ree, 1985).

While investigating the validity of ASVAB Forms 8, 9, and 10 for predicting final school grades, Wilbourn, Valentine, and Ree (1984) found that the AFQT added .16 to the average validity of the A composite, while adding only .07 to the M composite, .006 to the G composite, and .006 to the E composite. Such a relatively large increment to the validity of the A composite clearly indicates that the Services could do better in prediction for administrative or clerical-type military occupations.

Investigations of specific, tailored, screening systems using ASVAB aptitude indices in conjunction with other interest inventories and biographical, attitudinal, demographic and educational variables have led to some general conclusions about the usefulness of the ASVAB for prediction of training success (Flyer, 1988; Flyer & Elster, 1983; Guinn, Tupes, & Alley, 1970a, 1970b; Guinn, Wilbourn, & Kantor, 1977; Leisey & Guinn, 1977; Oslund & Clark, 1984; Valentine, 1977). The ASVAB aptitude indices predict training success very well by themselves, as evidenced from the criterion-related validity studies. Aptitude indices themselves seem to make the largest unique contribution to prediction of success in training (May, 1986; Valentine, 1977); but other variables make more of a contribution to other criteria, such as job performance or first-term attrition, the further away in time one moves from entry-level training (Hawley, Mullins, & Weeks, 1977). Specific, tailored prediction can almost always improve on the aptitude indices alone when non-cognitive ability variables--particularly educational status, specially designed aptitude measures, or interest measures or interest surrogates--are used in regression-weighted equations to predict training success.

Time-to-Completion (TTC)

Table 16 shows the mean validity for each of the M, A, G, E, and AFQT composites and the Selector Aptitude Indices (where these are different from either the MAGE or AFQT) for ASVAB Forms 8, 9, and 10, for the TTC criterion.

Table 16. ASVAB Forms 8, 9, and 10 Subtests and Composites Effect Sizes (Validities) Against Time-to-Completion (TTC)

	Mean r's	SD[a]	Total N	Number r's	BESD range	
			Subtests[b]			
GS	-.32	.15	41,970	9	.66	.34
AR	-.43	.19	41,970	9	.72	.28
WK	-.30	.11	41,767	8	.65	.35
PC	-.28	.11	41,970	9	.64	.36
NO	-.24	.06	41,970	9	.62	.38
CS	-.26	.06	41,970	9	.63	.37
AS	-.29	.09	41,970	9	.65	.35
MK	-.44	.23	41,970	9	.72	.28
MC	-.35	.13	41,970	9	.68	.32
EI	-.32	.10	41,970	9	.66	.34
			Composites[c]			
AFQT	-.30	.04	30,334	8	.65	.35
M	-.25	.03	30,334	8	.63	.37
A	-.28	.03	30,334	8	.64	.36
G	-.28	.05	30,334	8	.64	.36
E	-.36	.07	30,334	8	.68	.32

[a]Weighted by sample size.
[b]Definitions for abbreviations are found in Table A-2.
[c]Definitions for abbreviations are found in Table A-5.

These data show a finding that is repeated across forms and generations of the ASVAB: lower subtest and composite validities against this type of criterion. The data for this type of validity are derived almost exclusively from Navy studies and represent the time required for a new enlistee to complete a self-paced, entry-level technical training course. These data are generally from Navy Type BE/E schools (electronics schools) and the consistently lower validities may be due not only to greatly increased restriction in range of abilities for this type of school, but also to peculiarities in the nature of this criterion. There is often no incentive for smarter students to finish a self-paced course of instruction early, as they may have to wait in dormitories for the next block of instruction to begin, or they may even be assigned to undesirable special details such as sweeping or cleaning. There may be, therefore, some serious contamination in the criterion measure that is operating to restrict the amount of observed variance in the validity correlations. That is, time-to-completion of a self-paced course may be influenced by factors unrelated to a recruit's ability.

Job Performance Measures as Criteria

Because many different types of job performance measures (JPM) have been employed over the years, a simple and specific classification of these measures is not possible. Instead, all measures of job performance (with the exception of first-term attrition, which is discussed in a later section) were averaged together across all types of measures. More detailed discussion of validity of composites against specific performance measures will ensue where appropriate. This necessary averaging across varying types will have the effect of introducing variance in the observed validities of the ASVAB for prediction of job performance measures. Much of the JPM criterion information for ASVAB Forms 6 and 7 and ASVAB Forms 8, 9, and 10 is from the Army, which uses the Skill Qualification Test (SQT) as a job performance measure. The SQT has two components: a hands-on performance test and a written test of an incumbent's job knowledge.

Summary validity information for ASVAB Forms 8, 9, and 10 on the M, A, G, E, and AFQT composites, and Selector Aptitude Indices for prediction of the JPM criterion are presented in Table 17. These data are based on two studies--one Marine Corps study and one Navy study--and are not averaged across studies, but are averaged across the jobs within each of the two studies. The aggregated subtest and composite validities for ASVAB Forms 8, 9, and 10 show high values for predicting the JPM criterion. The subtest-level information comes from Maier and Hiatt (1984) and concerns the JPM validity of the ASVAB Forms 8, 9, and 10 subtests (corrected for restriction in range) for only the hands-on portion of a job performance test. The subtest validities are comparable, though somewhat lower in magnitude, to the corrected subtest validities for the criterion of final school grade.

Table 17. ASVAB 8, 9, and 10 Subtests and Composites Effect Sizes (Validities) Against Job Performance Measures

	Mean r	SD[a]	Total N	Number r's	BESD range	
			Subtests[b]			
GS	.55		294	2	.24	.76
AR	.44		294	2	.32	.68
WK	.52		294	2	.25	.75
PC	.61		294	2	.19	.81
NO	.50		294	2	.41	.59
CS	.40		294	2	.30	.70
AS	.45		294	2	.28	.72
MK	.53		294	2	.27	.73
MC	.57		294	2	.26	.74
EI	.52		294	2	.27	.73
			Composites[c]			
AFQT	.35	.04	16,283	5	.32	.68
M	.44	.04	65,193	5	.28	.72
A	.44	.09	65,193	4	.28	.72
G	.46	.03	65,193	3	.27	.73
E	.47	.03	65,193	3	.27	.73
SAI	.47	.05	65,193	9	.27	.73

Note. The data for subtests are from An evaluation of using job performance tests to validate ASVAB qualification standards (CNR-89) by M. H. Maier and C. M. Hiatt, 1984, Alexandria, VA: Center for Naval Analyses. The data for composites are from Validation of current alternative Armed Services Vocational Aptitude Battery (ASVAB) area composites, based on training and Skill Qualification Test (SQT) information in fiscal year 1981 and 1982 (ARI-TR-651,AD-A156 807) (p. 22) by D. H. McLaughlin, P. G. Rossmeissl, L. L. Wise, D. A. Brandt, and M. Wang, 1984, Alexandria, VA: Army Research Institute.

[a]No SD calculated for subtests.

[b]Definitions for abbreviations are found in Table A-2.

[c]Definitions for abbreviations are found in Table A-5.

The JPM corrected composite validities are of lower magnitude than the subtest validities, but are less variable than the subtest coefficients. The validity coefficients were corrected for restriction in range, but the study reports were not specific as to the type of correction (multivariate or univariate), nor to the type of population. These validities result from a single study by McLaughlin et al. (1984). Most noticeable is that the AFQT composite has lower validity for the job performance criterion compared to validities of the MAGE composites. Such was not the case in the prediction of final school grade, where the AFQT validity coefficients were similar to those for the other MAGE composites. The SQT performance measure is taken at a later point in time--after completion of initial training in the recruit's first term. It may be that general trainability becomes a less important factor later in time, and specific abilities and/or experience assume greater importance the longer a recruit stays in a particular job. That the more specific aptitude area composites predict such criteria better than the AFQT does is interesting and needs to be explored in future validity research.

Though the composite results presented in Table 17 are based on only one study, they are based on large sample sizes. These data represent analyses of the written portion of the SQT for over 65,000 FY 81 and FY 82 Army recruits.

The effort to develop reliable and cost-effective job performance criteria came as a result of the norming error referred to earlier in the historical perspective section of this report. During its inquiry concerning the norming error (DoD, 1980), Congress learned that aptitude scores were validated only against training success and not against criteria of job performance. Congress subsequently required the Department of Defense to establish a link between aptitudes (as measured by the ASVAB) and job performance. This continuing criterion development effort has yielded useful validity data on the prediction of a variety of job performance measures. A series of annual reports by the Department of Defense was included in the present review (e.g., DoD, 1981b, 1987). These reports document the Services' efforts to develop hands-on performance measures in order to establish linkages to entry-level aptitudes. Considerable resources are being expended to develop hands-on performance measures. Some of the measures used by the Services have been developed and used in the past as performance measures, and where available, summary validity information of ASVAB subtests and composites was included in the aggregate validity data presented in Table 17.

In a programmatic effort to understand the relationship of aptitudes to job performance, a number of studies (Fox, Taylor, & Caylor, 1969; Vineberg, Sticht, Taylor, & Caylor, 1971; Vineberg & Taylor, 1972; Vineberg, Taylor, & Caylor, 1970; and Vineberg, Taylor, & Sticht, 1970) examined the relationship between aptitude groupings based on the AFQT (three levels - high, medium, and low aptitude) and three criteria of job performance (supervisor ratings, job sample tests, and job knowledge tests). Table 18 presents correlations for two aptitude levels, CAT IV and non-CAT IV aptitude groups, from Vineberg and Taylor (1972). Note that individuals at the CAT V level are not accessed and the non-CAT IV group contains individuals at the CAT I to CAT III levels. The subjects in the study were further divided into high and low reading groups, as shown in Table 19. The authors concluded that, because of the high correlations between job sample tests and job knowledge tests, job knowledge tests could be used in lieu of job sample tests--where skill required for the job sample was minimal and the job knowledge required to actually perform the work or task was clearly specified and tested by the job knowledge tests. There appears to be little separation between AFQT categories until the effect of experience (as measured by months on the job in this study) is partialed out of the relationship between the criterion measures. The fact that this study used adjacent category research tends to obscure the magnitude of the relationship in question, in that the division of the groups into adjacent AFQT categories may serve to attenuate any observed relationship. For example, if one were investigating the relationship between height and basketball ability, and obtained two samples drawn by dividing a pro–basketball team in half, one would probably observe a very low correlation between ability (however measured) and height--opposite to the true state of affairs. Still, these early results provide an interesting backdrop for more recent efforts to develop on-the-job performance measures and because of the obtained relationship between job knowledge and job performance measured by job sample tests.

Table 18. Correlations of Job Sample and Job Knowledge for Category IV and Non-Category IV Subgroups[a]

	Armor Crewman		Repairman		Supply Specialist		Cook	
Sample	N[b]	r	N	r	N	r	N	r
Zero-Order Correlations								
Entire Sample	368	.68	360	.59	380	.72	366	.58
Category IV	186	.65	178	.59	188	.69	186	.54
Non-Category IV	182	.66	182	.54	192	.72	180	.54
Partial Correlations[c]								
Entire Sample	368	.49	360	.49	380	.65	366	.50
Category IV	186	.34	178	.47	188	.58	186	.42
Non-Category IV	182	.49	182	.45	192	.64	180	.45

Note. From Performance in four Army jobs by men at different aptitude (AFQT) levels: Relationships between performance criteria (TR-72-23) (p. 22) by R. Vineberg and E.N. Taylor, 1972, Alexandria, VA: Human Resources Research Organization.

[a]All correlations are significantly different from zero ($p \leq .05$).

[b]Ns for subsamples will occasionally not add to total for entire sample. Where pertinent data were missing in Army records, the case was omitted from analysis.

[c]Correlations with the effects of Months On the Job (MOJ) partialled out.

Table 19. Correlations of Job Sample and Job Knowledge for Low Reading and High Reading Ability Subgroups[a]

	Armor Crewman		Repairman		Supply Specialist		Cook	
Sample	N[b]	r	N	r	N	r	N	r
Zero-Order Correlations								
Entire Sample	368	.68	360	.59	380	.72	360	.58
Low Reading	178	.65	174	.45	186	.67	176	.42
High Reading	190	.57	186	.47	194	.64	190	.54
Partial Correlations[c]								
Entire Sample	368	.49	360	.49	380	.65	366	.50
Low Reading	178	.34	174	.44	186	.59	176	.49
High Reading	190	.51	186	.48	194	.64	190	.41

Note. From Performance in four Army jobs by men at different aptitude (AFQT) levels: Relationships between performance criteria (TR-72-23) (p. 22) by R. Vineberg and E. N. Taylor, 1972, Alexandria, VA: Human Resources Research Organization.

[a]All correlations are significantly different from zero ($p \leq .05$).

[b]Ns for subsamples will occasionally not add to total for entire sample. Where pertinent data were missing in Army records, the case was omitted from analysis.

[c]Correlations with the effects of MOJ partialled out.

Attrition-Related Studies

Much of the research relating ASVAB aptitude scores to training attrition and first-term attrition criteria exists in the form of categorical rates and data; i.e., changes in attrition rates across Services by AFQT categories. Thus, a summary of effect size in terms of correlations is not appropriate. First-term attrition could be considered a job performance criterion, but is discussed separately here for the sake of clarity. The relevance of first-term attrition was mentioned previously: Premature loss of an individual represents a loss of investment by the Services and the taxpayers. As Lawrence (1984) noted, first-term attrition is also the most extensively researched job performance criterion. Buddin (1984) estimated that across the Services the average first-term attrition rate was 30%.

The classes of predictors of first-term attrition can be broadly divided into characteristics of the individual and characteristics of the job. Many studies have explored numerous individual characteristics as predictors. These studies include attempts to build manpower models describing the attrition among the Services (Albert, 1980; Armor, Fernandez, Bers, Schwarzbach, Moore, & Cutler, 1982; Fernandez & Garfinkle, 1984; Flyer & Elster, 1983; Marcus & Lockman, 1981; May & Mayberry, 1986; Wardlaw, 1983), as well as other studies simply relating individual characteristics, general economic factors, and demographic variables to first-term attrition (Buddin, 1984; Camara & Lawrence, 1987; Flyer, 1988; Grafton & Horne, 1985; Grissmer & Kirby, 1985; Guinn, 1977; Martin, 1977; Vitola, et al., 1977). These studies lead to the fundamental conclusion that though many variables are related to first-term attrition, three remain paramount: education (high school diploma, graduate or not), aptitude, and age (Buddin, 1984).

Rosenthal and Lawrence (1988) and Lawrence (1988) found that job characteristics such as working conditions, physical demands, etc. were positively related to first-term attrition, but obtained conflicting results regarding the interaction of aptitude and job characteristics in predicting attrition. Rosenthal and Lawrence (1988) found that the prediction of first-term attrition using job characteristics was essentially the same for high aptitude recruits (defined as AFQT Category I & II recruits) as for medium aptitude recruits (defined as AFQT Categories IIIa and IIIb). Lawrence (1988), however, found that aptitude did moderate the relationship between job complexity (categorized as high, medium, and low) and attrition. Above-average recruits (as defined in the Rosenthal and Lawrence study) had lower attrition rates and higher promotion rates as the complexity of the job increased. Though promotion rates were lower and attrition higher in the less complex or easier jobs, the high aptitude recruits tended to have more favorable attrition and promotion rates overall (Lawrence, 1988).

Validity of the High School Composites

The validity of the ASVAB high school composites for use as counseling tools in career exploration depends on the model of counseling employed. One model proposes that ability composites have a strong relationship with success in training in an occupational area, and that knowledge about a person can be used to counsel the individual about chances of successful completion of training in a variety of occupational areas. This model is referred to as the "Prediction of Success Model."

Another widely used model is the Profile Similarity Model, which uses the same aptitude information, but bases recommendations for career exploration on similarity of patterns of personal characteristics of those counseled to those in the career field (Prediger, 1987a, 1987b). This model depends much more heavily on the existence of differential validity among measures of specific ability.

Our review of the literature identified no published study of the predictive validity of the ASVAB high school composites for civilian career success or occupational choice other than a number of studies whose results were published in the Technical Supplement to the Counselor's Manual for ASVAB 14 (DoD, 1984b). Because there were no other published studies which contained more recent validity information for ASVAB Form 14, aggregated validity information on Form 14 cannot be presented.

The DoD relies on generalization of the validity of the ASVAB from military to civilian occupations to support the use of the ASVAB in the high school testing program. Table 20 shows the averaged validities for the high school ASVAB form against a criterion of final school grade in 54 Marine Corps technical training courses. All data for this table come from Maier and Truss (1985), who calculated the Form 14 composites on a sample of over 16,000 Marine Corps recruits who completed ASVAB Form 8, 9, or 10. The subtest composition of the high school composites were as previously indicated in Table 5.

Table 20 shows the uniformly high average composite validities are about one standard deviation below the theoretical limit of validity if one assumes a criterion reliability of .60 (see Schmidt and Hunter, 1977, for a rationale for using this assumed value of criterion reliability). The somewhat lower validities of the additional, factor-based composites (T - Technical composite; S - Speed composite)--added for the sake of comparison by Maier and Truss (1985)--runs counter to the arguments and results reported by Hunter (1984) and Hunter et al. (1985), which showed that factor-based composites had higher validities. This result may be an artifact of the way in which Hunter et al. (1985) constructed their factor-based composites, essentially building the most 'g'-saturated composite possible with a subset of the ASVAB subtests. General cognitive ability, or 'g' factor, has been shown to be a very good predictor of most military job training success criteria (Ree & Earles, 1990).

Table 20. ASVAB Form 14 Effect Sizes (Validities) for High School Composites Against Final School Grade

Composites[a]	Mean r	SD[b]	Total N	Number r's	BESD range	
Academic						
AA	.60	.09	16,478	6	.20	.80
Q	.59	.10	16,478	6	.21	.79
V	.57	.08	16,478	6	.21	.79
Occupational						
B & C	.59	.11	16,478	6	.21	.79
M & C	.57	.10	16,478	6	.22	.78
HS & T	.61	.08	16,478	6	.20	.80
E & E	.61	.10	16,478	6	.19	.81
Factor						
T	.52	.11	16,478	6	.24	.76
S	.45	.09	16,478	6	.27	.73

Note. All data are from Validity of the Armed Services Vocational Aptitude Battery Forms 8, 9, and 10 with application to Forms 11, 12, 13, and 14 (CNR-102) by M. H. Maier and A. R. Truss, 1985, Alexandria, VA: Center for Naval Analyses.

[a]Definitions of abbreviations are found in Table 5.

[b]Weighted by sample size.

The positive manifold indicated by the intercorrelations shown in Table 21 from the Counselor's Manual (AACD, 1984) has been taken as an indication of the lack of differential validity in the ASVAB. Prediger (1987a, 1987b) has maintained that the Prediction of Success Model is not as useful for counseling purposes as the Profile Similarity Model, mainly because of a lack of differential validity, and because of a lack of a formal, standard criterion of occupational success. The disadvantage of the Profile Similarity Model, however, is that it provides no empirical index of the "closeness" of any two profiles.

Table 21. Intercorrelation of High School Composites for the 1980 Youth Population

Composite	AA	Verbal	Math	MC	BC	EE
Academic Ability (AA)						
Verbal	.93					
Math	.91	.78				
Mechanical and Crafts (MC)	.82	.78	.79			
Business and Clerical (BC)	.91	.88	.87	.71		
Electronics and Electrical (EE)	.92	.88	.93	.91	.86	
Health, Social, and Technology	.96	.90	.89	.93	.86	.94

Note. From Counselors Manual for the Armed Services Vocational Aptitude Battery Form 14 (p. 90) by the American Association for Counseling and Development, 1984, North Chicago, IL: U.S. Military Entrance Processing Command.

The implicit framework from which the Services operate in their selection and classification systems is the theory of differential classification explained earlier. From this theoretical position, one maximizes whatever differential validity one has by maximizing the predictive validity of each composite. Accordingly, the intercorrelations between composites do not matter. The Profile Similarity Model requires, or is much more dependent on, the existence of differential validity in a given battery of aptitude measures. It is not a matter of which model is better. It is a question of the best use of available information to counsel an individual, and aptitude information about the predicted success of an individual in an entry-level training program is useful.

Prediger (1987a) found that ASVAB Form 14 Job Cluster Scales, comprised of ASVAB Form 14 composite scores and self-estimated ratings of ability, better differentiated between occupational groups than did the ASVAB Form 14 composites alone. These results should be interpreted only in the light of the counseling model used. The use of any additional information would likely improve prediction, and the Profile Similarity Model uses other types of information besides aptitude scores. Prediger (1987a, 1987b) examined the ASVAB's use in counseling for use in the Profile Similarity Model, yet the model used by the DoD is the Prediction of Success Model. The Prediction of Success Model enjoys empirical success, as demonstrated in the previous section of this review. Though there are problems associated with defining occupational success, one cannot discount the strong empirical relationships between ASVAB predictors and success in entry-level training. Also, the validities are based on very large numbers of young people of an age where they are making entry-level, occupational choices, not unlike their civilian counterparts. The issue, it appears, therefore, is not one of the absence or presence of differential validity, nor is it one of choice of counseling models; the issue is one of counseling practice in which all information of potential use to the student and the counselor is made available and properly interpreted. The addition of more information to a counseling system can be expected to enhance its overall validity. The existence of valid predictors of entry-level occupational training, based on large, representative samples of American youth should play a meaningful role in a high school student's career exploration.

As noted in the previous sections, the well-established predictive utility of the ASVAB and its composites for success in training, and its usefulness in the prediction of job performance measures as well, provides compelling evidence that the ASVAB high school composites are useful predictors of training success in a variety of civilian occupations.

The validity of the ASVAB high school composites and subtests (using ASVAB Form 14) for prediction of success in civilian occupations was addressed by Armstrong, Chalupsky, McLaughlin, and Dalldorf (1988). Their study started as a predictive validation effort, investigating the predictive validity of ASVAB Form 14 for a sample of job incumbents in 12 entry-level civilian occupations. The subtests and composites were to be validated against carefully developed Behaviorally Anchored Rating Scales, BARS, (see Smith & Kendall, 1963, for a discussion of BARS). The incumbents were to be rated by their supervisors using the BARS. The ASVAB was administered and the BARS developed; however, because the Office of Management and Budget would not approve the use of the rating scales, a substitute criterion had to be found and the study had to be redesigned.

As a substitute for civilian occupational success criteria, Armstrong et al. (1988) used Army data from an Army effort (called "Project A") to develop job performance criteria. The Project A data consisted of ASVAB validation information for a job performance criterion, Skill Qualification Test scores. Civilian supervisors in the 12 selected civilian occupations for which the authors had validity data rated task analyses of the corresponding Army occupations to obtain estimates of the overlap, in terms of job requirements, between the 12 civilian occupations and the Army occupational specialties. Sufficient overlap was found in 9 of the 12 civilian occupations. Data from this analysis are presented in Table 22 along with the estimated validities for the nine matching civilian occupations (N = 1,328). The validities are not corrected for restriction in range. The validities are of the same magnitude as the validities of the ASVAB for prediction of JPM criteria measures reported previously.

Table 22. Validity from Military Occupations Studied in Project A

Civilian occupation	Corresponding Military Occupation Army (MOS)	Percent of supervisors matching	Project A[a] validity	N
Bookkeeper/ Accounting Clerk	Accounting Specialist (73D)	50	.70	72
Bus Driver	Motor Transport Operator (64C)	100	.59	14,917
Computer Operator	Computer/ Machine Operator (74D)	86	.64	545
Diesel Mechanic	Heavy-Wheel Vehicle Mechanic (63S)	100	.74	941
Firefighter	Firefighter (51M)	100	.72	72
Licensed Practical Nurse	Medical Specialist (91A)	86	.73	392
Line Installer/ Cable Splicer	Wire Systems Installer (36C)	67	.51	2,907
Operating Engineer	Heavy Construction Equipment Operator (62E)	100	.64	233
Word Processing Machine Operator	Administrative Specialist (71L)	55	.64	9,509

Note. From Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 753) (p. 59) by T. R. Armstrong, A. B. Chalupsky, D. H. McLaughlin, and M. R. Dalldorf, 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

[a]Based on the relationship between ASVAB subtests and Skill Qualification Test (SQT) scores (except for Medical Specialist, where the criterion was a composite of hands-on and job knowledge tests). No weighting of subtests nor corrections for restriction of range.

Because these are estimates of validity from a military study, this result is to be expected. Armstrong et al. (1988) suggested that the validities taken together with the high degree of match between civilian and military occupational task analysis show that the ASVAB probably predicts performance on civilian jobs as well as it does for military jobs.

Armstrong et al. (1988) did not stop with the substitution of validities for military job performance for civilian job performance. They also attempted to estimate the validities of the ASVAB for the civilian occupations directly from the available data. This was accomplished with estimated validities using the predicted likelihood that an individual would be in an occupation based on the observed frequencies of individuals' occupational membership in their data set. These validity estimates are presented in Tables 23, 24, and 25. Armstrong et al. (1988) used Clemans' Lambda (Clemans, 1958) to estimate the validity of the ASVAB for the 12 civilian occupations.

Table 23. ASVAB Subtest and Composite Validity Coefficients Estimated by Clemans' Lambda for Best Linear Combinations (Civilian Validation and Youth Cohort)

	Subtests		Composites	
Occupation	Civilian validation	Youth cohort	Civilian validation	Youth cohort
Bookkeeper/Accounting Clerk	.50	.53	.51	.52
Bus Driver	.17	.30	.13	.28
Cosmetologist	.57	.34	.57	.33
Diesel Mechanic	.73	.58	.73	.58
Electronics Assembler	.30	.56	.25	.55
Electronics Technician	.74	.62	.71	.60
Firefighter	.42	.36[a]	.41	.26[a]
Operating Engineer	.56	.62	.52	.57
Line Installer/Cable Splicer	.57	.54[a]	.57	.46[a]
Computer Operator	.35	.36	.35	.38
Licensed Practical Nurse	.53	.50	.54	.44
Word Processing Machine Operator	.67	.53	.66	.50

Note. From Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 758) (p. 56) by T. R. Armstrong, A. B. Chalupsky, D. H. McLaughlin, and M. R. Dalldorf, 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

[a]Based on a sample of fewer than 20.

Table 24. ASVAB Subtest Validity Coefficients[a] Estimated by Clemans' Lambda

Occupation	Subtests[b]								
	GS	AR	NO	CS	AS	MK	MC	EI	VE
Bookkeeper/ Accounting Clerk		.20	.32	.36		.16			.30
Bus Driver	.01				.10			.05	.04
Cosmetologist			.02	.02					
Diesel Mechanic	.04	.10			.68		.48	.49	
Electronics Assembler									
Electronics Technician	.28	.58	.39	.08	.39	.58	.45	.71	.23
Firefighter	.27	.17	.09		.33	.20	.35	.19	.17
Operating Engineer					.42		.10	.15	
Line Installer/ Cabler Splicer	.38	.30	.19	.12	.62	.18	.38	.57	.40
Computer Operator		.05	.21	.21		.03			
Licensed Practical Nurse	.10								.15
Word Processing Machine Operator			.26	.31				.00	

Note. From Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 758) (p. 54) by T. R. Armstrong, A. B. Chalupsky, D. H. McLaughlin, and M. R. Dalldorf, 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

[a]Only results based on positive relations between skills and occupations are presented in this table.

[b]Definitions for abbreviations are found in Table A-2.

The Cleman's Lambda values shown in Table 23 were based on the regression of all ASVAB subtests and composites and represent the best-weighted linear combinations of subtests and composites, for both the 1980 American Youth sample and the sample of men and women in the Armstrong et al. (1988) study. The regressions were done separately for males, females, and the total sample.

The researchers maintained that the results depicted in Tables 23 though 25 showed the ASVAB to be a valid discriminator among job incumbents in this study. Based on these results, they noted that the civilian job-holders most identifiable by the best-weighted linear combinations of ASVAB subtests and composites are the diesel mechanics and electricians. They noted also that bus drivers could not be distinguished from the other 12 occupations in this study.

Table 25. ASVAB Composite Validity Coefficients[a] Estimated by Clemans' Lambda

	Composites[b]						
Occupations	Acad	Verb	Math	Mech	Bus	Elec	Health
Bookkeeper/ Accounting Clerk	.25	.15	.18		.33	.03	.06
Bus Driver	.00	.01		.02			
Cosmetologist							
Diesel Mechanic	.01		.02	.51		.17	.22
Electronics Assembler							
Electronics Technician	.44	.26	.59	.56	.38	.60	.47
Firefighter	.18	.24	.19	.31	.15	.24	.28
Operating Engineer				.14			
Line Installer/ Cable Splicer	.35	.41	.23	.50	.25	.39	.38
Computer Operator	.02		.04		.11		
Licensed Practical Nurse		.13					
Word Processing Machine Operator				.11			

Note. From Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 758) (p. 54) by T.R. Armstrong, A.B. Chalupsky, D.H. McLaughlin, and M.R. Dalldorf, 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

[a]Only results based on positive relations between skills and occupations are presented in this table.

[b]Acad is the Academic Composite; Verb, the Verbal; Math, the Mathematic; Mech, the Mechanical; Bus, the Business & Clerical; Elec, the Electronics and Electrical; and Health, Health, Social, and Technology.

In an effort to address the question of differential validity in the ASVAB, Armstrong et al. (1988) performed a discriminant function analysis of the ASVAB scores for their sample of civilian job incumbents. After controlling for the effects of gender, they found that the first four characteristic roots (eigenvalues) were statistically significant (using Wilk's criterion). The authors then used the multiple discriminant functions to determine the percent of job incumbents that would be correctly classified.

Table 26 presents the percent of correctly classified civilian job incumbents based on the discriminate function developed in the Armstrong et al. (1988) study. The effect of gender on occupational choice made by the job incumbents can be observed in these results. The percentages in the last column are based on ASVAB scores with total gender means subtracted. These results show that removing the effect of gender reduces the prediction of occupational membership considerably, but there is still significant correct classification based on ASVAB scores alone. Again, the ASVAB appears to be the most discriminative within the electronics occupations.

Table 26. Percentages of Correct Classifications Based on the Subtest Discriminant Functions

Occupation	Including gender variation %	Excluding gender variation %
Bookkeeper/Accounting Clerk	37.3	26.6
Bus Driver	8.9	4.4
Cosmetologist	26.1	17.1
Diesel Mechanic	47.9	32.2
Electronics Assembler	28.5	33.9
Electronics Technician	76.9	73.0
Firefighter	25.9	13.4
Operating Engineer	29.5	31.3
Line Installer/Cable Splicer	38.6	34.0
Computer Operator	10.8	22.8
Licensed Practical Nurse	33.3	31.0
Word Processing Machine Operator	37.2	12.4

Note. From Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 758) (p. 53) by T. R. Armstrong, A. B. Chalupsky, D. H. McLaughlin, and M. R. Dalldorf, 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

In a further attempt to understand the relationship of other variables on the variation of ASVAB scores obtained in their study, Armstrong et al. (1988) analyzed the correlations of the high school composite scores with age and job tenure. Table 27 presents these correlations. These patterns of correlations are similar, indicating no clear-cut effect of age and tenure on the ASVAB composite scores.

Results of a high school composite validity study using Army occupational data from Hanser, Arabian, and Martin (1984), as reported in Hunter, Crosson, and Friedman (1985), are presented in Table 28. These validities are for Army recruits in the indicated broad family groups. The composites are classified into factor-based and occupational composites.

Table 27. Correlations of DoD Student Testing Composite Standard Scores with Age and Job Tenure

Composite	Age	Job Tenure
Academic Ability	0.11	0.05
Verbal Ability	0.12	0.06
Math Ability	-0.01	-0.02
Mechanical and Crafts	0.09	0.11
Business and Clerical	-0.04	-0.06
Electronics and Electrical	0.09	0.06
Health, Social, and Technology	0.06	0.05

Note. From Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 758) (p. 40) by T. R. Armstrong, A. B. Chalupsky, D. H. McLaughlin, and M. R. Dalldorf, 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Table 28. Validity Coefficients of the High School Composites for Various Occupational Areas in the Army

Occupational area	Number of jobs	Number of people	Factor					Occupational				
			AA	V	Q	T	S	M&C	B&C	E&E	HS&T	G
Clerical	7	5,385	64	59	62	51	50	55	<u>62</u>	62	62	63
Skilled Technician	5	5,367	52	48	51	47	38	49	50	53	<u>53</u>	53
Surveillance/ Communication	4	3,530	50	49	48	52	32	54	46	53	<u>54</u>	54
Operator/Food	5	7,724	51	49	47	50	34	53	46	52	<u>54</u>	53
Combat	16	13,904	46	45	44	45	32	48	43	48	<u>49</u>	49
Mechanical Maintenance	4	2,463	44	42	43	46	30	<u>48</u>	40	47	48	48
Electronics	7	4,958	43	42	42	44	28	46	40	<u>46</u>	46	47
Field Artillery	4	6,306	40	38	39	42	28	44	37	42	<u>44</u>	43
General Maintenance	3	596	39	35	42	39	29	41	38	42	<u>42</u>	42
Total	55	50,233	48	45	46	46	33	49	45	49	50	50

AA = Academic Ability
V = Verbal
Q = Quantitative
T = Technical
S = Speed

M&C = Mechanical and Crafts
B&C = Business and Clerical
E&E = Electronics and Electrical
HS&T = Health, Social, and Technology
G = General Cognitive Ability

Note. From The Validity of the Armed Services Vocational Aptitude Battery (ASVAB) for civilian and military job performance (p. 118) by J. E. Hunter, J. J. Crosson, and D. H. Friedman, 1985, Washington, DC: Department of Defense. Decimals omitted and largest row values underlined.

Factor-based composites are those composites that were defined in terms of subtest composition, based on the results of factor analytic studies. Occupational composites are those composites that were defined based on the best-regression-weighted subtests for prediction of military training school success. Factor-based composites were included in their analyses, although these composites are not used operationally in the high school testing system. These results are provided for comparison with the results of the Armstrong et al. (1988) study using the validity estimates presented in Table 25. The comparison shows that the obtained validities for military occupations are generally higher than the validity estimates obtained from the Armstrong et al. (1988) civilian validation. Where direct comparisons are possible (based on liked-named occupations such as electronics--electronics technicians; mechanical maintenance--diesel mechanic), the validity estimates provided in the Armstrong et al. (1988) work appear to be lower than the validities obtained in the Hanser et al. (1984) study.

Synthetic Validity

Some studies identified in the literature have explored techniques for estimating the validities of ASVAB composites (Mullins, Earles, & Ree, 1981; Weisen & Seigel, 1977). Weisen and Seigel (1977) employed job analytic data for a group of Navy jobs and predictive validity information on ASVAB Forms 6 and 7 to estimate ASVAB subtest-level scores that were predictive of poor, average, and superior performers. The authors met with moderate success for five of the ASVAB Form 6 and 7 subtests: WK, AR, MC, SI, and EI.

Mullins, Earles, and Ree (1981) employed a different strategy in that they sought not so much to estimate validity or develop "synthetic" validity as to weight predictive validity data for final technical training course data by the difficulty of the technical school. The importance of taking into account the differing entry-level standards was emphasized by the authors. They pointed out that in the usual predictive validity paradigm, the final school grade for a very difficult school (high aptitude entry standard) is treated the same as that for a very low difficulty school (low aptitude entry requirement). Mullins et al. (1981) developed a method of adjusting final school grades according to the difficulty level of the school and placed all school grades on a single continuum. They then recalculated the criterion measure, recomputed the selector aptitude indices, and compared them to the indices computed in the usual manner. Their results indicated that the adjusted criterion measures and aptitude indices weighted by technical school difficulty predicted final school grades better on cross-validation than did the traditionally computed (unweighted by difficulty) aptitude indices. This finding is consistent with the results reported by Lawrence (1988) on the effects of job complexity as moderating the relationship between aptitudes and first-term attrition.

Summary of Criterion-Related Validity Studies

With the exception of their validity for the TTC criterion, the criterion-related validity of the ASVAB subtests and composites is substantial. For final school grade, the corrected values for ASVAB Forms 8, 9, and 10 range from .37 for the A composite to .54 for the E composite. The comparable, but slightly lower values of the same composites and subtests against JPM is reassuring. The AFQT loses some validity relative to the aptitude area composites against the JPM measures, probably as a function of the lower reliabilities of the JPM criterion measures. The validity coefficients of ASVAB Forms 6 and 7 for prediction of JPM measures (see Appendix F) are uncorrected for range restriction, but still indicate a substantial relationship of the subtests and composites to JPM measures. The relationship of the validity of the AFQT to the aptitude area measures against JPM criteria was reversed for ASVAB Forms 6 and 7. The reasons for this are not clear. The lower averaged validity of the A composite for ASVAB Forms 6 and 7 is a finding consistent with validity results using final school grade as a criterion.

The validity of the ASVAB Forms 8, 9, and 10 for the TTC criterion is the lowest of all three training success criteria, but still appreciable. The lower values of the validity coefficients for this criterion are a consistent finding, with similar results indicated for ASVAB Forms 6 and 7 (see Appendix E). These findings may be indicative of artifacts of the training situation and are perhaps best attributed to a particular type of criterion contamination unique to this type of criterion measure.

The validity of the high school composites for use in the DoD Student Testing Program depends on the extent to which the demonstrated validity of the high school composites generalize from military occupations to similar civilian occupations. The evidence reviewed here indicates that the validities may generalize quite well. The one study with results that bear on the validity of the current high school ASVAB Form 14 indicates that a substantial number of the civilian jobs studied (9 of 12 civilian occupations) substantially overlap with military counterparts--based on judged similarity of task analyses of the military occupations by civilian supervisors. Estimated validities from this study were appreciable, but appeared to be underestimates of the actual validities. The obvious shortcoming in the research on ASVAB validity for the student testing program is the lack of a single, high quality, predictive validation study of the ASVAB against an acceptable job performance criterion.

IV. CONTENT VALIDITY STUDIES

The Standards (APA, 1985) define content-related validity evidence in terms of the demonstrated representativeness of the test items on a given measure to the universe or content domain. Implied in this definition of content validity is the assumption that content validity is established by the process by which the test is constructed. This assumption follows closely the requirements of content validity listed by Ebel (1983): (a) the explicit definition of the ability to be measured; (b) clear, explicit definition of the tasks that make up the test; and (c) the rationale for using a particular test-task to measure the defined ability. The establishment of content validity using Ebel's definition is a process of tying test questions to specific abilities. A fundamental tool in this process is the taxonomy of the test. Studies which provided empirical evidence that ASVAB subtest items "belonged" in relevant taxonomic categories were, for the purposes of this review, considered relevant to questions of the content validity of the battery. Factor analytic studies at the item level exemplify this type of study.

The use of subject-matter experts to generate raw test items according to an explicit taxonomy of content areas fulfills part of the requirement for a content-valid test. Other requirements for content validity are satisfied in the manifestation of desired or acceptable psychometric properties such as appropriate difficulty levels, biserial correlations of the test items, and acceptable reliability. Yet other aspects of content validity, such as the demonstration of appropriate correlation (or lack of correlation) with other measures of the same (or differing) aptitude or construct, are usually measures of construct validity.

The development of all ASVABs has been documented in reports which vary in detail. Bayroff and Fuchs (1970) documented the initial development of Form 1, which was based on the judged equivalence for Service-specific aptitude measures; there were no formal taxonomies for the ASVAB Form 1. Vitola and Alley (1968) documented the validity of Air Force composites derived from ASVAB subtests on Form 1.

For ASVAB Forms 5, 6, and 7, as documented in Jensen, Massey, and Valentine (1976), the "plan" for the new ASVABs consisted of the judgment that all cognitive areas represented in the common Services' classification tests used prior to ASVAB Form 1 development were adequately covered by the new forms. Again, there was no formal taxonomy for these batteries. It was not until ASVAB Forms 8, 9, and 10 that a formal, defined taxonomy was established and built around the content areas existing in this series of forms.

The content of the ASVAB Form 8ax (a precursor to the operational ASVAB Form 8a--the anchor battery for the current score-scale), which was administered to a nationally representative sample of American Youth in the Fall of 1980 (DoD, 1982b), was to set the pattern of the ASVAB for a decade. The development of a new score-scale based on a representative sample of 1980 American youth had a direct and lasting effect on the content validity of the ASVAB. Because ASVAB Form 8ax was the ASVAB version used for standardization of the 1980 score-scale (Maier & Sims, 1986), its content and the content of its sister versions (Forms 8b, 9a, 9b, 10a, and 10b) defined the content domain of all subsequent ASVABs and was published after the operational implementation of those forms. The more detailed version of the taxonomy has served as the basis of the development of ASVAB Forms 11, 12, and 13 (Andberg, Stillwell, Prestwood, & Welsh, 1988; Prestwood et al., 1985); Forms 15, 16, and 17 (Ree et al., in press); and the not yet fully developed Forms 18 and 19 (Curran & Palmer, in press) and Forms 20, 21, and 22 (Palmer, Curran, & Haywood , in press).

As part of the normal ASVAB development process, raw items are developed, edited, and tried out on successively broader ability ranges of military recruits until final, operational-length candidate ASVAB forms are calibrated on a sample of recruits. The final stage of development consists of an Initial Operational Test and Evaluation (IOT&E) where the new forms are administered and calibrated on a full ability range, large sample of applicants. At each stage of the developmental process, items are culled or replaced by items in the same taxonomical category as defined by ASVAB Form 8a, with items that are matched to the p-values and item-test biserial correlations of ASVAB Form 8a. The anchor subtests are always administered with experimental or candidate ASVAB items to obtain appropriate item statistics for matching (Andberg et al., 1988).

It is important in establishing content validity that performance--on tests as a whole and on test items specifically--is free from unintended influences of factors irrelevant to the measurement of the intended ability. More explicitly, quantitative evidence of content validity of the ASVABs comes from studies by Bock and his associates using the large representative sample of American youth from the Profile of American Youth Study (DoD, 1982b). A study by Bock and Moore (1984) investigated ethnic, gender, and demographic influences on test performance. Bock and Moore (1984), examining subgroup performance on individual subtests on the ASVAB Form 8ax used in the Profile of American Youth Study, found that scores on the 10 subtests of the ASVAB were significantly correlated with education level, gender, sociocultural group, mother's education level, and region of the country. In understanding these results, one must bear in mind that the purpose of the Bock and Moore study was to examine the pattern of performance differences in a representative sample of American youth. The ASVAB appeared to be measuring those abilities and experience that are the result of specialized, role-typical education and experience resulting from self-selection into educational tracks rather than an artifact of the test itself--especially in the case of sex differences. Bock and Moore (1984) related all the patterns of test score performance to the literature on the influence of background and biological factors on ability test performance. Their findings related to sex and ethnicity will be discussed later in the section on subgroup validity and equity. Although probably equally appropriate to the section on construct validity, some factor analytic results will be discussed in the context of content validity.

Bock and Mislevy (1981) examined the Profile of American Youth data set using Item Response Theory (IRT) methods to investigate possible test bias and the amount of information provided throughout the ability range on each of the subtests. Bock and Mislevy summarized the item characteristics of each of the power subtests and identified one badly flawed item on the Paragraph Comprehension (PC) subtest. The flawed item was "widowed" in the pagination of the test booklet; i.e., left by itself on the back side of the last page of the PC subtest. They also examined the tests for inordinately high guessing on the part of subjects in the Profile data base and concluded:

> Data from responses to the ASVAB are free from major defects such as high levels of guessing or carelessness, inappropriate levels of difficulty, cultural test-question bias and inconsistencies in test administration procedures. They provide a sound basis for the estimation of population attributes such as means, medians, and percentile points in the youth population as a whole and in subgroups defined by age, sex, and race/ethnicity. (p. 51)

These findings provide evidence of the overall content validity of the ASVAB, but the subtest-by-subtest results of the subtest information curves are most interesting. Bock and Mislevy summarized these subtest information results for ASVAB Form 8ax in the nationally representative sample of American youth, ages 16 - 23, as follows:

> Targeting varies from one subtest to another: the shortest test, Paragraph Comprehension, provides precise information at a level about one standard deviation below the mean, but not as much information for subjects very far above the mean. Mathematics Knowledge, on the other hand, is more informative about subjects above the mean than below the mean.
>
> Subtests with relatively high precision for subjects with low abilities include General Science, Word Knowledge, Paragraph Comprehension, Numerical Operations, and Coding Speed. These subtests would be particularly well-suited for initial selection decisions. Subtests with less precision for subjects with low abilities--i.e., "floor" effects--are Arithmetic Reasoning, Auto and Shop Information, Mathematics Knowledge, and Electronics Information.
>
> Subtests with particularly high precision for subjects with high abilities include General Science, Arithmetic Reasoning, Auto and Shop Information, Mathematics Knowledge, Mechanical Comprehension, and Electronics Information. These subtests would be especially well-suited for placement decisions for subjects who have already been selected. A Subtest with little precision for subjects with high abilities--i.e., a 'ceiling' effect--is Paragraph Comprehension. (p. 23)

Bock, Gibbons, and Muraki (1985) reported the results of a full-information item factor analysis of the eight power subtests of the ASVAB using a sample of 1,178 cases from the National Longitudinal Study of Labor Force Participation that were administered the ASVAB as part of the <u>Profile of American Youth Study</u>. The purpose was to describe a method for item factor analysis using marginal maximum likelihood estimation with the EM algorithm. Though the purpose of the study was methodological and the ASVAB is not defined as unidimensional, results of the study indicated that five of the eight power tests on the ASVAB depart significantly from unidimensionality. The results are discussed in terms of scoring for an adaptive test. The important point for the content validity of the ASVAB is that any adaptive test using Item Response Theory that presents a single score for an ability should be unidimensional; that is, it should contain test items that are all drawn from the same universe or dimension. Although the impact of violations of this assumption are more serious for adaptive tests, Bock et al. (1985) nevertheless provided pertinent data for the content validity of the ASVAB.

The Bock et al. (1985) study showed that five ASVAB subtests contained a statistically significant second factor. The Promax factors (the second factors) were intercorrelated with the first factor in the range of .74 (for GS) to .86 (for MK). The five subtests were General Science, Arithmetic Reasoning, Word Knowledge, Auto and Shop Information (which, in earlier versions of the ASVAB was two separate subtests), and Mathematics Knowledge.

Table 29 shows the types of content that define the factors within those five subtests identified as multidimensional in Bock et al. (1985). The results of the Bock et al. (1985) study are consistent with the findings of Bock and Mislevy (1981). The second factor in the five subtests accounted for between 1% and 4% of the common variance.

Table 29. Content of ASVAB Subtests with Two Factors

Subtests[a]	Factor I	Factor II
GS	Physical Sciences	Life Sciences
AR	Arithmetic Reasoning	Business Arithmetic (Calculation of Interest)
WK	Unknown or unspecifiable	
AS	Automotive Information	Shop Information
MK	Formal Algebra	Numerical Calculations and Mathematical Reasoning

Note. The data are from Full information item factor analysis by R. D. Bock, R. Gibbons, and E. Muraki, 1985, Chicago, IL: National Opinion Research Center.
[a]Definitions of abbreviations are found in Table A-2.

Linn, Hastings, Hu, and Ryan (1988) conducted an item-level study of the Differential Item Functioning (DIF) on ASVAB Form 14 for over 40,000 high school students. This item–level DIF study is mentioned here under content validity because the measurement of extraneous variables, such as race or gender, is one aspect of content validity. The results of the Linn et al. (1988) study indicated that AR and MK had few items with significant DIF related to ethnicity or gender. The Linn et al. (1988) results suggested that specialized vocabulary may play a role in gender-related differential item functioning, especially in Mechanical Comprehension and Electronics Information subtests, but the authors pointed out that such studies may not be of much practical use to test-builders. The results of the

Linn et al. (1988) study were complementary to the findings of Bock and Moore (1984) and Bock et al. (1985). The important finding for the content validity of the ASVAB was that the occurrence of DIF in the battery was minimal, and thus the confounding attributable to extraneous variables, such as gender- and ethnic-related differences, was also minimal or nonexistent. Other evidence as to whether or not the ASVAB subtests measure variables unrelated to the specified abilities comes from research on the ASVAB Form 14 used in the DoD Student Testing Program.

Table 30 displays the correlations of ASVAB Form 14 with age and job tenure from the civilian validation study by Armstrong et al. (1988). The relationships depicted in Table 30 indicate there is not a particularly strong relationship between any of the subtests and age or time on the job. In this study, 72% of the sample of 1,315 were under age 35, and individuals between the ages of 18 and 29 constituted 56% of the sample.

Table 30. Correlations of ASVAB Subtest Standard Scores with Age and Job Tenure

Subtest	Age	Job Tenure
General Science	0.13	0.08
Arithmetic Reasoning	0.06	0.03
Word Knowledge	0.20	0.10
Paragraph Comprehension	0.01	-0.02
Numerical Operations	-0.18	-0.17
Coding Speed	-0.16	-0.13
Auto and Shop Information	0.11	0.12
Math Knowledge	-0.07	-0.07
Mechanical Comprehension	-0.03	0.04
Electronics Information	0.16	0.16
Verbal	0.15	0.07

Note. From Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 758) (p. 40) by T. R. Armstrong, A. B. Chalupsky, D. H. McLaughlin, and M. R. Dalldorf, 1988, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

V. CONSTRUCT VALIDITY STUDIES

As defined in the Standards (APA, 1985), any evidence of validity can be taken as a form of construct validity. Direct indications of construct validity are usually derived from factor analytic studies, studies comparing the test performance of examinees on two batteries that purportedly measure the same thing, and studies that factor analyze intercorrelations among the subtests of two test batteries. Similarity of factors (i.e., like-named subtests loading on the same factors across batteries) is frequently presented as evidence of the construct validity of the subtests comprising the batteries.

Numerous studies have included factor analyses of the ASVAB. Strict comparison of results among studies should be made with caution due to the fact that different researchers used differing types of factor extraction and factor rotation (oblique or correlated, and orthogonal or uncorrelated factors). Comparisons of factor analytic results between first-generation ASVABs (Forms 1-3), second-generation ASVABs (Forms 5, 6, and 7) and third-generation ASVABs (Forms 8 - 14) are provided in Andberg et al. (1988); DoD (1984b); Fletcher and Ree (1976); Maier and Grafton (1981); Ree et al. (1982); and Sims and Hiatt (1983). These studies allow for the comparison of factor solutions among the generations of ASVAB.

Results of each of these and other studies can be summarized in terms of the number of factors found, the type of solution, the subtests composing the obtained factor solutions, the variance accounted for by the factor solutions, and the criteria used for factor loadings.

Table 31 presents summary results for the factor analyses of ASVAB Forms 2 and 5 by Fletcher and Ree (1976). The Sims and Hiatt results (1983) of a factor analyses of ASVAB Forms 6, 7, and 8 are shown in Table 32. These results are illustrative of the typical factor solutions found in other studies (Diehl, 1981; Fischl, Ross, & McBride, 1979; Kass, Mitchell, Grafton, & Wing, 1982; Moreno, Wetzel, McBride, & Weiss, 1983; Sims & Mifflin, 1978; Stoloff, 1983).

A four-factor solution for ASVABs in the third-generation ASVAB Forms 8 through 14, and a five-factor solution for second-generation ASVAB Forms 5, 6, and 7 (see Table 32), are the most interpretable factor solutions found across studies using the most common types of factor extraction methods and rotation. These analyses involved principal factors analysis with squared multiple correlations in the diagonals and subtest intercorrelations in the off-diagonals. The solution was then rotated to a Varimax criterion for an orthogonal solution.

Table 31. Factor Analysis for ASVAB Forms 2 and 5

Factor 1 Technical Information Forms			Factor 2 Scholastic Information Forms			Factor 3 Attention to Explicit Rules Forms			Factor 4 Spatial Perception Forms		
	2	5		2	5		2	5		2	5
TK[a]	.818		WK	.763	.793	AD		.826	SP	.791	.721
SI	.785	.735	GS		.681	NO		.757	MC		.540
AI	.762	.768	MK		.664	CS	.743		AR	.482	.416
EI	.712	.647	AR	.620	.614				MK		.435
MC	.530	.629	EI		.491						
GI		.530	GI		.485						
GS		.447	MC		.433						

Note. From Armed Services Vocational Aptitude Battery (ASVAB) correlational analysis, ASVAB Form 2 versus ASVAB Form 5 (AFHRL TR-76-70, AD-AO32 593) (p. 17) by J. Fletcher and M. J. Ree, 1976, Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.
[a]Definitions of abbreviations are found in Table 1.

Table 32 results show that a four-factor solution is repeatedly found in factor analytic research of the ASVAB. In the analyses of ASVAB Form 8, the deletion of Space Perception from the third generation of ASVABs accounts for the lack of a fifth factor that is typically found in factor analyses of ASVAB Forms 5, 6, and 7. Despite the fifth factor related to spatial ability, the repeated finding of four, common, correlated, interpretable factors across ASVAB generations provides some support for the assumption that validity results can be generalized among ASVAB forms and generations (DoD, 1984b). These four correlated factors are generally interpreted in the literature as Verbal, Speed, Technical, and Quantitative factors.

Table 32. Factor Pattern Matrix from Joint Factor Analysis[a] of ASVAB Forms 6, 7 and 8

	Rotated Factor Loading[b]									
	Factor 1 (verbal)		Factor 2 (speed)		Factor 3 (technical)		Factor 4 (quantitative)		Factor 5 (space)	
Sub-test[c]	6/7	8	6/7	8	6/7	8	6/7	8	6/7	8
GS	.52	.65	---	---	---	---	-.30	---	---	---
WK	.90	.97	---	---	---	---	---	---	---	---
PC	---	.68	---	---	---	---	---	---	---	---
CS	---	---	---	.55	---	---	---	---	---	---
NO	---	---	.72	.70	---	---	---	---	---	---
AR	---	---	---	---	---	---	-.64	-.69	---	---
MK	---	---	---	---	---	---	-.88	-.90	---	---
MC	---	---	---	---	.38	.42	-.33	---	.32	.29
EI	.35	.42	---	---	.46	.41	---	---	---	---
AI	---	---	---	---	.95	.84[d]	--	---	---	---
SI	---	---	---	---	.65	---	---	---	---	---
GI	.56	---	---	---	---	---	---	---	---	---
AD	---	---	-53	---	---	---	---	---	---	---
SP	---	---	---	---	---	---	---	47	---	---

Note. From A joint factor analysis of ASVAB Forms 5/6/7 and Forms 8/9/10 (CNA-83-3178/10) (p. 5) by W.H. Sims and C.M. Hiatt, 1983, Alexandria, VA: Center for Naval Analyses.

[a]2,025 cases who tested on both ASVAB 6/7 and ASVAB 8. The subtest correlation matrix was factor analyzed using a principal factor solution followed by oblique rotation.

[b] Values less than .30 were omitted.

[c] Definitions of abbreviations are found in Tables 1 and 2.

[d] In ASVAB 6/7, the AI and SI subtests were scored separately. In ASVAB 8/9/10, they were combined to form the AS.

Using ASVAB Forms 8, 9, and 10, Ree et al. (1982) extracted varying numbers of factors with orthogonal (to Varimax criterion) and oblique rotations (to Kaiser-Harris Type 2 criterion). The subtest composition of the four-factor solution, with oblique rotation from Ree et al. (1982), is presented in Table 33.

Table 33. ASVAB Form 8a Common Factors (Oblique Solutions)

	Factor Loadings							
	Verbal		Speed		Quantitative		Technical	
Subtest[a]	I[b]	I[c]	II[b]	II[c]	III[b]	III[c]	IV[b]	IV[c]
GS	.54	.54	-.04	-.05	.26	.21	.27	.29
AR	.21	.07	.14	.11	.59	.69	.15	.15
WK	.70	.95	.08	.03	.13	-.02	.16	.01
PC	.62	.68	.17	.16	.15	.08	.12	-.04
NO	.13	-.03	.57	.79	.19	.12	-.08	.03
CS	.07	.06	.56	.81	.10	-.06	.20	.00
AS	.23	.00	.01	.05	.04	-.10	.68	.94
MK	.10	.08	.17	.06	.62	.85	.12	-.05
MC	.13	-.03	.00	.03	.29	.27	.58	.68
EI	.33	.28	.02	-.04	.14	.10	.56	.62

[a]Definitions of abbreviations are found in Table A-2.
[b]Ree, Mullins, Mathews, and Massey (1982), n = 19,359 Service applicants 1980.
[c]Stoloff (1983), n = 9,173 18-23 yr old (DoD reference population).

Table 34 presents the factor intercorrelations from Ree et al. (1982) and Stoloff (1983). Values presented in the first column are from Ree (1982) and those in the second (same Roman numeral as the corresponding factor from Ree et al.) are from Stoloff (1983). The relatively high loadings for GS, WK, and PC are similar in both studies. These subtests, along with EI, formed what both authors called a Verbal factor. A quantitative factor was made up of AR and MK in both, as well as a Technical factor comprised of AS, MC, and EI. The speeded tests, NO and CS, comprised the Speed factor.

Table 34. Factor Intercorrelation Matrix for ASVAB Form 8a from Two Studies[a]

Factors	Verbal	Quantitative	Technical	Speed
Verbal	1.00	.60	.54	.31
Quantitative	.72	1.00	.51	.25
Technical	.62	.58	1.00	.45
Speed	.68	.65	.31	1.00

[a]Below diagonal from Stoloff (1983); above diagonal from Ree, Mullins, Mathews, and Massey (1982).

The factor loadings, as well as the subtests defining the four factors, are very similar in the two studies. The Ree et al. (1982) study used a sample of approximately 20,000 military applicants, whereas the Stoloff sample came from the DoD reference population of 9,173 18- to 23-year-old males and females (DoD, 1982b).

Stoloff (1983) used squared multiple correlations in the principal diagonal of the ASVAB subtest intercorrelation matrix as initial estimates of the commonality and then factored the intercorrelation matrix using Principal Components. Stoloff decided that extraction of four factors best reproduced the original intercorrelation matrix. These four extracted factors were then rotated using oblique Oblimin rotation. The four factors accounted for 87% of the total variance in the ASVAB, and were rotated to a factor pattern with loadings as represented in Table 34.

Aside from the similarity of the factor loadings, the factor intercorrelations were universally lower in the Ree et al. (1982) analysis, with the exception of the intercorrelation between the Speed and Technical factors . The four-factor solution accounted for 87% of the common variance in the Stoloff (1983) study and about 74% of the common variance in the Ree et al. (1982) study. It is important for the reader to bear in mind the nature of the restriction in the range of abilities of the samples in specific studies. The effect of the selector composites in the military selection and classification system is to reduce the amount of observed variance in the intercorrelations, thus changing the factor analytic results. The Ree et al. (1982) results are in close agreement with the results obtained by Kass et al. (1982) on a sample of over 98,000 Army applicants, and those obtained by Moreno et al. (1983) on a sample of 356 male Marine Corps recruits.

The finding of a four-factor solution was repeated in the developmental work for ASVAB Forms 11, 12, and 13 by Andberg et al. (1988) on a sample of approximately 120,000 applicants for the Military Services. The orthogonal Varimax rotation after the extraction of four principal factors accounted for over 96% of the common variance among the subtests for ASVAB Forms 11, 12, and 13. The interpretation of the factors was similar to previous findings for the third-generation ASVABs; that is, Verbal, Quantitative, Speed, and Technical factors were found.

Stoloff (1983) best summarized the results of the factor analytic work on the ASVABs by noting that high positive intercorrelations of the ASVAB factors suggested that a single underlying ability, or 'g,' was measured by the subtests. Other factor analytic work reported by Hunter (1983, 1984) and Hunter et al. (1985) replicated and extended the notion advanced by Stoloff that each ASVAB subtest measures a significant amount of general cognitive ability (GCA or psychometric 'g').

Hunter's work is now discussed separately because it involved replication and re-analysis of much of the previous conventional factor analytic research on the ASVAB and because it tied that research to factor analytic research on another multiple-aptitude battery from the civilian sector, the General Aptitude Test Battery (GATB).

Hunter et al. (1985) used three analytic techniques to support his conclusion about the ASVAB as a measure of GCA. Hunter et al. (1985) corrected ASVAB subtest validities for unreliability and for restriction in range, but did not apply any corrections for sampling error to the validity data. None of the other usual meta-analytic corrections for other sources of error, such as sampling error, in observed validity correlations were used in this particular study. Instead, Hunter et al. (1985) used confirmatory factor analysis procedures to obtain relatively error-free estimates of specific abilities, second-order abilities, and GCA. Here Hunter et al. (1985) made a distinction between specific <u>abilities</u> (as measured by ASVAB and GATB subtests) and general second-order <u>aptitudes</u> (which they seemed to treat collectively as measures of General Cognitive Abilities in their discussion). The distinction between 'g', and Hunter et al.'s (1985) GCA and second-order aptitudes becomes blurred in their treatment and discussion of those constructs in that study.

Next, Hunter et al. (1985) used confirmatory factor analysis to verify a hierarchical factor model, then path analysis to examine the "causal linkages" between aptitudes and job performance. Hunter et al.'s (1985) results showed that the causal path to job performance was linked only to GCA, and that there were no direct causal linkages to job performance from specific aptitudes as measured by ASVAB subtests.

The Hunter et al. (1985) results have a bearing on the construct validity of the ASVAB for a number of reasons unrelated to theoretical position on validity generalization. Hunter et al. (1985) replicated the factor analytic results obtained on large military data sets from the Maier and Grafton (1981), Sims and Hiatt (1981), and Kass et al. (1982) studies. These studies spanned two generations of ASVABs (ASVAB Forms 6 and 7, and ASVAB Forms 8, 9, and 10) with differing subtest content. Hunter et al. (1985) concluded that the ASVAB was a better measure of GCA than the GATB because the ASVAB added a 9% increase in validity (from .55 to .60) over that of the GATB.

Hunter's preliminary work on the validity generalization of the ASVAB (Hunter, 1983, 1984) had shown that the ASVAB measures GCA better than most civilian aptitude batteries. However, the important point from Hunter's validity generalization work, from the standpoint of the present review, is that it demonstrated that the ASVAB measured the same ability or abilities as another widely used aptitude battery, the GATB. Hunter et al. (1985) made another point in their study; namely, that there are some occupations for which specific abilities make a contribution to job performance beyond that of GCA. Just how measures of perceptual ability are considered measures of specific ability when they too have typically high GCA saturation is a question left unanswered by Hunter et al. (1985). Measures of spatial aptitude appear to contribute to validity over and above that contribution made by measures of 'g.' The issue of the contribution of specific and general abilities to prediction of job performance has interested psychologists for a number of years, and was discussed previously in conjunction with the high school testing program and the issue of differential validity of the ASVAB.

A recent study by Wothke, Bock, Curran, Fairbank, Augustine, Gillet, and Guerrero (in preparation) examined the relation of the subtests of ASVAB Form 13c (ASVAB 8a, the reference ASVAB) to the Kit of Factor-Referenced Cognitive Tests (KIT). This factor analytic study of the intercorrelations of the 10 ASVAB subtests and 46 cognitive ability tests from the KIT was based on a sample of Air Force basic trainees. A matrix sampling scheme was used to pair every test with every other test. Joint factor analysis of the data indicated three factors accounted for the ASVAB subtest intercorrelation structure, and six factors accounted for the intercorrelations among the KIT tests. Simultaneous analysis of the two batteries showed most of the factor-space of the ASVAB fits within the factor-space of the KIT. There were two exceptions: Associative memory and figural fluency factors of the KIT were not covered by the ASVAB factor-space.

The ASVAB and Measures of Literacy

The relationship of ASVAB composites to measures of literacy or reading ability has enjoyed much empirical exploration over the years. This attention has largely been due to the need of the Services' training communities to gauge the ability of new recruits to comprehend written materials presented in technical manuals, technical orders, and other written forms of instruction required for job-related tasks or job-related instruction.

The DoD concern about military job demands for literacy, and the relation of reading ability measures to ASVAB aptitude measures, led to the realization that the acquisition of job knowledge was heavily dependent on reading skills and that literacy demands varied by type of military occupation (Burkett, 1977; Caylor, Sticht, Fox, & Ford, 1973).

In the process of establishing a linkage between job and literacy demands, other variables were shown to be important for the predicting of both training success and job performance, such as quality of supervision, types of experience on the job, etc. Still, the focus for selection and classification research and validation became the assessment of reading ability of recruits. Mathews, Valentine, and Sellman (1978) administered the Literacy Assessment Battery (LAB) and other reading tests to over 4,500 applicants for the Armed Services in order to determine the reading skills of applicants and the relationship of the ASVAB General (G) composite to the LAB and to other commercial reading assessment batteries like the Gates-MacGinitie and the Nelson-Denny. Sticht, Hooke, and Caylor (1982) examined the LAB in the context of a "special selection test" for use in screening applicants for Service; however, the strong relationship of the ASVAB Forms 6 and 7 'G' or General Technical composite and the AFQT to reading grade levels as measured by the Gates-MacGinitie ($r = .74$, uncorrected) and the Nelson-Denny (.65, uncorrected) precluded its use as an additional selection test. The cost of additional testing time at applicant processing centers could not be justified in terms of the small increments to the already substantial validity of the ASVAB.

The necessity of predicting and reporting reading ability from selection and classification aptitude measures has a long history in military research. Madden and Tupes (1966) related a reading ability scale to the Airman Qualifying Examination. Mathews et al. (1978) examined ASVAB Forms 6 and 7 aptitude indices in terms of predictive scores on other widely used reading tests. Also, the Air Force developed the Air Force Reading Abilities Test (AFRAT) and related that to ASVAB aptitude composites, the Nelson-Denny, the Test of Adult Basic Education (TABE), and the Gates-MacGinitie (Mathews & Roach, 1983).

A comprehensive study by Waters, Barnes, Foley, Steinhaus, and Brown (1988) examined the relation of six nationally used reading tests to various subtests of the ASVAB. The goal was to calibrate one ASVAB composite (from ASVAB Forms 11, 12, and 13) to each reading test in order to ultimately develop a single Reading Grade Level (RGL) score-scale for use by DoD in reporting the reading abilities of new recruits. The meaning of the resultant RGL score-scale is unknown because calibrating or equating two tests which measure different constructs does not preserve the meaning of either test.

In the Waters et al. (1988) study, over 20,000 applicants were administered either the Gates-MacGinitie, the Nelson-Denny, the Adult Basic Education Examination (ABLE), the Air Force Reading Abilities Test (AFRAT), the Test of Adult Basic Education (TABE), or the Stanford Test of Academic Skills (TASK). The results of factor analyses of these six tests of reading ability and the ASVAB subtests are shown in Table 35. The results indicated that the VE composite should be the anchor composite for the RGL score-scale that represented an equated average of five of the RGL score-scales from the study sample. The study authors decided to omit the Nelson-Denny from the average because descriptive results of the score distributions were too deviant from those of the other five tests to warrant its inclusion in the final analysis.

The results of the factor analyses indicated in Table 35 replicate the usual finding of four factors in the ASVAB with similar subtest composition of the factors. Whether this factor structure was independently arrived at as the most interpretable by the authors, or whether it was "imposed" on the results because the authors extracted the four principal factors found in past research, is not clear from the description of the study. In any event, the composition of the Verbal factor was WK, PC, and GS. Table 35 also shows the corresponding reading tests' loadings with the ASVAB subtests' loadings on the four factors. The extraction of the four factors were somehow arrived at after what appears to be six separate Principal Components analyses of the six instruments. The GS subtests appear on the Verbal factor for four of the reading tests shown in Table 35, yet Waters et al. (1988) decided not to include the GS subtest in the equating because it did not measure the same abilities as did WK and PC. From Table 35, which displays all the loadings above .30, it is apparent that all six instruments have a strong Verbal component or factor, and that the ASVAB subtests WK and PC are loaded heavily on the common Verbal factor. The two exceptions were the Nelson-Denny and the TASK, where PC did not load on the Verbal factor.

Table 35. Factor Analyses of Reading Total Scores and ASVAB Subtest Scores[a]

Test	Eigen-value	Subtest[b]		(Loading)		Factor
ABLE	6.22	WK (.81)	ABLE (.47)	GS (.39)	PC (.30)	Verbal
	1.65	CS (.93)	NO (.32)			Speed
	0.72	AS (.91)	EI (.38)	MC (.36)		Technical
	0.60	MK (.85)	AR (.36)			Quantitative
AFRAT	6.25	WK (.86)	AFRAT(.68)	GS (.48)	PC (.39)	Verbal
	1.60	CS (.93)	NO (.32)			Speed
	0.70	AS (.91)	EI (.35)	MC (.33)		Technical
	0.62	MK (.85)	AR (.39)			Quantitative
G-M	6.24	AS (.91)	EI (.36)	MC (.35)		Technical
	1.67	CS (.93)	NO (.33)			Speed
	0.73	WK (.86)	G-M (.78)	GS (.52)	PC (.42)	Verbal
		AR (.31)				
	0.59	MK (.84)	AR (.38)			Quantitative
N-D	6.10	AS (.92)	EI (.36)	MC (.35)		Technical
	1.67	CS (.93)	NO (.32)			Speed
	0.75	WK (.91)	N-D (.35)	GS (.33)		Verbal
	0.60	MK (.84)	AR (.36)			Quantitative
TABE	6.15	WK (.86)	TABE(.71)	GS (.49)	PC (.41)	Verbal
	1.64	CS (.94)	NO (.32)			Speed
	0.76	AS (.91)	EI (.36)	MC (.34)		Technical
	0.61	MK (.85)	AR (.36)			Quantitative
TASK	6.06	AS (.91)	MC (.36)	EI (.35)		Technical
	1.70	CS (.94)	NO (.33)			Speed
	0.75	WK (.76)	TASK(.37)			Verbal
	0.65	MK (.84)	AR (.37)			Quantitative

Note. From Estimating the reading skills of military applicants: Development of an ASVAB to RGL conversion table (HumRRO-FR-PRD-88-22) (p.50) by B. K. Waters, J. D. Barnes, P. P. Foley, S. D. Steinhaus, and D. C. Brown, 1988, Alexandria, VA: Human Resources Research Organization.

[a]Factor loadings >.30.

[b]Definitions for abbreviations are found in Table A-2.

Special Military Selection and Classification Measures

Service-specific needs have periodically resulted in implementation of specific aptitude measures for prediction of training success and job performance.

For example, the Analysis Aptitude Test (Mathews, 1977) showed substantial validity ($r = .58$) for Air Force and Army students in a radio communications course. Still, the ASVAB Forms 6 and 7 subtests of WK, AR, and SP added significant validity (14%) to that of the Analysis Aptitude Test against a criterion of final school grade.

Wilbourn, Guinn, and Leisey (1976) validated non-verbal aptitude measures along with ASVAB-based aptitude composites in an effort to replicate earlier results obtained by Wilbourn and Guinn (1973). The earlier study had shown that non-verbal measures added significant increments to validity over that provided by aptitude measures alone. The 1976 study attempted to replicate these findings, using a sample of 13,584 male Air Force recruits, and found that non-verbal aptitude measures added larger increments in validity for lower ability airmen than for higher ability airmen. These results hint at the notion that validity for a set of measures may not be constant throughout the ability range, a concept explored later in a study by Lee and Foley (1986). The Lee and Foley (1986) study was primarily concerned with the effect of the predictor means on obtained validity coefficients when the correlations are corrected for range restriction. However, the concept that validity of aptitude measures may be different in different ability ranges is one worth pursuing in future research efforts.

Another special selection screening test is the English Diagnostic Test (EDT), which was developed in an effort to better select candidates for joint-Service journalism-related courses. Results of two studies--Park, Mathews, and Ree (1985) and Booth-Kewley (1984a)--indicated that the ASVAB General composite had higher predictive validity than the EDT against final school grade, with the $R^2 = .634$ (corrected for restriction in range) for the General composite. The EDT added only .006 to the R^2. These results are consistent with those of Mathews et al. (1978), in which the ASVAB General composite was found to be a good predictor of reading ability.

The ASVAB aptitude composites have a good record for prediction of training success when compared to other measures, as indicated in the previous research. However, Stoker, Hunter, Batchelor, and Curran (1987) compared five special measures of specific aptitudes (MCAT - Multiplex Controller Aptitude Test; OCT - Object Completion Test; RBT - Rotated Blocks Test; PAT - Perceptual Abilities Test; and the EMT - Electrical Maze Test) with MAGE ASVAB aptitude indices (AIs) for prediction of training success in the Air Traffic Controller career field. The results indicated that two of the five measures (MCAT and RBT) explored in the experimental validity study were better predictors of training success than the ASVAB AIs, and added significant predictive validity over that of the ASVAB AIs alone. The significant increment in R^2 using the MCAT and RBT (.068 against a criterion of pass/fail in training) indicates that there was some job-specific criterion variance unpredicted by ASVAB composites alone.

The study by Stoker et al. (1987) also examined Air Traffic Controller post-training attrition (within the first year) in relation to both the ASVAB aptitude indices and the five experimental measures. First-year attrition was primarily due to the success or failure on the Federal Aviation Agency (FAA) licensing examination, a fairly objective index of job performance. Results indicated that scores on the ASVAB AIs were not significantly related to post-training success. However, scores on the PAT were significantly different for the success and failure groups. The authors suggested that the abilities required for success in training may be different from those required for post-training success.

Because the Air Traffic Controller career field is one requiring perceptual abilities, the above results may simply point to a need to predict that portion of the criterion space with measures of perceptual ability. This gap in specific abilities measured by the current generation of the ASVAB is consistent with the observations of Hunter et al. (1985) and Schmidt et al. (1987).

The finding that prediction of training success may differ from prediction of later first-term success was also supported by the results of Hawley et al. (1977) using specially developed job performance tests for Air Force jet engine mechanics. ASVAB general knowledge aptitude subtests (such as Arithmetic Reasoning and Word Knowledge) were more highly related to training success than to performance criterion tests administered after entry-level training; and information subtests of the ASVAB were more related to the performance criterion tests for experienced mechanics than were other ASVAB aptitude indices. The study results also indicated that only one of the job performance tests developed in Hawley et al. (1977) was significantly related to training success. These results indicated differences between the acquired abilities related to performance and the abilities required for success in training. The ASVAB aptitude measures seemed to lose some predictive validity as the criteria became more job-performance-oriented and the further removed the performance measures were from the initial training.

The ASVAB has also been validated with an interest inventory, the Vocational Interest-Career Examination (VOICE), developed by the Air Force to predict job satisfaction (Alley, Wilbourn, & Berberich, 1976). The results of that study indicated that the ASVAB aptitude indices added no predictive validity to the interest inventory for the prediction of job satisfaction, not a surprising finding in that aptitude measures are not theoretically related to job satisfaction in any strong manner.

Comparison Studies of the ASVAB with Civilian Multiple-Aptitude Batteries

Different forms of the ASVAB have been systematically compared to other civilian, commercially available batteries in an effort to establish a nomological net for ability constructs. These relationships to other multiple-aptitude batteries have been summarized elsewhere (DoD, 1984a, 1984b; Hunter et al., 1985; McGrevy, Knouse, & Thompson, 1974). The following discussion summarizes the correlations found among ASVAB subtests and composites with the GATB, the Differential Aptitude Test (DAT), the Flanagan Industrial Test (FIT), the Flanagan Aptitude Classification Test (FACT), the California Achievement Test (CAT), and the Wechsler Adult Intelligence Scale (WAIS). The correlational data between the ASVAB and other multiple-aptitude batteries have not been corrected for restriction in range for any of the subtests or composites of the batteries. The lack of positive manifold for some of the comparisons may be an artifact of the restriction in range, which will tend to reduce the variability in measures. Also, none of the measures were corrected for unreliability.

The correlations between subtests in the GATB and ASVAB Form 5 from a study by Kettner (1976) are presented in Table 36. The correlations indicate that the GATB and the ASVAB subtests measure similar cognitive abilities. The correlations between subtests with similar content are all moderate to high, from about .59 between ASVAB Form 5 Mechanical Comprehension and GATB Dimensional Space to .67 between Dimensional Space and ASVAB Space Perception. Other high correlations are found between ASVAB

Arithmetic Reasoning and GATB Arithmetic Reasoning (.74) and ASVAB Form 5 Word Knowledge and GATB Vocabulary (.73). There is also a moderate correlation of the ASVAB General Science with GATB Vocabulary (.60).

The correlations in Tables 37 and 38 between subtests and composites of the DAT, respectively, and ASVAB Form 14 show strong relationships to subtests and composites claiming to measure the same cognitive abilities. For example, ASVAB Mechanical Comprehension and DAT Mechanical Reasoning correlate .73; ASVAB Arithmetic Reasoning and DAT Abstract Reasoning correlate .65; DAT Numerical Ability and ASVAB Arithmetic Reasoning correlate .79. ASVAB General Science correlates .72 with the DAT measure of Verbal Reasoning. The ASVAB speeded subtests Numerical Operations and Coding Speed correlate moderately (.43) with DAT Clerical Speed and Accuracy and ASVAB Coding Speed. There is a lower correlation (.26) for the ASVAB, Business and Clerical, and Civilian Occupational composite against DAT Clerical Speed and Accuracy.

Tables 39 and 40 show generally lower overall correlations between the subtest and composites of the FIT/FACT and the ASVAB than was the case with the GATB or the DAT. These low correlations may be due in part to the low variances of the FIT/FACT in the sample on which the correlations are based (Friedman et al., 1986).

Tables 41 and 42 indicate the subtest and composite correlations between the CAT and the ASVAB Form 14. As with the DAT and the GATB, the relations between the CAT and the ASVAB subtests and composites are somewhat predictable. The CAT Total Reading correlates .86 with the ASVAB Verbal. The lowest AFQT correlation with a total CAT score is .76 (CAT-Total Math). The lowest AFQT correlation with a CAT subtest (Spelling) is .58. The CAT Total Math correlates .86 with the ASVAB Math composite. ASVAB General Science correlates .70 with the CAT Reading Vocabulary and the Reading Comprehension subtests, and .73 with Total Reading.

The AFQT based on the ASVAB Form 3 was administered with the Wechsler Adult Intelligence Scales (Full-Scale IQ, Performance IQ and Verbal IQ) in a study by McGrevy et al., (1974). That study also examined the relationship of these tests to the Airman Qualifying Examination (AQE Form J, a predecessor of the first ASVAB used by the Air Force for selection and classification) for a sample of 100 Blacks and Whites.

Table 36. Average Correlations[a] Between ASVAB Form 5 and GATB Subtests

ASVAB Subtests	GATB Subtests							
	NC	CO	DS	VO	TM	AR	FM	MM
GI	.16	.19	.19	.34	.12	.26	.12	.01
NO	.56	.68	.37	.42	.50	.53	.42	.40
AD	.34	.31	.30	.15	.44	.19	.31	.40
WK	.31	.44	.35	.73	.20	.61	.22	.04
AR	.40	.44	.51	.59	.28	.74	.41	.08
SP	.30	.40	.67	.37	.29	.48	.44	.12
MK	.44	.66	.49	.64	.36	.73	.45	.14
EI	.16	.51	.50	.42	.20	.40	.24	.01
MC	.20	.32	.59	41	.17	.50	.35	-.04
GS	.23	.33	.45	.60	.19	.49	.21	.02
GB	.20	.31	.37	.54	.19	.43	.19	-.01
SI	.23	.26	.47	.36	.29	.40	.34	.08
AI	.09	.16	.42	.23	.13	.21	.25	-.02

ASVAB-5 Sul .ests

General Information (GI)
Numerical Operations (NO)
Attention to Detail (AD)
Word Knowledge (WK)
Arithmetic Reasoning (AR)
Space Perception (SP)
Math Knowledge (MK)
Electronics Information (EI)
Mechanical Comprehension (MC)
General Science (GS)
General Biological Science (GB)
Shop Information (SI)
Automotive Information (AI)

GATB Subtests

Name Comparison (NC)
Computation (CO)
Dimensional Space (DS)
Vocabulary (VO)
Tool Matching (TM)
Arithmetic Reasoning (AR)
Form Matching (FM)
Mark Making (MM)

Note. From Technical supplement to the Counselor's Manual for ASVAB Form 14 (p. 39) by Department of Defense, 1984b, North Chicago, IL: Military Entrance Processing Command.

[a]Coefficients reported are the average of 11th and 12th grade male and female students. The total sample was 616.

Table 37. ASVAB Form 14 and DAT Correlation Coefficients[a] for ASVAB Subtests

DAT Subtests	ASVAB Subtests[b] GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
Verbal Reasoning-VR	.72	.75	.78	.72	.23	.22	.47	.73	.67	.48
Numerical Ability-NA	.64	.79	.67	.66	.41	.35	.40	.78	.57	.42
Abstract Reasoning-AR	.58	.65	.62	.62	.30	.28	.39	.66	.57	.40
Clerical Speed and Accuracy-CSA	.03	.10	.04	.07	.35	.43	-.03	.13	.03	-.01
Mechanical Reasoning-MR	.66	.62	.63	.60	.20	.12	.63	.58	.73	.59
Space Relations-SR	.61	.66	.59	.59	.16	.19	.49	.67	.66	.50
Spelling-S	.53	.54	.60	.57	.32	.36	.27	.54	.39	.35
Language Usage-LU	.68	.67	.76	.72	.20	.26	.39	.67	.55	.48
VR + NA	.73	.82	.78	.74	.33	.30	.47	.80	.63	.49

Note. From Technical supplement to the Counselor's Manual for the ASVAB Form 14 (p. 36) by Department of Defense, 1984b, North Chicago, IL: Military Entrance Processing Command.

[a]Based on 1,338 students.

[b]Definition of abbreviations are found in Table A-2.

Table 38. ASVAB Form 14 and DAT Correlation Coefficients[a] for ASVAB Composites

DAT Subtests	ASVAB Composites[b] AFQT	AA	VBL	MTH	MC	BC	EE	HST
Verbal Reasoning-VR	.78	.82	.80	.78	.69	.75	.79	.81
Numerical Ability-NA	.80	.79	.71	.82	.65	.78	.77	.77
Abstract Reasoning-AR	.69	.69	.66	.69	.60	.68	.67	.70
Clerical Speed and Accuracy-CSA	.16	.08	.06	.12	.03	.26	.08	.07
Mechanical Reasoning-MR	.65	.68	.69	.63	.76	.58	.72	.75
Space Relations-SR	.66	.68	.65	.70	.69	.63	.72	.73
Spelling-S	.64	.62	.61	.57	.46	.65	.58	.58
Language Usage-LU	.76	.78	.78	.71	.62	.73	.74	.75
VR + NA	.84	.86	.81	.85	.72	.81	.84	.84

Note. From Technical supplement to the Counselor's Manual for the ASVAB Form 14 (p. 36) by Department of Defense, 1984b, North Chicago, IL: Military Entrance Processing Command.

[a]Based on 1,338 students.

[b]Definition of abbreviations are found in Table 5.

Table 39. ASVAB Form 14 and FIT/FACT Correlation Coefficients[a] for ASVAB Subtests

FIT/FACT Subtests	ASVAB Subtests[b]									
	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
Arithmetic	.35	.48	.35	.37	.49	.52	.07	.51	.26	.23
Electronics	.38	.36	.32	.30	.08	.04	.43	.30	.43	.49
Expression	.41	.34	.46	.40	.31	.36	.01	.43	.16	.16
Judgment and Comprehension	.46	.46	.51	.50	.21	.25	.08	.48	.30	.29
Reasoning	.52	.67	.55	.58	.30	.29	.17	.70	.44	.35
Mechanics	.40	.28	.31	.21	.03	-.06	.65	.18	.51	.50
Scales	.41	.52	.43	.40	.34	.41	.26	.49	.41	.35
Tables	.30	.44	.36	.42	.47	.55	-.01	.45	.20	.18
Vocabulary	.48	.45	.56	.48	.12	.15	.21	.47	.35	.33
Coding	.30	.33	.31	.41	.36	.38	-.15	.41	.17	.16

Note. From Technical supplement to the Counselor's Manual for the ASVAB Form 14 (p. 38) by Department of Defense, 1984b, North Chicago, IL: Military Entrance Processing Command.

[a]Based on 1,029 students.

[b]Defintions of abbreviations are found in Table A-2.

Table 40. ASVAB Form 14 and FIT/FACT Correlation Coefficients[a] for ASVAB Composites

FIT/FACT Subtests	ASVAB Composites[b]							
	AFQT	AA	VBL	MTH	MC	BC	EE	HST
Arithmetic	.53	.47	.40	.52	.32	.59	.46	.44
Electronics	.34	.38	.37	.35	.53	.27	.46	.44
Expression	.47	.44	.47	.41	.21	.52	.40	.37
Judgment and Comprehension	.53	.54	.55	.50	.34	.51	.50	.50
Reasoning	.66	.69	.62	.73	.50	.64	.67	.66
Mechanics	.25	.31	.34	.25	.60	.16	.42	.42
Scales	.54	.53	.46	.54	.47	.56	.53	.53
Tables	.53	.46	.41	.47	.25	.59	.41	.41
Vocabulary	.52	.55	.57	.49	.41	.47	.52	.53
Coding	.44	.38	.39	.40	.15	.48	.36	.34

Note. From Technical supplement to the Counselor's Manual for the ASVAB Form 14 (p. 38) by Department of Defense, 1984b, North Chicago, IL: Military Entrance Processing Command.

[a]Based on 1,029 students.

[b]Definitions of abbreviations are found in Table 5.

Table 41. ASVAB Form 14 and CAT Correlation Coefficients[a] for ASVAB Subtests

CAT Subtests	ASVAB Subtests[b]									
	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
Reading Vocabulary-RV	.70	.65	.81	.70	.39	.31	.27	.65	.48	.39
Reading Comprehension-RC	.70	.66	.79	.74	.41	.35	.30	.66	.50	.39
Spelling	.41	.49	.53	.47	.37	.36	.04	.52	.23	.17
Language Mechanics-LM	.47	.57	.59	.55	.42	.39	.07	.62	.33	.22
Language Expression-LE	.60	.64	.72	.68	.42	.37	.19	.66	.41	.32
Mathematics Computation-MC	.51	.70	.55	.57	.48	.39	.15	.75	.39	.28
Mathematics Concepts and Applications-MCA	.62	.80	.67	.67	.49	.40	.26	.83	.52	.37
Reference Skills	.55	.63	.64	.64	.46	.44	.17	.64	.41	.31
Total Reading (RV + RC)	.73	.69	.83	.73	.42	.35	.30	.69	.52	.41
Total Language (LM + LE)	.58	.65	.71	.67	.45	.41	.15	.69	.41	.30
Total Math (MC + MCA)	.59	.79	.66	.65	.51	.42	.22	.83	.48	.35

Note. From Technical supplement to the Counselor's Manual for the ASVAB Form 14 (p. 35) by Department of Defense, 1984b, North Chicago, IL: Military Entrance Processing Command.

[a]Based on 1,681 students.

[b]Definitions of abbreviations are found in Table 2.

Table 42. ASVAB Form 14 and CAT Correlation Coefficients[a] for Composites

	ASVAB Composites[b]							
CAT Subtests	AFQT	AA	VBL	MTH	MC	BC	EE	HST
Reading Vocabulary-RV	.79	.79	.81	.69	.56	.72	.73	.75
Reading Comprehension-RC	.79	.80	.82	.70	.58	.75	.73	.76
Spelling	.58	.56	.52	.53	.29	.59	.49	.48
Language Mechanics-LM	.65	.64	.59	.63	.37	.67	.58	.58
Language Expression-LE	.75	.75	.73	.68	.49	.72	.67	.69
Mathematics Computations-MC	.71	.70	.60	.77	.48	.72	.69	.65
Mathematics Concepts and Application-MCA	.81	.82	.73	.86	.61	.80	.80	.78
Reference Skills	.72	.71	.68	.67	.47	.73	.65	.66
Total Reading (RV + RC)	.83	.84	.86	.73	.60	.77	.77	.79
Total Language (LM + LE)	.80	.75	.73	.71	.47	.75	.69	.69
Total Math (MC + MCA)	.76	.80	.70	.86	.58	.79	.78	.75

Note. From Technical supplement to the Counselor's Manual for the ASVAB Form 14 (p. 35) by Department of Defense, 1984b, North Chicago, IL: Military Entrance Processing Command.

[a]Based on 1,681 students.

[b]Definitions for abbreviations are in Table 5.

Intercorrelations from the McGrevy et al. study are shown in Tables 43 and 44. They are presented in the present review for historical and general interest, and to illustrate the process by which subsequent versions of the ASVAB were systematically related not only to previous Armed Services selection and classification aptitude batteries (c.f., Frankfelt, 1970), but to other well-established individual intelligence tests as well. The stronger relationship of the AFQT to Full-Scale IQ and the extent of the differences between Blacks and Whites display a commonly found pattern (Jensen, 1980).

The work of Hunter et al. (1985) and Hunter's earlier work (Hunter, 1983; 1984) reexamined GATB and ASVAB data in the context of validity generalization. Their conclusion, resulting from analysis of several large data sets from the military using ASVAB data and from the U.S. Employment Service (USES) using GATB, was that although both batteries measure the same general cognitive ability, the ASVAB has more subtests that are 'g'-saturated and consequently is a better measure of 'g' than the GATB. Table 45 presents the results of the Hunter et al. (1985) re-analysis of five large military data sets with estimates of General Cognitive Ability for the GATB and the ASVAB. These average validities represent relatively "pure" measures of what Hunter et al. (1985) called "key aptitudes." Key aptitudes are relatively uncontaminated measures of general cognitive abilities and differ in subtest composition from the regression-based selector composites used by the Services.

Table 43. Intercorrelations of WAIS IQs, AFQT Scores, and AQE Aptitude Indices for 100 Black Air Force Enlistees

Test	1	2	3	4	5	6	7	8
1. WAIS Verbal IQ								
2. WAIS Performance IQ	.46							
3. WAIS Full-Scale IQ	.88	.83						
4. AFQT	.38	.44	.48					
5. AQE-Admin	.27	.19	.27	.26				
6. AQE-Elec	.26	.26	.30	.37	.31			
7. AQE-Gen	.23	.29	.30	.36	.62	.55		
8. AQE-Mech	.10	.23	.19	.30	.29	.40	.63	

$p \leq .05 = .195$ $p \leq .01 = .254$

Note. From Relationships among an individual intelligence test and two Air Force screening and selection tests (AFHRL-TR-74-25, AD-781 033) (p. 8) by D. F. McGrevy, S. B. Knouse, and R. A. Thompson, 1974, Lackland AFB, TX: Manpower Personnel and Training Division, Air Force Human Resources Laboratory.

Table 44. Intercorrelations of WAIS IQs, AFQT Scores and AQE Aptitude Indices for 100 White Air Force Enlistees

Test	1	2	3	4	5	6	7	8
1. WAIS Verbal IQ								
2. WAIS Performance IQ	.54							
3. WAIS Full-Scale IQ	.91	.84						
4. AFQT	.64	.62	.71					
5. AQE-Admin	.60	.28	.52	.51				
6. AQE-Elec	.69	.60	.74	.75	.56			
7. AQE-Gen	.77	.44	.71	.64	.71	.73		
8. AQE-Mech	.45	.50	.53	.67	.34	.69	.56	

$p \leq .05 = .195$ $p \leq .01 = .254$

Note. From Relationships among an individual intelligence test and two Air Force screening and selection tests (AFHRL-TR-74-75, AD-781 033) (p. 8) by D. F. McGrevy, S. B. Knouse, and R. A. Thompson, 1974, Lackland AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Table 45. The Average Validity of Various Composite Scores Estimating General Cognitive Ability from Five Large Military Studies

Authors	Service	Number of Jobs	GATB[a]	QV[b]	QVT[c]
Maier and Fuchs (1972)	Army	103	57	60	63
Maier and Grafton (1981)	Army	35	55	58	61
Maier and Truss (1983)	Marine Corps	33	59	61	61
Sims and Hiatt (1981)	Marine Corps	33	52	56	57
Thorndike (1957)	Air Force	46	53	54	57
Total		250			
Average			55.2	57.8	60.2

Note. From The validity of the Armed Services Vocational Aptitude Battery (ASVAB) for civilian and military job performance (p. 94) by J. E. Hunter, J. J. Crosson, and D. H. Friedman, 1985, Washington, DC: Department of Defense.

[a]GATB = AR + WK.

[b]QV = (AR + MK) + (WK + GS).

[c]QVT = (AR + MK) + (WK + GS) + (MC + EI).

Hunter et al. (1985) also pointed out that the ASVAB contains no measure of psychomotor ability and no measure of perceptual ability. However, Hunter et al. (1985) failed to discuss how these specific abilities and their presence in the GATB relates to the greater validity of the ASVAB (which has none of these measures). Moreover, in the hierarchical factor analysis of ASVAB subtests, Hunter et al.(1985) did not include four of the ten ASVAB subtests, on the grounds these four subtests were not reliable measures of the specific or general abilities which Hunter et al. (1985) deemed important to include in their analyses (i.e., they believed the subtests were contaminated with a number of different types of error including measurement error, sampling error, and the effects of restriction in range).

One of the abilities excluded from the Hunter et al. (1985) analyses was Perceptual Speed, which might be adequately assessed by the Speed factor typically found in factor analyses of the ASVAB (with NO and CS subtests loading on the factor). Hunter et al. (1985) also excluded AS and PC from their hierarchical factor analysis, largely on the grounds that the intercorrelation of AS with the other subtests was not parallel with the other subtests. PC was omitted from the analysis because of its lesser reliability than that of WK, its companion on the Verbal factor.

Importantly, the work of Hunter et al. (1985) has shown that the issue of specific and general abilities in prediction is not settled, and that the ASVAB owes much of its predictive validity to the high General Cognitive Ability (GCA) saturation of the power subtests. Based on the results of their analyses, Hunter et al. (1985) maintained that the usefulness of ability measures in prediction of job performance or training success is due to the measurement or estimation of GCA. Further, they maintained that the ASVAB provides as good or better an estimate of GCA than does the GATB, and that use of the ASVAB as an estimate of GCA results in an increase in general validity of .55 to .60 over that of the GATB.

The validity generalization of the ASVAB and the studies that deal with that topic seem to devolve to a single issue, that of the usefulness of specific and general abilities in prediction. Whether or not there is specific variance (situational specificity) in military or civilian jobs, or whether there is sufficient differential validity in the ASVAB to predict what situational, job-specific variance may exist, remains to be explored.

Validity Generalization Studies

The ASVAB has been the subject of a number of more traditional validity generalization studies beyond the study by Hunter et al. (1985). The ASVAB high school testing program relies heavily on the validity generalization of the ASVAB from military occupations to the same or similar civilian occupations. Much of the validity generalization underpinning the ASVAB's use as a counseling tool hinges on the presence or absence of job or situational specificity (the doctrine that validity of employment tests varies from job to job or situation to situation because the factor structure of job performance is situation-specific). The literature on validity generalization is too broad to cover in toto here, but it should be noted that those validity generalization studies which directly address the situational specificity issue (i.e., Linn, Harnisch, & Dunbar, 1981; Pearlman, Schmidt, & Hunter, 1980; Schmidt, Hunter & Pearlman, 1981; and Schmidt et al., 1988) all point to the absence of situational specificity. That is, all support the hypothesis that the validity of a test or test battery for one job is good for any similar type job, in any other setting.

The problem of situational specificity was addressed directly by Foley (1986) using a predecessor to the ASVAB, the Basic Test Battery (BTB). Foley found that grouping Navy jobs in different ways had no moderating effect on the variability of the aggregated validity coefficients. Foley further maintained, based in part on his results (displayed in Table 46), that the traditional practice of validating each selector aptitude composite against training success in each individual military occupation is unnecessary. The validity of the BTB (and hence, the highly similar ASVAB) generalizes across military occupations.

The doctrine of situational specificity is not to be entirely laid to rest, however. In a 1985 study, Dunbar, Mayekawa, and Novick used simultaneous estimation of regression weights for ASVAB Forms 6 and 7 subtests, which then made up the operational selector composites for Marine Corps training courses, to examine a method of Bayesian simultaneous estimation of multiple regression in m-groups. Dunbar et al. (1985) tested the difference of within-group slopes and intercepts of regression equations developed for groups or clusters of Marine Corps jobs. Their purpose was to determine the extent to which regression equations developed for specialties within an occupational cluster were indeed interchangeable. They examined the differences in predictive accuracy of single equations developed for Marine Corps training courses in the Clerical, Electrical, and Mechanical areas on a sample of 14,649 Marine Corps recruits. The results of their application of simultaneous estimation and comparison of regression weights within occupational clusters revealed that there were few differences among specialties in the Clerical cluster, but some courses in other clusters had significantly different regression equations when compared to other courses within the cluster.

Table 46. BTB Validity Generalization Results Based on Individually Corrected Validities

Composite[a]	Mean r	SD[b]	CV[c](90%)	n(N)
M	.54	.08	.44	245 (199,254)
A	.51	.07	.42	379 (255,490)
G	.58	.08	.48	383 (258,519)
E	.60	.09	.49	196 (164,684)

Note. The data are from Validity generalization of Navy selector composites (NPRDC-TR-86-17) by P. P. Foley, 1986, San Diego, CA: Navy Personnel Research and Development Center.

[a]Abbreviations are defined in Table 3.
[b]Weighted by sample size.
[c]Credibility values.

The mixed results from the Dunbar et al. (1985) study suggest that common regression weights for aptitude variables are appropriate for some, but clearly not all, Marine Corps specialties within an occupational cluster or family. Some specialties within the Mechanical and Electrical job clusters retained different weights, thus indicating some situational specificity involved in the prediction of final course grades. It would appear, based on these findings, that results of specific validity studies should generalize only to jobs that have similar or common regression weights for the prediction of training success.

The Dunbar et al. (1985) results are consistent with those obtained by Alley et al. (1988) using Air Force recruits attending 211 different Air Force technical training schools. The total sample consisted of 154,844 Air Force recruits, with individual course sizes ranging from a minimum of 100 to 15,584, and an average N per course of 734.

Using a similar analytic strategy, though not the same m-group regression method used by Dunbar et al. (1985), Alley et al. (1988) hierarchically clustered Air Force jobs on the basis of the similarity of job-specific regression equations using ASVAB subtests as predictors of final school grade (FSG). Results of this study indicated six clusters of jobs, or job families. Four of the clusters were roughly equivalent to the four basic Air Force selection and classification composites, the now familiar M, A, G, and E groupings. Two other clusters with differing characteristics emerged from the regression equation groupings in the Alley et al. (1988) study. One cluster was characterized by uniformly low regression coefficients across all ASVAB subtests. The other had correspondingly high weights for the subtests. Alley et al. (1988) interpreted this latter cluster as indicating an emerging need for "generalists who demonstrate a relatively broad range of talents across the whole domain of abilities as measured by the ASVAB" (p. i).

Role of General Cognitive Ability

The results of other validity generalization studies on the ASVAB are found in Stermer (1988), who focused on the AFQT and selector composites, and in Jones (1988), who focused on the contribution of general cognitive ability to the criterion-related validity of the ASVAB subtests, using samples of Air Force recruits. The results of the Stermer (1988) study showed that the AFQT had high predictive validity for the Air Force occupations included in the sample. Results of this study also indicated that there was some differential validity for the Mechanical, Electronics, and General Air Force selector composites.

In another study employing Air Force validity data for ASVAB Forms 11, 12, and 13 for prediction of final training school grades, Jones (1988) employed meta-analytic techniques to aggregate validities within Air Force occupational specialties grouped into M, A, G, and E job families or clusters, across a sample of 37 USAF jobs (N = 24,482).

A principal components analysis of the ASVAB subtest intercorrelations in the 1980 sample of American youth was done and the first principal component was computed. Jones (1988) then averaged the validities across all the Air Force occupational specialties used in the study, and rank-ordered the validities. Jones (1988) then correlated the rank-ordering of the validities with the rank-ordering of the ASVAB subtests on the first principal component (taken as an index of 'g' saturation for the subtests in this study). Her results indicated that the rank-orderings were significantly correlated, showing 'g' to be a significant predictor of entry-level training success in Air Force occupations. Results of the Jones (1988) study did not, however, provide an indication as to whether specific abilities added significant validity to prediction of this criterion.

The Stermer (1988) validity generalization study more directly addressed the issue of differential validity in the ASVAB composites. That study investigated the hypothesis that the AFQT, as a measure of GCA, would add significant validity to the prediction of military training success over that of the four Air Force aptitude selector composites (M, A, G & E). Stermer analyzed validity results for a sample of Air Force recruits (N = 29,619), and further hypothesized that the AFQT, as a measure of GCA, would have higher validity for females than for males.

Table 47 indicates the results of the Stermer (1988) study in terms of the proportion of variance in the criterion scores accounted for both by the selector composites and by the AFQT. ASVAB Forms 8, 9, and 10 composites and the AFQT validity coefficients were corrected for restriction in range and unreliability. The increments to validity were directly related to the GCA saturation of the composites. The lack of incremental validity for the Administrative or A composite over the AFQT was attributed to the composite's lower GCA saturation. This finding is consistent with the criterion-related validity results analyzed in the present report, which show the A composite has generally lower validity than do the other three composites for prediction of training success in their respective job clusters.

Table 47. Estimates of Proportion of Variance Accounted for by the Composites versus the AFQT

Composite[a]	r^2 Composite		r^2 AFQT	$r^2_c - r^2_a$ [b]
M	.6939	=[c]	.6585	.0354[c]
A	.1870	<	.7329	-.5409
G	.8358	>	.7250	.1108
E	.8536	>	.7029	.1327

Note. From Meta-analysis of Armed Services Vocational Aptitude Battery composite validity data (p. 57) by N.S. Stermer, 1988, Unpublished master's thesis, St. Mary's University, San Antonio, TX.

[a]Abbreviations are defined in Table 3.

[b]Confidence intervals indicate no significant difference between the composite versus the AFQT for these groups; thus, the r^2 differences may be interpreted as zero.

[c]Difference between r^2 composite and r^2 AFQT.

Stermer's (1988) analysis attributed the lower GCA saturation to the inclusion of the speeded tests in the A composite. This line of reasoning is similar to that used by Hunter (1984) and Hunter et al. (1985) in defining the role of general cognitive ability in prediction of job performance. However, Stermer's (1988) results are inconsistent with those of Hunter and his associates, who estimated that including a speed factor would raise the validity of the best 'g' composite from .55 to .59 for clerical occupations. Results from Stermer (1988) also indicated that the increment was moderated by gender, with females showing a larger increase in validities for the AFQT in A jobs than males.

Confidence intervals were estimated from uncorrected validities and the corrections applied to the endpoints as indicated in Table 48. Results indicated that the validity corrected values were .78 + .09, for females and .65 + .06 for males. These intervals did not overlap; therefore, Stermer concluded the validities were higher for females.

This gender-moderated effect, especially for clerical composites, was similar to results of a study by Dunbar and Novick (1984) in which individually constructed regression equations for Marine Corps clerical training courses were found to be significantly different between genders, with equations for one group of schools showing overprediction of female performance and another cluster of schools exhibiting underprediction.

Table 48. Corrected Validities and Confidence Intervals for Comparison Conditions

	Composites			AFQT	
Composite[a]	Mean r	95% CI		Mean r	95% CI
M	.8330	.8189 to .8477	=	.8115	.7981 to .8287
A	.4325	.3701 to .4859	<	.8561	.8331 to .8768
G	.9142	.9030 to .9202	>	.8515	.8416 to .8660
E	.9239	.9002 to .9218	>	.8384	.8279 to .8673

Note. From Meta-analysis of Armed Services Vocational Aptitude Battery composite validity data (p. 56) by N. S. Stermer, 1988, Unpublished master's thesis, St. Mary's University, San Antonio, TX.

[a]Abbreviations are defined in Table 3.

Ree (1990) extended the Jones (1988) study to explore the contribution of specific and general ability to prediction of training success. Ree, like Jones, used principal components analysis, regressing FSG onto the 10 principal components of ASVAB subtests determined in the 1980 Youth Population Sample Study (DoD, 1982b). Ree then calculated the 10 principal components scores of the Air Force recruits in the study (N = 78,049) for 89 different Air Force occupational specialties. The smallest sample was 274, and the largest was 8,384. The first component in Ree's research was used as an estimate of GCA, or 'g,' and the remaining components were used as estimates of specific ability.

Ree's results indicated that the first principal component (his estimate of 'g') had the strongest relationship to the criterion of FSG, entering first in the within-specialty, stepwise regression equations for all 89 of the Air Force specialities he analyzed. Other specific components, however, added validity over that of 'g.' The average increment to R^2 was .02 for prediction of training success.

Rossmeissl and Stern (1983) examined the validity generalization of the ASVAB Forms 8, 9, and 10 for 11 Army occupations. The N's for this study ranged from 91 to 613. Table 49 presents the results of their validity generalization study. The authors noted that large proportions of the variation in the observed validity coefficients were due to sampling error and that these proportions were greater for subtests than for composites (58% for subtests versus 40% for composites). The authors did not perform any of the other two usual corrections applied in meta-analytic studies--those for restriction in range and for unreliability in the criterion and predictors.

Implications of Validity Generalization Results and the Role of General and Specific Abilities

Based on the validity generalization results, it would seem important for the Services' classification system, as well as for high school counselors, to understand that 'g' or GCA is prepotent in prediction. Attempts to classify individuals on the basis of aptitude composite scores or patterns of scores is useful; however, because the predictive power of the 'g' component of any measure of specific aptitude is probably greater than that of the specific component by a factor of 5 or 6, it is difficult to describe the differential validity of a set of specific measures.

Yet, there does seem to be some contribution to prediction by ASVAB measures of specific abilities. The results of studies by Alley et al. (1988), Dunbar et al. (1985), Dunbar and Novick (1984), Hunter et al. (1985), Jones (1988) and Ree (1990) all indicate there is differential validity in the ASVAB for prediction of training success, and that the difference is generally small in terms of predicted criterion variance. Estimates of specific abilities' contribution to prediction of final training course grades from these studies is 1% to 15%.

Table 49. Validity Generalization Results for ASVAB Forms 8, 9, and 10 Subtests and Composites

	Observed average r	Observed variance in r	Percentage variance due to sampling error	Residual variance in r	Estimated mean true validity
			Subtests[a]		
GS	.2449	.0073	57.5	.0031	.6054
AR	.3222	.0124	31.4	.0085	.6480
WK	.2412	.0119	35.3	.0077	.5865
PC	.2433	.0095	45.3	.0052	.5659
NO	.0725	.0079	62.5	.0030	.4863
AS	.2235	.0059	73.6	.0016	.4875
MK	.3205	.0143	28.0	.0103	.6302
MC	.2485	.0079	54.2	.0036	.5623
EI	.2500	.0064	65.6	.0022	.5829
CS	.1191	.0039	125.9	-.0114	.4366
			Composites		
AFQT	.3324	.0155	24.0	.0118	.6694
Clerical	.2238	.0066	67.7	.0021	.5848
Motor Maintenance	.3214	.0111	33.1	.0074	.6490
Operators/ Foods	.3248	.0100	36.9	.0063	.6637
Electronics	.3770	.0116	29.7	.0081	.6968
Surveillance/ Communications	.2903	.0065	61.8	.0025	.6403
Skilled Technical	.3463	.0130	27.8	.0094	.6877

Note. From The application of meta-analytic techniques in estimating selection/classification parameters, by P.G. Rossmeissl and B.M. Stern, 1983. In Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 fiscal year (ARI-TR-660) (p. 429-430), by N. K. Eaton, M. H. Goer, J. H. Harris and L. M. Zook, 1983, Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

[a]Definitions of abbreviations are found in Table A-2.

Other evidence for the existence of situation-specific criterion variance that may or may not be well predicted by the ASVAB comes indirectly from the validity studies for specific selector screening tests reviewed under the comparison section of the present report. That other specific screens showed significant increments to validity over ASVAB selector composites for specific military occupational specialties like Air Traffic Controllers (Stoker et al., 1987) indicates the need for prediction of situationally specific abilities for at least some occupations.

VI. SUBGROUP VALIDITY ANALYSES

General

Measures may be valid, but not necessarily equitable for various subgroups. Analysis of the equity of a measure is a necessary part of the analysis of its validity. Equity analyses are done in order to determine if a given measure is equally valid for different subgroups.

Unlike large civilian organizations, however, the Military has some unique features that present special problems for equity analyses. For example, there is a legal issue in relation to gender differences in military selection and classification. Though the types of careers open to women have increased over the years, females are still barred by law from entry into occupational specialties identified as combat jobs--infantrymen, field artillerymen, tank gunners, crewmembers on ships of war and combat aircraft, etc. The determination of what constitutes a viable career choice for a woman is not the issue for this review. However, those issues are part of the context for analyzing the validity of the composites in regard to gender differences. As the role of women in society has changed, so have the types of military jobs for which women apply. The summary validity information for gender-related data is not as nearly complete as for ethnicity. This indicates lack of female members in certain military occupational areas that may be the result of either culturally or statutorily enforced selection.

ASVAB Subgroup Reliabilities

The reliabilities of the the ASVAB subtests and composites for the different subgroups of interest are presented in Tables 50 and 51. Table 50 shows alternate forms reliabilities of the subtests and composites for Forms 11a with Forms 9 and 10, from the study by Palmer et al. (1988), for Blacks, Whites and Hispanics. Table 51 presents reliability information for males and females from the same study.

The reliabilities appear to be fairly uniform across subgroups, with the subtest reliabilities for the Hispanics tending to be slightly higher than those for either the Blacks or the Whites. The composite reliabilities are similarly uniform. The uniform reliabilities across subgroups indicates relatively equal precision of measurement for the ASVAB subtests and composites for the ethnic and gender groups.

The validity information on ethnic differences will be presented at two levels: score-level differences and item-level differences. Although score-level differences in validity coefficients have been observed and will be reported here, differences in validity coefficients do not indicate the presence of bias. Validity coefficients are very prone to artifactual distortion. This is the reason for using linear models analyses to detect bias.

Table 50. Alternate Forms Reliability (r) of Subtests and Composites of ASVAB Form 11a with ASVAB Forms 9a, 9b, 10a, and 10b for Blacks, Hispanics and Whites

	Blacks r	Whites r	Hispanics r
Subtests[a]			
GS	.77	.80	.81
AR	.81	.84	.86
WK	.85	.85	.87
PC	.65	.66	.68
NO	.65	.69	.71
CS	.69	.65	.75
AS	.73	.80	.81
MK	.76	.82	.86
MC	.65	.70	.73
EI	.59	.64	.73
Composites[b]			
M	.85	.89	.89
A	.84	.83	.87
G	.90	.89	.92
E	.89	.89	.92
AFQT	.90	.89	.92

Note. From Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (AFHRL-TR-87-48, AD-A191 658) (p. 14) by P. Palmer, D. D. Hartke, M. J. Ree, J. R. Welsh, and L. D. Valentine, Jr., 1988, Brooks AFB, Texas: Air Force Human Resources Laboratory.
[a]Definitions for abbreviations are in Table A-2.
[b]Definitions for abbreviations are in Table A-5.

Table 51. Alternate Forms Reliability (r) of Subtests and Composites of ASVAB Form 11a with ASVAB Forms 9a, 9b, 10a, and 10b for Males and Females

Subtests[a]	Males r (11a)	Females r (11a)	Composites[b]	Males r (11a)	Females r (11a)
GS	.84	.80	M	.91	.85
AR	.87	.86	A	.87	.83
WK	.88	.88	G	.93	.92
PC	.71	.68	E	.93	.92
NO	.70	.64	AFQT	.93	.92
CS	.73	.70			
AS	.83	.69			
MK	.86	.82			
MC	.77	.69			
EI	.71	.56			

Note. From Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (AFHRL-TR-87-48, AD-A191 658) (p. 13) by P. Palmer, D. D. Hartke, M. J. Ree, J. R. Welsh, and L. D. Valentine, Jr., 1988, Brooks AFB, Texas: Air Force Human Resources Laboratory.

[a]Definitions for abbreviations are found in Table A-2.

[b]Definitions for abbreviations are found in Table A-5.

Given that a test battery is relatively free of bias, one still needs to examine the functioning of the battery or test in a particular selection system for differences in prediction due to subgroup membership. It is incorrect to assume that differences in validity coefficients alone constitute a demonstration of test bias. A more complete definition of equity involves examination of the selection system for differences in prediction for subgroups of interest. Equity data will be presented in this review in both forms. First, the results of two large studies will be presented which show the differences in validity coefficients for various ethnic and gender subgroups. Then the data examining test bias based on regression models will be presented and discussed. Because simple mean differences on the predictors alone do not constitute evidence of test bias, studies that deal with such mean differences on the ASVAB without discussion of criterion differences or differences in validity will not be discussed.

Studies of Ethnicity Differences

Tables 52 and 53 contain ethnicity and gender validity data for final school grade from two large studies--Wilbourn et al. (1984) for Air Force recruit data, and Booth-Kewley et al. (1984) for Navy recruit data. Other published studies have reported ethnicity validity data for ASVAB Forms 8, 9, and 10 (Maier & Truss, 1984; Rossmeissl & Brandt, 1984) but data were presented in a form not amenable for convenient summary.

Differences in mean validities between White and Black subgroups from the Wilbourn et al. (1984) study are evident from inspection of Tables 52 and 53. These differences in mean validities are not consistent with the literature addressing racial effects on ASVAB validity (cf., Bock & Moore, 1984; DoD, 1984a, 1984b), nor can they be explained by subgroup differences in reliabilities. Their results, however, may be explained by relative restriction in ranges of abilities and consequent reduction in variance of scores of the two groups in the Air Force sample. This reduction of variance produces an artificial reduction in the observed correlations.

Table 52. ASVAB Forms 8, 9, and 10 AFQT Validities for Gender and Ethnic Group Membership Against Final School Grade (FSG)

Study	White r^a(SD) n	Black r^a(SD) n	Male r^a(SD) n	Female r^a(SD) n
1[b]	.41(.14) 3,346	.20(.12) 715	.37(.20) 2,816	.42(18) 633
2[c]	.41(.10) 24,256	.29(.14) 4,118	.42(.10) 26,259	.37(.12) 2,925

Note. All correlations uncorrected for restriction in range.
[a]averaged validity.
[b]Data are from Predictive validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, 10 against 100 Navy schools (NPRDC-TR-85-15) by S. Booth-Kewley, P. P. Foley, and L. Swanson, 1984, San Diego, CA: Navy Personnel Research and Development Center.
[c]Data are from Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to the Air Force technical school final grades (AFHRL-TP-84-8, AD-A144 213) by J. M. Wilbourn, L. D. Valentine, Jr., and M. J. Ree, 1984, Brooks AFB, TX: Air Force Human Resources Laboratory.

Table 53. ASVAB Forms 8, 9, and 10 Validities for M, A, G, E Composites by Ethnicity and Gender Against FSG

Study	White r^a(SD) n	Black r^a(SD) n	Male r^a(SD) n	Female r^a(SD) n
		Mechanical		
1[b]				
2[c]	.23(.11) 917	.44(.12) 7,986	.44(.11) 8,609	.25(.15) 534
		Administrative		
1				
2	.07(.11) 977	.25(.10) 2,076	.26(.09) 2,221	.1(.14) 949
		General		
1	.20(.10) 620	.41(.17) 2,907	.38(.22) 2,464	.48(.11) 484
2	.29(.10) 2,007	.45(.07) 8,618	.45(.08) 9,806	.42(.10) 1,133
		Electronics		
1				
2	.38(.12) 217	.48(.09) 5,575	.49(.09) 5,623	.33(.22) 309

Note: All correlations uncorrected for restriction in range.
[a]Averaged validity.
[b]Data are from Predictive validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, 10 against 100 Navy schools (NPRDC-TR-85-15) by S. Booth-Kewley, P. P. Foley, and L. Swanson, 1984, San Diego, CA: Navy Personnel Research and Development Center.
[c]Data are from Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to the Air Force technical school final grades (AFHRL-TP-84-8, AD-A144 213) by J. M. Wilbourn, L. D. Valentine, Jr., and M. J. Ree, 1984, Brooks AFB, TX: Air Force Human Resources Laboratory.

The Administrative (A) composite appears to have the greatest difference between the two groups. But Wilbourn et al. (1984) pointed out that the data used in their Air Force validation study suffered from severe restriction in range due to selection, and the validity coefficients were not corrected.

Other studies such as Maier and Truss (1984) indicate that the ASVAB is relatively free from subgroup bias. The results of their analyses are presented in Table 54. Using regression analysis and general linear models tests, Maier and Truss (1984) examined Marine Corps validity data from ASVAB Forms 5, 6, and 7, as well as ASVAB Forms 8, 9, and 10, for 34 Marine Corps courses.

One can get an estimate of the magnitude of the effect of social or gender subgroup on the validity of the occupational composites in Table 54 by examining the differences between validities of the composite by itself and the multiple correlation of the aptitude composite score which includes a racial grouping or gender grouping variable. One can see that the magnitude of the effect on validity due to group membership is very small, generally ranging from .00 to .06.

The results of the general linear models analysis indicated that there was no difference in validity between Blacks and Whites for the occupational composites used in the high school testing program. Maier and Truss (1984) found no systematic over- or underprediction of performance based on their sample of Marine Corps recruits. The same was not true for gender differences. Maier and Truss (1984) found some underprediction of female performance in two traditionally female occupations--Food Service and Administrative/Clerical Marine Corps occupational specialties.

Findings from the majority of the research on test equity and the ASVAB more closely parallel those obtained in the Maier and Truss (1984) study. The Maier and Truss (1984) results are also consistent with previous equity research using ASVAB Forms 5, 6, and 7.

McLaughlin et al. (1984) examined ASVAB Forms 8, 9, and 10 for ethnicity and gender differences in a very large study of Army recruits (N = 65,193). Their analyses examined the differences between subgroup and common regression lines. They showed that the patterns of the relationships among subgroup, majority, and common regression lines were very similar in the region near the then-current minimum aptitude cutoff scores. These results indicate that in regions of the aptitude scale that are most important in making personnel decisions, there are few or no differences among groups. Thus, selection on ASVAB predictors with existing minimum aptitude qualifying scores would not likely result in any adverse impact on one group. However, McLaughlin et al. (1984) noted regression line differences that resulted in slight overprediction of criterion performance of Black soldiers when the common regression line was used.

Table 54. Effects of Social Grouping on Validity of Aptitude Composites

Sample	Regression Weight[a]				Validity[b]	
Title	Composite[c]	Race	Education	Gender	Composite	Multiple
			Mechanical Maintenance			
Engineer Equipment Operator	.24**[d]	1.46	2.17*	1.18	.33	.34
Combat Engineer	.38**	2.38*	2.58**	--	.58	.60
Automotive Mechanic	.40**	1.29	5.09**	1.88	.58	.61
Aircraft Mechanic	.36**	-3.21	2.40**	-3.90*	.38	.40
Helicopter Mechanic	.25**	-2.21	.97	--	.28	.29
Tracked Vehicle Repair	.41**	-.07	3.33**	-7.88	.50	.54
Aircraft Maintenance	.36**	-.69	2.74**	-.99	.43	.44
Electrical Equipment Repair	.36**	-1.22	-.01	-.95	.42	.42
Airfield Services	.25**	4.82**	1.78	7.60**	.40	.46

Table 54 (Continued)

Sample Title	Regression Weight[a] Composite[c]	Race	Education	Gender	Validity[b] Composite	Multiple
			Clerical			
Administration	.31**	.41	2.16*	-3.27**	.44	.48
Administration	.25**	1.38	2.47**	-2.77*	.32	.35
Communications Center	.29**	-.46	.98*	-1.92	.36	.37
Supply Stock Control	.45**	1.33	2.29*	-.60	.54	.55
Intelligence/ Operations	.38**	-2.90	.14	-1.68	.40	.42
Supply	.36**	-.01	3.50**	-1.83	.40	.42
Finance and Accounting	.50**	-.54	.34	-.44	.52	.52
			Electronics Repair			
Radio Operator	.28**	.66	3.39**	-.47	.32	.36
Basic Electronics	.48**	-.50	3.32**	-1.40	.47	.48
Basic Electricity and Electronics	.52**	-1.12	2.39**	-.77	.45	.45
			General Technical			
Ammunition Storage	.38**	-.13	6.37**	-1.39	.48	.53
Logistics	.25**	1.99	2.89	-2.89*	.27	.33
Food Service	.32**	3.19**	1.28	-4.41**	.43	.48
Aviation Ordnance	.48**	-1.11	3.71**	.90	.43	.44

Table 54 (Concluded)

Sample Title	Regression Weight[a]				Validity[b]	
	Composite[c]	Race	Education	Gender	Composite	Multiple
			Combat			
Rifleman	.26**	1.77**	-.28	--	.39	.40
Rifleman	.26**	1.18	2.57**	--	.27	.30
Machine Gunner	.23**	2.08*	-.04	--	.33	.34
Machine Gunner	.21**	1.24	2.64	--	.27	.30
Mortarman	.30**	2.23*	-.58	--	.43	.44
Mortarman	.15*	.91	2.21	--	.18	.21
Assaultman	.25**	3.15**	.65	--	.36	.38
Assaultman	.25**	-1.07	1.52	--	.27	.28
			Field Artillery			
Fire Control	.35**	.45	3.26**	--	.48	.50
Amphibian Crew	.34**	1.34	3.18**	--	.42	.50
Anti-Air	.32**	3.56**	-.45	--	.43	.45

Note. From Validity of the Armed Services Vocational Aptitude Battery Forms 8, 9, and 10 with applications to Forms 11, 12, 13, and 14 (CNR-102) (p. 39-40) by M. H. Maier and A. R. Truss, 1985, Alexandria, VA: Center for Naval Analyses.

[a]Regression Coefficients are shown for appropriate aptitude composite and subgroups (Whites versus Blacks and others; high school graduates versus nongraduates; males versus females);sample values are used.

[b]Validity coefficients are shown for aptitude composite score by itself and multiple correlation for aptitude composite score plus subgroups. These are uncorrected for restriction in range.

[c]The appropriate composite was used for each occupational group.

[d]Regression weights significant at the 1-percent level are shown by **; those significant at the 5-percent level are shown by *.

That there are mean differences of aptitude composite scores between Blacks and Whites is consistent with the majority of the literature on tests of mental ability. Group mean differences on tests are not necessarily an indicator of bias, any more than group mean differences in height should be considered an indicator of bias in yardsticks. Eitelberg (1981) documented a pattern of mean differences in AFQT percentiles--differences typically found in the military testing experience with the AFQT. The Eitelberg (1981) review of the literature on subgroup differences in tests of mental ability indicated the average (median) AFQT score for male, non-prior-service non-White enlistees' was about 25 percentile points below the median AFQT score for White non-prior-service males (Eitelberg, 1981, p. 9). In a previous discussion of the correlation of AFQT and the WAIS IQ, it was noted that the pattern of correlations between the AFQT and the IQ measures was similar for both Whites and Blacks. Jensen (1980) also indicated mean differences in IQ scores between Blacks and Whites at about 1.2 standard deviation units (of the majority group). This was about the order of magnitude of the differences found by Eitelberg (1981).

Studies of Gender Differences in Validity

As noted previously, the study of gender differences in validity in the Military is complicated not only by the traditional issues of social and cultural constraints which direct women's occupational choices, but by the statutory prohibitions against women in combat-related jobs.

Table 51 shows the alternate forms reliability estimates of ASVAB subtests and composites for males and females. The pattern of reliability estimates for the subtests is generally similar between males and females, with the three notable exceptions of the Auto and Shop Information (AS), Mechanical Comprehension (MC), and Electronics Information (EI) subtests which show lower reliabilities for women. These differences in reliabilities indicate differences in the precision of measurement in the ASVAB at the subtest level between men and women. The Services make no personnel decisions based solely on ASVAB subtest scores. Only composites are used. The M, A, G, and E composites' reliabilities, as well as that of the AFQT, are more uniform and consistent for males and females.

Tables 52 and 53, previously presented and discussed in connection with ethnic differences, contain information relevant to gender differences in predictive validity of the ASVAB. There are differences in the mean validities between males and females from the

Wilbourn et al. (1984) and Booth-Kewley et al. (1984) studies on both the AFQT and the MAGE composites. The lower validity of the M composite for women found in the Wilbourn et al. (1984) study is consistent with the lower reliabilities for women of the subtests which make up the M composite (MC + GS + 2AS). These findings are reasonably consistent with the previously presented results of the Maier and Truss (1985) study (Table 54). The lower averaged validity of females in the A composite cluster of jobs is somewhat surprising, but replicated elsewhere (see Dunbar & Novick, 1984). The A or Administrative composite is used to select recruits for clerical or administrative military jobs. The subtests which comprised this composite were VE + NO + CS. The two speeded subtests, which have the lowest of the subtest reliabilities and show the most subgroup variation, make up two-thirds of the composite. This might explain some of the observed difference for males and females, though available evidence seems to indicate that the speeded subtests are equally unreliable for both gender groups.

Maier and Curia (1986) examined previous validity studies that indicated that gender differences in expected performance were most pronounced in Clerical and Food Service specialties. Four courses from these fields were examined using regression equations to predict final course grade. Variables in the equations were aptitude composite scores, level of education, race, interest, and gender. Results indicated that in the Administrative Clerk course women had consistently higher expected performance than did men, taking into account race, education level, and interest. This result can be taken as a form of bias, given that the females' performance on the criterion of final school grade was underpredicted. Maier and Curia (1986) noted that though females consistently outperformed males in this specialty, there were no differences in the accuracy of predictions for males and females as indexed in the results of their study by the consistently smaller standard errors of estimate for females. The two regression equations had significantly different intercepts, however.

McLaughlin et al. (1984) reported very slight differences in validity patterns for males and females. These differences were .06 after correction for restriction in range. These Army researchers' results also replicated a familiar pattern of male/female differences in that females' performance was consistently underpredicted in clerical or administration-type jobs. The performance of females in more non-traditional E- or M-type jobs was consistently overpredicted by the male regression line or the common regression line.

Differential Item Functioning (DIF)

DIF is relevant to validity in at least one aspect: Items on a test should be free from contamination by variables unrelated to those being measured. Test items should be free from influence from sources of variation from variables such as gender or ethnic group membership; hence, DIF is included in this review of ASVAB validity studies. Only one study was published on DIF on the ASVAB (Linn et al., 1988). Its primary purpose was to examine and compare 27 different indices of DIF. Although performed on a large sample of high school students who took ASVAB Form 14, the comparisons are relevant to this topical area. Specifically, Linn et al. (1988) noted that there were few general conclusions one could make about the items that were found to function differently for Blacks, Whites and Hispanics; or males and females. They did offer some general observations that could serve as a basis for future research.

Specifically, on the General Science subtest, Whites tended to do better on physical science items than did Blacks or Hispanics, whereas the reverse was true for life science items. On Word Knowledge, Whites tended to do slightly better on words commonly found in science texts. On the Mechanical Comprehension and Electronics Information subtests, those items displaying the greatest DIF between males and females were items that seem to be part of a specialized vocabulary. Just how important knowledge of those words may be to the measurement of the constructs in question is a difficult judgment to make.

In general, the Linn et al. (1988) study found a few ASVAB items that favored a focal group (a focal group in DIF research is generally taken as the minority subgroup in ethnicity studies and the male group in studies of gender-related DIF), and some favoring a reference group (defined as the majority ethnic group--usually Whites--or males in studies of gender-related DIF). They also found that because the occurrence of DIF was counterbalanced within a subtest, on average there was no constant advantage for any subgroup.

Linn et al. (1988) noted that it is difficult to make generalizations from one DIF study, and that what is needed is a large body of such studies to generate conclusions useful to test-builders.

In a study done on ASVAB Form 8ax as part of the Profile of American Youth Study (DoD, 1982b), Bock and Mislevy (1981) examined the test items on ASVAB with respect to cultural and gender bias. Though not actually a DIF study in the current sense of the term, it represented a seminal piece of research into the general topic of item bias on the ASVAB. Bock and Mislevy (1981) investigated cultural or ethnicity bias on ASVAB Form 8a using Item Response Theory (IRT) techniques to explore the null hypothesis that if the ASVAB items were measuring the same underlying variable for all subjects regardless of group membership, then the item parameters estimated from the responses of any subgroup of subjects would be statistically equivalent. The authors tested this hypothesis using a Chi-Square test for the equality of the item parameters. They concluded that:

> Test item cultural bias, with regard to sex and racial/ethnic groups, was not apparent in the ASVAB power subtests, with two minor exceptions: (1) The General Science subtest is a composite of three distinct variables, Physical Science, Health Science, and Biology. The use of a single General Science scores tends to overestimate the Health Science ability of males but underestimate their physical science abilities, while the reverse is true for females. (2) Word Knowledge shows certain items that are slightly but systematically easier for Whites, while others are slightly but systematically easier for Blacks and Hispanics. The words that were relatively easier for Whites tended to be more 'literary.' As indicated by the analyses of subject-fit indices, the disturbances caused by these group-by-item interactions are not severe. (p. 43)

Summary of Equity Studies

Perhaps the easiest part of a summary of equity studies of the ASVAB at either the score level or the item level is the observation that there are not many of them. Although the Military has its own unique problems regarding the classification and assignment of females, the usual difficulties that preclude civilian organizations from adequate equity studies--that of insufficient sample sizes for minorities--does not present an excuse for the Services. There are adequate sample sizes for many more studies of equity than were reviewed here. Moreover, there is enough, though spotty, evidence of some under- and overprediction of minority groups for some specific military occupations to warrant more thorough investigations of the equity of Service-specific selector composites. McLaughlin et al. (1984), Maier and Truss (1984) and Wilbourn et al. (1984) presented results which, on the whole, present a picture of a reasonably equitable selection and classification system. There are, however, nagging exceptions where slight over- or underprediction (there does not seem to be a clear, consistent pattern) does occur for Blacks in specific specialties.

There are more notable differences discovered when one examines the evidence for differences in gender-related prediction. The Maier and Curia (1986) study and other evidence points to more consistent, troublesome differences related to gender in the Administrative and Food Service specialties. Consistent underprediction of females for these specialties should have sparked a search for more valid predictors for females in these military occupational areas. There are many questions unanswered by the available validity research on both gender and ethnicity equity issues.

Based on equity studies at the item level, the ASVAB is an equitable test insofar as it appears to be free of racial or ethnic bias in the large majority of the test items. The Services should develop a systematic study of DIF on the ASVAB. Given the large sample sizes in enlistment testing research, there is much that can be learned about real between group differences in item functioning as well as undesirable differential item functioning. Many more careful, systematic equity studies at both the score level and the item level need to be conducted. Particular attention should be directed toward incorporating the study of DIF into the various developmental stages of the ASVAB.

VII. SUMMARY AND CONCLUSIONS

Criterion-Related Validity

A number of different types of criteria were discussed in the present review: final technical school grade, first-term (less than 4-year initial tour) attrition, job knowledge tests, hands-on-performance tests, and performance ratings. All these measures are relevant to particular personnel and training policy decisions at differing points in time in a new recruit's tenure in the Service, just as they are relevant to differing types of manpower policy decisions. The criterion problem for military manpower policy-makers is not so much that of finding a meaningful criterion as it is determining which of the several available are most relevant to the decisions at hand. Other considerations of national, social and economic policy, as well as current social and economic conditions, often outweigh or at least complicate any straightforward use and interpretation of ASVAB validity information in making selection and classification decisions.

In spite of complicating factors, the Military has enjoyed the benefit of massive amounts of data on the validity of its cognitive aptitude tests. From the studies reviewed here, a number of summary statements and conclusions can be made about the criterion-related validity of the ASVAB aptitude composites.

The clearest conclusion from this review is that the ASVAB-based AFQT and MAGE composites are valid predictors of enlisted, entry-level final technical school grades. Results from empirical, criterion-related studies have consistently shown that these five composites are correlated with final school grades (FSG) in the range of .55 to .60 (corrected for restriction in range). The order of magnitude of these coefficients for these composites does not seem to vary by Service.

The validity of the AFQT and MAGE composites for a training success criterion of time-to-completion (TTC) of self-paced entry-level courses is consistently lower than that for prediction of FSG. These validities ranged from -.25 to -.36 (corrected for restriction in range).

The averaged validity for each of the five ASVAB-based composites for combined job performance criteria is somewhere between the validities for FSG and TTC. Averaged validities ranged from .35 to .47 (corrected for restriction in range).

To the Military, success in training is, and will continue to be, the most important criterion for prediction. The large and expensive military training systems have to be able to estimate the abilities of those who elect to apply. The Services need to select and classify appropriately the most trainable applicants. The ASVAB composites have a clearly demonstrated validity for that purpose.

Content Validity

The present review of the published reports documenting the development of the current generation of the ASVAB indicates that these forms are content valid. This conclusion is based in part on the process by which ASVAB forms are built. The third generation of the battery is carefully tied to a published taxonomy based on the reference form. Item-level factor analytic work based on a nationally representative sample of American youth indicates that each ASVAB subtest is relatively free from the confounding measurement of variables unrelated to the subject-matter content of the subtest.

Construct Validity

The ASVAB subtests show significant positive relationships to like-named subtests of other multiple-aptitude batteries. Moreover, years of factor analytic studies of successive operational forms of the ASVAB show the same factor pattern and the same relationships among the subtests of the ASVAB.

The issue of the validity of the ASVAB for use as a counseling and career exploration tool for students and counselors in the DoD Student Testing Program is discussed at some length because the validity of the ASVAB for civilian occupational choice and success must generalize from the well-established relationships with military occupational training success to the civilian world of work. Such an exercise depends on arguments founded in validity generalization and situational specificity.

The issue of the validity of the ASVAB for purposes of counseling high school students in career exploration is not as controversial as may at first be supposed. The ASVAB, as demonstrated in military settings for military occupations, is definitely a valid predictor of success in entry-level training for military occupations. The extent to which the validity of the ASVAB generalizes to civilian occupations is less clear, but the evidence suggests that the validity of the ASVAB does generalize to similar civilian occupations.

What is perhaps arguable is the idea that the ASVAB validity generalizes because of its high 'g' saturation. The validity generalization results of Hunter and his associates (1985) indicate that the ASVAB is indeed a more valid predictor of civilian job performance or proficiency than is the GATB largely because the ASVAB is a better measure of 'g.' Hunter also maintained that the causal path from ability to job performance is through job knowledge; and job knowledge is best predicted by measures of general cognitive abilities.

Opposed to this view are those who believe in the doctrine of situational specificity and hold strong to the notion that humans have differing, and separate, cognitive abilities. Accordingly, they believe use of measures of the specific abilities necessarily leads to better prediction of criteria composed of complex and specific patterns of abilities necessary for successful job performance.

Thorndike (1985) reanalyzed three sets of data on the predictive validity of civilian and military multiple-aptitude batteries and found that a common factor ('g') explained between 60% and 120% more of the common systematic variance as was explained by regression weighted composites, in cross-validation. Thorndike argued that this supports the widespread validity of 'g' as a predictor of job performance. He further noted that it is the widespread and generalizable validity of 'g' that accounts for the generalizability of cognitive tests across job situations. Thorndike estimated, however, that about 10% to 15% of additional variance is likely to be accounted for with regression-weighted composites of tests of specific ability.

There is solid and clear evidence for some type of situational specificity and the subsequent need to measure specific abilities or patterns of specific abilities in prediction of job performance. This evidence comes from the results of Hunter et al. (1985) and Schmidt et al. (1987) in their analysis of the increments to validity that would accrue if subtests of perceptual ability were added to the ASVAB. The evidence also comes from the work by Dunbar et al. (1985) and Alley et al. (1988), which indicates that not all jobs have the same, homogeneous regression equations. The data reviewed here on gender and ethnic differences also indicated differing patterns of prediction for some specific types of jobs. Last, but not least, the press of practical necessity has led some Services to search for, develop, validate, and use other supplemental selection and classification screening tests for specific military occupations that experience unusual or unacceptably high attrition from training.

The problem is to determine which jobs require specific validation and what specific abilities need to be assessed in order to successfully predict training success and to assess the cost benefits, etc. What is needed in terms of future research at this point are studies which examine the contribution to predictive validity of specific and general cognitive ability within military jobs (Ree & Earles, 1990). All the studies reviewed here, with the exception of Ree's study, examined validity across jobs. This procedure can have the effect of averaging-out any existing differences in the patterns of subtest validities within a given job or occupational specialty.

A number of specific recommendations for future research emerge from the present review. The gaps in validity research work are particularly evident in relation to the DoD Student Testing Program. Though the results of this review support the use of validity generalization of ASVAB validities from military to civilian jobs, there is sufficient indication of situational specificity in the military job setting to warrant several good criterion-related validation efforts of ASVAB for some types of civilian training criteria and job performance criteria (e.g., well-developed supervisor's rating scales such as BARS).

The DoD Student Testing Program, as well as the operational testing program, could benefit from analyses of the equity of the ASVAB in a civilian-sector validation study. The effects of restriction in range on equity analysis for military occupations could be better gauged; and, in particular, gender differences in predictions for civilian and military occupations could be brought into sharper focus. Moreover, ethnic or gender-related DIF could be better understood by comparing the occurrence (or lack of) DIF in either civilian or military samples.

In general, the ASVAB is a relatively unbiased test, but the military test equity research tends to be concerned only with composite score-level bias. Although the military studies of test equity suffer the same problems as do civilian sector studies, the sample sizes and restriction in range problems in the military are not nearly as great or as intractable as they usually are in the civilian sector. This is more than just the usual call for more score-level and item-level research on the equity of the ASVAB. Given the wealth of data available, there is a serious gap in ASVAB equity research.

More research is needed on the contribution of specific versus general abilities to the prediction of all types of criteria. The issue of the differential validity of the ASVAB is perhaps most relevant to the high school testing program; but, again, both the Services' recruits and the nation's high school students could benefit from a clearer determination of the contributions of specific abilities and general ability to successful job performance. It seems a worthy conclusion, based on the validity data reviewed here, that measures of general cognitive abilities predict about 60% to 70% of the criterion variance in training-success-type criteria. Solid evidence exists, however, that for some jobs, specific abilities may predict additional criterion variance. As more reliable job performance measures are developed, the role of specific abilities and the amount and type of situational specificity for given jobs may become more definable and hence more predictable.

The Services have a definite need to continue with the traditional validation studies, just as they have a need to investigate for other predictors. In that it seems apparent that--with the exception of Figural Fluency and Associative Memory--cognitive aptitude testing has gone just about as far as it will go in the prediction of success in military training, the use of non-cognitive predictors needs to be more fully explored.

REFERENCES

Albert, W. G. (1980). Predicting involuntary separations of enlisted personnel (AFHRL-TR-79-58, AD-A082 995). Brooks AFB, TX: Computational Sciences Division, Air Force Human Resources Laboratory.

Alley, W. E., Treat, B. R., & Black, D. E. (1988). Classification of Air Force jobs into aptitude clusters (AFHRL-TR-88-14, AD-A206 610). Brooks AFB, TX: Information Sciences Division, Air Force Human Resources Laboratory.

Alley, W. E., Wilbourn, J. M., & Berberich, G. L. (1976). Relationships between performance on the Vocational Interest-Career Examination and reported job satisfaction (AFHRL-TR-76-89, AD-A040 754). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

American Association for Counseling and Development. (1984). Counselors Manual for the Armed Services Vocational Aptitude Battery Form 14. Chicago, IL: United States Military Entrance Processing Command.

American Psychological Association. (1985). Standards for educational and psychological testing. Washington, DC: Author.

Andberg, M.A., Stillwell, W.G., Prestwood, J.S., & Welsh, J. R. (1988). Initial operational test and evaluation (IOT&E) of ASVAB Forms 11, 12, and 13: Parallelism of the new forms (AFHRL-TR-87-65 (I), AD-A198 460). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Armor, D. J., Fernandez, R. L., Bers, K., Schwarzbach, D. S., Moore, S. C., & Cutler, L. E. (1982). Recruit aptitude and Army job performance: Setting enlistment standards for infantrymen (R-2874-MRAL). Santa Monica, CA: Rand Corporation.

Armstrong, T. R., Chalupsky, A. B., McLaughlin, D. H., & Dalldorf, M. R. (1988). Armed Services Vocational Aptitude Battery: Validation for civilian occupations (AFHRL-TR-88-20, AD-A198 758). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Automated Sciences Group and CACI-INC-Federal. (1988). CAT ASVAB program concept of operation and cost/benefit analysis. Fairfax, VA: Department of the Navy, Office of the Chief of Naval Operations.

Bayroff, A. G., & Fuchs, E. F. (1970). The Armed Services Vocational Aptitude Battery (Tech Research Rep. No. 1161). Arlington, VA: Army Behavorial and Systems Research Laboratory.

Bock, R. D., Gibbons, R., & Muraki, E. (1985) Full information item factor analysis. Chicago, IL: National Opinion Research Center.

Bock, R. D., & Mislevy, R. J. (1981). The Profile of American Youth: Data Quality Analysis of the Armed Services Vocational Aptitude Battery. Chicago, IL: National Opinion Research Center.

Bock, R. D., & Moore, E. G. (1984). Profile of American Youth: Demographic influences in ASVAB test performance. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Installations, and Logistics).

Booth-Kewley, S. (1984a). Validation of Armed Services Vocational Aptitude Battery (ASVAB) of the English Diagnostic Test (EDT) for performance in basic journalist (JO) 'A' school (MPL-TN-84-3). San Diego, CA: Navy Personnel Research and Development Center.

Booth-Kewley, S. (1984b). Validation of Armed Services Vocational Aptitude Battery (ASVAB) selection criteria for crytologic technician collection (CTR) "A" school (MPL TN-84-2). San Diego, CA: Navy Personnel Research and Development Center.

Booth-Kewley, S. (1984c). Validation of the Armed Services Vocational Aptitude Battery (ASVAB) selection criteria for strategic weapons systems electronics "A" school (NPRDC-TR-84-22). San Diego, CA: Navy Personnel Research and Development Center.

Booth-Kewley, S. (1985). An empirical comparison of the accuracy of univariate and multivariate corrections for range restrictions (NPRDC-TR-85-19). San Diego, CA: Navy Personnel Research and Development Center.

Booth-Kewley, S., Foley, P. P., & Swanson, L. (1984). Predictive validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, 10 against 100 Navy schools (NPRDC-TR-85-15). San Diego, CA: Navy Personnel Research and Development Center.

Brogden, H. E. (1955, September). Least-squares estimates and optimal classification. Psychometrica, pp. 249-252.

Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981). Assessment of numerical skills of Navy enlisted personnel (TM-82-4). Orlando, FL: Navy Training Evaluation and Analyses Group.

Buddin, R. (1984). Analysis of early military attrition behavior (R-3069-MIL). Santa Monica, CA: The Rand Corporation.

Burkett, J. R. (1977). Functional literacy demands of Air Force jobs. Paper presented at the ASVAB Working Group Meeting. Washington, DC: Office of the Assistant Secretary of Defense (Manpower and Reserve Affairs).

Camara, W. J., & Lawrence, J. H. (1987). Military classification and high aptitude recruits (HumRRO-FR-PRO-87-21). Alexandria, VA: Human Resources Research Organization.

Caylor, J.S., Sticht, T.G., Fox, L.C., & Ford, J.P. (1973). Methodologies for determining reading requirements of military occupational specialties (HumRRO-TR-73-5). Alexandria, VA: Human Resources Research Organization.

Clemans, W. V. (1958). An index of item-criterion relationships. Educational and Psychological Measurement, 18, 167-172.

Curran, L. T., & Palmer, P. (1990). Armed Services Vocational Aptitude Battery (aSVAB): Item overlength and operational length devleopment of Forms 18 and 19. (AFHRL-TR-89-74). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Curtis, J. S., Booth-Kewley, S., & Swanson, L. (1984). Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 in two Hull Maintenance Technician Class "A" Schools (NPRDC Ltr). San Diego, CA: Navy Personnel Reserch and Development Center.

Department of Defense. (1969). Project One Hundred Thousand: Characteristics and performance of "New Standards" men. Washington, DC: Office of the Secretary of Defense.

Department of Defense. (1980). Aptitude testing of recruits (Report to the House Committee on Armed Services). Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics).

Department of Defense. (1981a). Department of Defense efforts to develop quality standards for enlistment (Report to the House and Senate Committees on Armed Services). Washington, DC: Office of the Assistant Secretary of Defense.

Department of Defense. (1981b). Implementation of new Armed Services Vocational Aptitude Battery and actions to improve the enlistment standards process (Report to the House and Senate Committees on Armed Services). Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics).

Department of Defense. (1982a). Conference on Joint-Service efforts to link enlistment standards and job peformance. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics).

Department of Defense. (1982b). Profile of American Youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery. Washington, DC: Office of the Assistant Secretary of Defense.

Department of Defense. (1984a). Armed Services Vocational Aptitude Battery (ASVAB) Test Manual (DoD 1304.12AA). Chicago, IL: Military Entrance Processing Command.

Department of Defense. (1984b). Technical supplement to the Counselor's Manual for ASVAB 14. North Chicago, IL: Military Entrance Processing Command.

Department of Defense. (1984c). Joint-Service efforts to link enlistment standards to job performance: Third annual report to the House Committee on Appropriations. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Installations and Logistics).

Department of Defense. (1985). A report to the House Committee on Appropriations: Fourth annual report to Congress on Joint-Service efforts to link enlistment standards to job performance. Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel).

Department of Defense. (1986). A report to the House Committee on Appropriations: Fifth annual report to Congress on Joint-Service efforts to link enlistment standards to job performance. Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel).

Department of Defense. (1987). A report to the House Committee on Appropriations: Sixth annual report to Congress on Joint-Service efforts to link enlistment standards to job performance. Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel).

Department of Defense. (1989). Joint-Service efforts to link enlistment standards to job performance: Recruit quality and military readiness (A report to the House Committee on Appropriations). Washington, DC: Office of the Assistant Secretary of Defense, (Force Management and Personnel).

Department of the Navy. (1981). Validation of ASVAB 6/7 selector composites for the Avionics Group (AV) and AE Class "A" school (NPRDCL, 3 Sep 1981).

Department of the Navy. (1982). Validation of ASVAB Forms 6 and 7 against school performance in Cryptologic Technician (CT) "A" schools (NPRDCL, 22 Feb 1982).

Diehl, G. E. (1981). An analysis of factor patterns in the Armed Services Vocational Aptitude Battery for high school seniors. Unpublished doctoral dissertation, Auburn, AL: Auburn University.

Dunbar, S. B., Mayekawa, S., & Novick, M. R. (1985). Simultaneous estimation of regression functions for Marine Corps technical training specialties (ONR-TR-85-1). Chicago, iL: Office of Naval Research.

Dunbar, S.B., & Novick, M.R. (1984). Gender-Differentiated Prediction: Marine Corps clerical specialties (TR-84-1). Iowa City, IA: The University of Iowa.

Eaton, N. K., Goer, M. H., Harris, J. H., & Zook, L. M. (1983). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 Fiscal Year (ARI-TR-660). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Ebel, R. L. (1983). The practical validation of tests of ability. Education Measurement: Issues and Practices, 3, 7-10.

Eitelberg, M. J. (1981). Subpopulation differences in performance on tests of mental ability: Historical review and annotated bibliography (HumRRO-TM-81-3). Washington, DC: Office of the Assistant Secretary of Defense (Manpower Reserve Affairs and Logistics).

Eitelberg, M. J., Lawrence, J. H., Waters, B. K., & Perelman, L. S. (1984). Screening for service: Aptitude and education criteria for military entry. Arlington, VA: Human Resources Research Organization.

Fernandez, R. L., & Garfinkle, J. B. (1984). Setting enlistment standards and matching recruits to jobs using job performance criteria (R-3067-MIL). Santa Monica, CA: The Rand Corporation.

Fischl, M. A., Ross, R. M., & McBride, J. R. (1979). Development of factorially based ASVAB high school composites (Tech. Paper No. 306). Alexandria, VA: Army Research Institute for the Behavorial and Social Sciences.

Fletcher, J., & Ree, M. J. (1976). Armed Services Vocational Aptitude Battery (ASVAB) correlational analysis, ASVAB Form 2 versus ASVAB Form 5 (AFHRL TR-76-70, AD-A032 593). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Flyer, E. S. (1988). Characteristics and behavior of recruits assigned to highly sensitive positions (FR-8801). Alexandria, VA: Human Resources Research Organization, International.

Flyer, E. S., & Elster, R. (1983). First-term attrition among non-prior service enlisted personnel: Loss probabilities based on selected entry factors (NPS54-84-007). Monterey, CA: Naval Postgraduate School.

Foley, P. P. (1986). Validity generalization of Navy selector composites (NPRDC-TR-86-17). San Diego, CA: Navy Personnel Research and Development Center.

Fox, W. L., Taylor, J. E., & Caylor, J. S. (1969). Aptitude level and the acquisition of skills and knowledges in a variety of military training tasks (TR-69-6). Alexandria, VA: Human Resources Research Organization.

Frankfelt, E. (1970). Comparison of ACB and ASVAB clerical tests for use in the Army Qualification Battery (Tech. Rep. BESRL-RM-70-6). Arlington, VA: Army Behavioral and Systems Research Laboratory.

Friedman, D., Crosson, J., Streicher, A., & Messersmith, D. (1986). The construct validity of the Armed Services Vocational Aptitude Battery (ASVAB) Form 14: A comparison with CAT, DAT and FIT/FACT. Unpublished manuscript.

Friedman, D., Streicher, A., Wing, H., Grafton, F., & Mitchell, K. (1983). Reliability of scores for fiscal year 1981 Army applicants: Armed Services Vocational Aptitude Battery Forms 8, 9, 10 (Research Note 85-48). Alexandria, VA: Army Research Institute for the Behavorial Sciences.

Fruchter, D. A., & Ree, M. J. (1977). Development of the Armed Services Vocational Aptitude Battery; Forms 8, 9, and 10 (AFHRL-TR-77-19, AD-A039 270). Lackland AFB, TX: Personnel Research Division Air Force Human Resources Laboratory.

Grafton, F. C., & Horne, D. K. (1985). An investigation of alternatives for setting second-to-third tour reenlistment standards (ARI-TR-690, AD-A164 694). Alexandria, VA: Army Research Institute.

Greenberg, I. M. (1980). Mental standards for enlistment: Performance of Army personnel related to AFQT/ASVAB Scores. Monterey, CA: McFann and Gray Associates.

Grissmer, D. W., & Kirby, S. N. (1985). Attrition of nonprior-service reservists in the Army National Guard and Army Reserve (R-3267-RA, AD-A161 639). Santa Monica, CA: The Rand Corporation.

Grunzke, M. E., Guinn, N., & Stauffer, G. (1970). Comparative performance of low ability airmen (AFHRL TR-70-4, AD-A705 575). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Guinn, N. (1977). USAF attrition trends and identification of high risk personnel. In H.W. Sinaiko (Ed.), First Term Enlisted Attrition, Volume I: Papers (TR-3). Arlington, VA: Office of Naval Research.

Guinn, N., Tupes, E. C., & Alley, W. E. (1970a). Cultural subgroup differences in the relationships between Air Force aptitude composites and training criteria (AFHRL-TR-70-35, AD-715 922). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Guinn, N., Tupes, E. C., & Alley, W. E. (1970b). Demographic differences in aptitude test performance (AFHRL-TR-70-15, AD-710 618). Lackland AFB, TX: Personnel Research Division, Air Force Human Resource Laboratory.

Guinn, N., Wilbourn, J. M., & Kantor, J. E. (1977). Preliminary development and validation of a screening technique for entry into the security police career field (AFHRL-TR-77-38, AD-A043 919). Brooks AFB, TX: Peronnel Research Division, Air Force Human Resources Laboratory.

Hanser, L. M., Arabian, J. M., & Martin, C. L. (1984). Validity of the high school composites as predictors of job performance. Alexandria, VA: U.S. Army Research Institute.

Harris, D. A. (1976). Classification typology for predicting performance in Air Force technical training (AFHRL-TR-76-91, AD-A039 268). Lowry AFB, CO: Technical Training Division, Air Force Human Resources Laboratory.

Hawley, J. K., Mullins, C. J., & Weeks, J. (1977). Jet engine mechanic - AFSC 426X2: Experimental job performance tests (AFHRL-TR-77-73, AD-A053 302). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Hiatt, C.M., & Sims, W.H. (1980). Armed Services Vocational Aptitude Battery (ASVAB) and job performance. Arlington, VA: Center for Naval Analyses.

Hunter, J. E. (1983) Validity generalization of the ASVAB: Higher validity for factor analytic composites. Unpublished manuscript.

Hunter, J. E. (1984). The validity of the ASVAB as a predictor of civilian job performance. Unpublished manuscript.

Hunter, J. E., Crosson, J. J., & Friedman, D. H. (1985). The validity of the Armed Services Vocational Aptitude Battery (ASVAB) for civilian and military job performance HQUSAF/MPXOA (Contract No. F41689-83-C-0025). Washington, DC: Department of Defense.

Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). Meta-analysis: Cummulating research findings across studies. Beverly Hills, CA: SAGE.

James, L.R., Demaree, R.G., & Muliak, S.A. (1986). A note on validity generalization procedures. Journal of Applied Psychology, 71, 440-450.

Jensen, A.R. (1980). Bias in mental testing. New York: The Free Press.

Jensen, H. E., Massey, I. H., & Valentine, L. D., Jr. (1976). Armed Services Vocational Aptitude Battery development (ASVAB Forms 5, 6, and 7) (AFHRL-TR-76-87, AD-A037 522). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Jensen, H. E., & Valentine, L. D., Jr. (1976). Validation of ASVAB-2 against civilian vocational-technical high school criteria (AFHRL-TR-76-16, AD-A023 118). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Jones, G. E. (1988). Investigation of the efficacy of general ability versus specific abilities as predictors of occupational success. Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Kantor, J.E., & Guinn, N. (1975). Comparison of performance and career progression of high school graduates and non-graduates in the Air Force (AFHRL-TR-75-73, AD-A022 973). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Kass, R. A., Mitchell, K. J.,Grafton, F. C., & Wing, H. (1982). Factor structure of the Armed Services Vocational Aptitude Battery (ASVAB), Forms 8, 9, 10: 1981 Army sample (ART-TR-581). Alexandria, VA: U.S. Army Research Institute for the Behavorial and Social Sciences.

Kettner, N. (1976). Armed Services Vocational Aptitude Battery (ASVAB Form 5): Comparison with GATB and DAT tests (AFHRL-TR-76-78, AD-A035 305). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Lawrence, J. H. (1984). Education standards for military enlistment and the search for the successful recruits (HumRRO FR-PRD-84-4). Alexandria, VA: Human Resources Research Organization.

Lawrence, J. H. (1988). Job complexity and military performance: A comparison of high and average aptitude recruits (HumRRO FR-PRD-88-12). Alexandria, VA: Human Resources Research Organization.

Lawrence, J. H., Waters, B. K., & Perelman, L. S. (1983). Enlisted military selection: Impacts of changing aptitude standards since 1940. In Proceedings of the 24th Annual Meeting of the Military Testing Association. Alexandria, VA: Human Resources Research Organization.

Lee, R., & Foley, P. R. (1986). Is the validity of a test constant throughout the test score range?, Journal of Applied Psychology, 71(4); 644.

Leisey, S. A., & Guinn, N. (1977). Development of a screening methodology for entry into medical technical training courses (AFHRL-TR-77-49, AD-A048 119). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Linn, R. L., Harnisch, D. L., & Dunbar, S. B. (1981). Validity generalization and situational specificity: An analysis of the prediction fo first-year grades in law school. Applied Psychological Measurement, 5(3), 281-189.

Linn, R. L., Hastings, C. N., Hu, P. G., & Ryan , K. E. (1988). Armed Services Vocational Aptitude Battery: Differential item functioning on the high school form (AFHRL-TR-87 45, AD-A193 693). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Madden, H. L., & Tupes, E. C. (1966). Estimating reading ability from the AQE general aptitude index. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Maier, M. H. (1982). Issues for defining ASVAB 11/12/13/14 aptitude composites (A briefing presented to the ASVAB Working Group, 7-9 December 1982) (CNA 82-3199/27). Alexandria, VA: Center for Naval Analyses.

Maier, M. H., & Curia M. D. (1986). Validity of the ASVAB for predicting performance in Marine Corps training courses: Gender differences (CRM 86-179). Alexandria, VA: Center for Naval Analyses.

Maier, M. H., & Fuchs, E. F. (1969). Development of improved aptitude area composites for enlisted classification (BeSRL-TR-1159). Alexandria, VA: Army Research Institute, Behavioral and Social Sciences.

Maier, M. H., & Fuchs, E. F. (1972). Development and evaluation of a new ACB & Aptitude Area System (TR-239). Alexandria, VA: Army Behavior amd Systems Research Laboratory.

Maier, M. H., & Fuchs, E. F. (1974). A comparison of three-subtest AFQT and four-subtest AFQT (ARI-RM-74-5). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Maier, M. H., & Fuchs, E. F. (1978). Differential validity of the Army aptitude areas for predicting Army job training performance of blacks and whites (TP-312). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Maier, M. H., & Grafton, F. C. (1981). Aptitude composites for ASVAB 8, 9, and 10 (Research Report 1308). Alexandria, VA: U.S. Army Research Institute.

Maier, M. H., & Hiatt, C. M. (1984). An evaluation of using job performance tests to validate ASVAB qualification standards (CNR-89). Alexandria, VA: Center for Naval Analyses.

Maier, M. H., & Sims, W.H. (1986). The ASVAB score scales: 1980 and World War II (CNR-116). Alexandria, VA: Center for Naval Analyses.

Maier, M. H., & Truss, A. R. (1983). Validity of ASVAB Forms 8, 9 and 10 for Marine Corps training courses: Subtests and current composites (Memorandum No. 83-3107). Alexandria, VA: Center for Naval Analyses.

Maier, M. H., & Truss A. R. (1984). Validity of the occupational and academic composites for the Armed Services Vocational Aptitude Battery, Form 14, in Marine Corps Training Courses. (Memorandum No. 84-3043/2). Alexandria, VA: Center for Naval Analyses.

Maier, M. H., & Truss, A. R. (1985). Validity of the Armed Services Vocational Aptitude Battery Forms 8, 9, and 10 with applications to Forms 11, 12, 13, and 14 (CNR-102). Alexandria, VA: Center for Naval Analyses.

Marcus, A. J., & Lockman, R. F. (1981). Alternative enlistment standards (CRC-448). Alexandria, VA: Center for Naval Analyses.

Martin, A.J. (1977). Trends in DoD first term attrition. In H.W. Sinaiko (Ed.), First term enlisted attrition, Volume I: Papers (TR-3). Arlington, VA: Office of Naval Research.

Mathews, J. J. (1977). Analysis aptitude test for selection of airmen for the Radio Communications Analysis Specialist course: Development and validation (AFHRL-TR-77-74, AD-A051 962). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Mathews, J. J., & Roach, B.W. (1983). Reading abilities tests: Development and norming for Air Force use (ARHRL-TR-82-26, AD-A125 913). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Mathews, J. J., Valentine, L. D., Jr., & Sellman, W. S. (1978). Prediction of reading grade levels of service applicants from Armed Services Vocational Aptitude Battery (ASVAB) (AFHRL-TR-78-82, AD-A063 656). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

May, L. J. (1986). Educational quality requirements for Marine Corps enlisted personnel (CNR 121). Arlington, VA: Center for Naval Analyses.

May, L. J., & Mayberry, P. W. (1986). Rand cost-performance model for setting qualification standards: Preliminary comments (CRM-86-200). Arlington, VA: Center for Naval Analyses.

McGrevy, D. F., Knouse, S. B., & Thompson, R. A. (1974). Relationships among an individual intelligence test and two Air Force screening and selection tests (AFHRL-TR-74-25, AD-781 033). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

McLaughlin, D. H., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., & Wang, M. (1984). Validation of current alternative Armed Services Vocational Aptitude Battery (ASVAB) area composites, based on training and Skill Qualification Test (SQT) information in fiscal year 1981 and 1982 (ARI-TR-651, AD-A156 807). Alexandria, VA: Army Research Institute.

Moreno, K., Wetzel, C. D., McBride, J. R., & Weiss, D. J. (1983). Relationship between corresponding Armed Services Vocational Aptitude Battery (ASVAB) and Computerized Adaptive Testing (CAT) subtests (NPRDC-TR-83-27). San Diego, CA: Navy Personnel Research and Development Center.

Mullins, C. J., Earles, J. A., & Ree, M. (1981). Weighting of aptitude components based on differences in technical school difficulty (AFHRL-TR-81-19, AD-A102 045). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Mullins, C. J., Earles, J. A., & Wilbourn, J. M. (1979). Calculation of predictor composites in the absence of a criterion (AFHRL-TR-79-53, AD-A080 921). Brooks AFB, TX: Personnel REsearch Division, Air Force Human Resources Laboratory.

Mullen, B. & Rosenthal, R. (1985). BASIC meta-analysis: Procedures and programs. Hillsdale, NJ: Laurence Erlbaum Associates.

Nunnally, J. C. (1978). Psychometric theory. St. Louis, MO: McGraw-Hill Book Company.

Oslund, D. A., & Clark, J. S. A. (1984). Development of an enlistment standards model for the Navy Aviation Machinist's Mate (AD) rating. Monterey, CA: Naval Postgraduate School.

Palmer, P., Curran, L. T., & Haywood, C. S. (1990). Armed Services Vocational Aptitude Battery (ASVAB) Forms 20, 21, and 22: Items development. (AFHRL-TP-89-77). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D., Jr. (1988). Armed Services Vocational Aptitude Battery: Alternate forms reliability (Forms 8, 9, 10 and 11) (AFHRL-TR-87-48, AD-A191 658). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Park, R. K., Mathews, J. J., & Ree, M. J. (1985). English Diagnostic Test: Validation for journalism-related programs (AFHRL-TP-85-8, AD-A155 734). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Pearlman, K., Schmidt, F. L., & Hunter J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65(4), 373-406.

Plag, J. A., & Goffman, J. M. (1967). The Armed Forces Qualification Test: Its validity in predicting military effectiveness for Naval enlistees. Personnel Psychology, 20(3), 323-329.

Plag, J. A., Goffman, J. M., & Phelan, J. D. (1967). The adaptation of Naval enlistees scoring in mental group IV on the Armed Forces Qualification Test (Report Number 68-23, AD-699 504). San Diego, CA: Navy Medical Neuropsychiatric Research Unit.

Plag, J. A., Wilkins, W. L., & Phelan, J. D. (1968). Strategies for predicting adjustment of AFQT Category IV Navy and Marine Corps personnel (Report No. 68-28). San Diego, CA: Navy Medical Neuropsychiatric Unit.

Prediger, D. J. (1987a). Career counseling validity of the ASVAB job cluster scales used in DISCOVER. Iowa City, IA: American College Testing Program.

Prediger, D. J. (1987b, December). Validity of the new Armed Services Vocational Aptitude Battery job cluster scores in career planning. The Career Development Quarterly, pp. 113-125.

Prestwood, J.S., Vale, C.D., Massey, R.H., & Welsh, J.R. (1985). Armed Services Vocational Aptitude Battery: Development of Forms 11, 12, and 13 (AFHRL-TR-85-16, [Vol. I, AD-A160 584; Vol. II, AD-A160 585]). Brooks AFB, TX: Manpower Personnel Division, Air Force Human Resources Laboratory.

Ramsberger, P. F., & Means, B. (1987). Military performance of low aptitude recuits: A reexamination of data from Project 100,000 and the ASVAB misnorming period. Alexandria, VA: Human Resources Research Organization.

Ree, M. J., & Earles, J. A. (1990). Armed Services Vocational Aptitude Battery: The differential validity of a differential aptitude test. (AFHRL-TR-89-59). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ree, M. J., Mullins, C. J., Mathews, J. J., & Massey, R. H. (1982). Armed Services Vocational Aptitude Battery: Item and factor analyses of Forms 8, 9, and 10 (AFHRL-TR-81-55, AD-A113 465). Brooks AFB, TX: Manpower Personnel Division, Air Force Human Resources Laboratory.

Ree, M. J., Welsh, J. R., Earles, J. A., & Curran, L. T. (in press). Equating and implementation of Armed Services Vocational Aptitude Battery (ASVAB): Forms 15, 16, and 17 in the 1980 youth population metric. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Rosenthal, D. B., & Lawrence, J. H. (1988). Job characteristics and military attrition (FR-PRD-88-11). Alexandria, VA: Human Resources Research Organization.

Rosenthal, R., & Rubin, D.B. (1982). Comparing effect sizes of independent studies. Psychological Bulletin, 92, 500-504.

Rossmeissl, P. G., & Brandt, D.A. (1984). Subgroup variation in the validity of Army aptitude area composites. In N. K. Eaton, M. H. Goer, H. H. Harris, & L. M. Zook (Eds.), Improving selection and classification and utilization of Army enlisted personnel: Annual report 1984 Fical Year (ARI-TR-660) (pp. 361-413). Alexandria, VA: Army Research Institute for the Behavioral Sciences.

Rossmeissl, P. G., & Stern, B. M. (1983). The application of meta-analytic techniques in estimating selection/classification parameters. In N. K. Eaton, M. H. Goer, J. H. Harris & L. M. Zook (Eds.), Improving the selection, classification, and utilization of Army enlisted personnel: Annual Report, 1984 Fiscal Year (ARI-TR-660) (pp. 423-430). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62(5), 529-540.

Schmidt, F. L., Hunter, J.E., & Dunn, W.L. (1987). Potential increases from adding new tests to the Armed Services Vocational Aptitude Battery (NPRDC-DO-0053). San Diego, CA: Navy Personnel Research and Development Center.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red-herring. Journal of Applied Psychology, 66, 166-185.

Schmidt, F. L., Hunter, J. E., & Raju, N. S. (1988). Validity generalization and situational specificity: A second look at the 75% rule and Fisher's z transformation. Journal of Applied Psychology, 73(4), 665-672.

Shields, J. L., & Grafton, F. C. (1983). A natural experiment: Analysis of an almost unselected Army population. Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

Sims, W. H. (1978). An application of factor analysis to the construction of improved classification composites from the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 (CNA 78-3094). Arlington, VA: Center for Naval Analyses.

Sims, W. H., & Hiatt, C. M. (1981). Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 with applications to ASVAB Forms 8, 9, and 10 (CNSA 1160). Alexandria, VA: Center for Naval Analyses.

Sims, W. H., & Hiatt, C. M. (1983). A joint factor analysis of ASVAB Forms 5/6/7 and Forms 8/9/10 (CNA 83-3178/10). Alexandria, VA: Center for Naval Analyses.

Sims, W. H., & Mifflin, T. L. (1978). A factor analysis of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 (CNA 78-3092). Arlington, VA: Center for Naval Analyses.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Society for Industrial and Organizational Psychology, Inc. (1987). Principles for the validation and use of personnel selection procedures. (3rd ed.) College Park, MD: Author.

Stermer, N. S. (1988). Meta-Analysis of Armed Services Vocational Aptitude Battery composite validity data. Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Sticht, T. G., Hooke, L. R., & Caylor, J. S. (1982). Literacy, oracy, and vocational aptitude as predictors of attrition and promotion in the armed services (HumRRO-PP-2-82). Alexandria, VA: Human Resources Research Organization.

Stoker, P., Hunter, D. R., Batchelor, C. L., & Curran, L. T. (1987). Air traffic controller trainee selection (AFHRL TR-87-19, AD-187 497). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Stoloff, P. H. (1983). A factor analysis of ASVAB Form 8A in the 1980 DoD reference population (CNA 83-3135/25). Alexandria, VA: Center for Naval Analyses.

Thorndike, R.L. (1957). The optimum test composite to predict a set of criteria (Research Report AFPTRC-TN-75-103). San Antonio, TX: Air Force Personnel and Training Research Center.

Thorndike, R. L. (1985). The central role of general ability in prediction. Multivariate Behavioral Research, 20, 241-254.

Valentine, L. D., Jr. (1977). Prediction of Air Force technical training success from ASVAB and educational background (AFHRL-TR-77-18, AD-A041 735). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Valentine, L. D., Jr. & Mathews, J. J. (1977). Validity of high school composites from Armed Services Vocational Aptitude Battery (ASVAB) (Forms 6 and 7) for Air Force technical training (Staff Report). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Vineberg, R., Sticht, T. G., Taylor, E. N., & Caylor, J. H. (1971). Effects of aptitude (AFQT), job experience, and literacy on job performance: Summary of HumRRO work units utility and REALISTIC (TR-71-1). Alexandria, VA: Human Resources Research Organization.

Vineberg, R., & Taylor, E. N. (1972). Performance in four Army jobs by men at different aptitude (AFQT) levels: Relationships between performance criteria (TR-72-23). Alexandria, VA: Human Resources Research Organization.

Vineberg, R., Taylor, E. N., & Caylor, J. S. (1970). Performance in five Army jobs by men at different aptitude (AFQT) levels: I. Purpose and design of study (TR-70-18). Monterey, CA: Human Resources Research Organization.

Vineberg, R., Taylor, E. N., & Sticht, T. G. (1970). Performance in five Army jobs at different aptitude (AFQT) levels: II. Development and description of instruments (HumRRO-TR-70-20, AD-715-641). Alexandria, VA: Human Resources Research Organization.

Vitola, B. M., & Alley, W. E. (1968). Development and standardization of Air Force composites for the Armed Services Vocational Aptitude Battery (AFHRL-TR-68-110, AD-688 222). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Vitola, B. M., Guinn, N., & Wilbourn, J. M. (1977). Impact of various enlistment standards on the procurement-training system (AFHRL-TR-77-16, AD-040 752). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Vitola, B. M., Mullins, C. J., & Croll, P. R. (1973). Validity of the Armed Services Vocational Aptitude Battery, Form 1, to predict technical school success (AFHRL-TR-73-7, AD-767 578). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Wagner, M. P., Dirmeyer, R. P., Means, B., & Davidson, M. K. (1982). Analysis of aptitude, training, and job performance measures. Washington, DC: Department of Defense.

Wainer, H., & Braun, H. I. (1988). Test validity. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Wardlaw, W. E. (1983). Enlisted performance standards model for Operations Specialist rating. Monterey, CA: Naval Postgraduate School.

Waters, B. K., Barnes, J. D., Foley, P. P., Steinhaus, S. D., & Brown, D. C. (1988). Estimating the reading skills of military applicants: Development of an ASVAB to RGL conversion table (HumRRO-FR-PRD-88-22). Alexandria, VA: Human Resources Research Organization.

Waters, B. K., Lawrence, J. H., & Camara, W. J. (1987). Personnel enlistment and classification procedures in the military. Washington, DC: National Academy Press.

Weeks, J. L., Mullins, C. J., & Vitola, B. M. (1975). Airman classification batteries from 1948 to 1975: A review and evaluation (AFHRL-TR-75-78, AD-A026 470). Lackland AFB, TX: Air Personnel Research Division, Air Force Human Resources Laboratory.

Wegner, T. G., & Ree, M. J. (1985). Armed Services Vocational Aptitude Battery: Correcting the speeded subtests for the 1980 youth population (AFHRL-TR-85-14, AD-A158 823). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Wegner, T. G.,& Ree, M. J. (1986). Alternative Armed Forces Qualification Test composites (AFHRL-TP-86-27, AD-A173 027). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Weisen, J. P., & Siegel, A. I. (1977). Development and evaluation of a method for approximating ASVAB validity data. Unpublished manuscript. Wayne PA: Applied Psychological Services, Inc.

Weltin, M. M., & Popelka, B. A. (1983). Evaluation of the ASVAB 8/9/10 clerical composite for predicting training school performance (ARI-TR-594). Alexandria, VA: Army Research Institute.

Wherry, R.J., & Gaylord, R.H. (1943). The concept of test and item reliablity in relation to factor pattern. Psychometrika, 8, 247-264.

Wilbourn, J. M., & Guinn, N. (1973). Feasibility of using special measures in the classification and assignment of lower mental ability airmen (AFHRL-TR-73-31, AD-777 831). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Wilbourn, J. M., Guinn, N., & Leisey, S. A. (1976). Validation of non-verbal measures for selection and classification of enlisted personnel (AFHRL-TR-76-72, AD-A037 589). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Wilbourn, J. M., Valentine, L. D., Jr., & Ree, M. J. (1984). Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to the Air Force technical school final grades (AFHRL-TP-84-8, AD-A144 213). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Wiley, L. N., & Hahn, C. P. (1977). Task level job performance criteria dvelopment (AFHRL TR 77-75, AD-A055 694). Brooks AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory.

Wilfong, H. D. (1980). Technical Supplement to the High School Counselor's Guide. Ft. Sheridan, IL: Military Entrance Processing Command.

Wothke, W., Bock, R. D., Curran, L.T., Fairbank, B. A., Augustine, J. W., Gillet, A. H., & Guerro, C. (In press). A factor analytic examination of the Armed Services Vocational Aptitude Battery (ASVAB) and the Kit of Factor Referenced tests. (AFHRL-TR-In Press). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

APPENDIX A: SUBTESTS AND COMPOSITES DEFINITIONS FOR ASVAB FORMS 1, 2, 5, 6, 7, AND 8 THROUGH 17

Table A-1. Content of ASVAB Forms 6 and 7

Subtest (ASVAB Order)	Description	Number of items	Test time (mins)
General Information (GI)	General knowledge test, primarily on sports, outdoor activities, automobile mechanics and history	15	7
Numerical Operations (NO)	Speeded mathematical test requiring elementary addition, subtraction, multiplication, and division	50	3
Attention to Detail (AD)	Speeded test in which the examinee counts the number of Cs embedded in lines of Os	30	5
Word Knowledge (WK)	Vocabulary test	30	10
Arithmetic Reasoning (AR)	Arithmetic test requiring examinees to solve word problems	20	10
Space Perception (SP)	Pictorial test requiring knowledge of algebra, geometry, fractions, decimals and exponents	20	12
Electronics Information(EI)	Test requiring knowledge of electrical and electronic components, principles and symbols	30	15

Table A-1. (Concluded)

Subtest (ASVAB Order)	Description	Number of items	Test time (mins)
Mechanical Comprehension (MC)	Drawings illustrating mechanical principles	20	15
General Science (GS)	Test measuring knowledge in the physical (N = 10) and biological (N = 10) sciences	20	8
Shop Information (SI)	Test measuring knowledge about the use of tools and practices	20	8
Automotive Information (AI)	Test on automobile parts, operations and functions	20	10
Total		275	103

Table A-2. Subtest Content of ASVAB Forms 8 Through 17

Subtest (ASVAB Order)	Description	Number of items	Test time (mins)
General Science (GS)	Knowledge of the physical and biological sciences	25	11
Arithmetic Reasoning (AR)	Word problems emphasizing mathematical reasoning rather than mathematical knowledge	30	36
Word Knowledge (WK)	Understanding the meaning of words; i.e. vocabulary	35	11
Paragraph Comprehension (PC)	Presentation of short paragraphs followed by one or more multiple-choice items	15	13
Numerical Operations (NO)[a]	A speeded test of four arithmetic operations; i.e. addition, subtraction, multiplication and division	50	3
Coding Speed (CS)[a]	A speeded test of matching words and four-digit numbers	84	7
Auto & Shop Information (AS)	Knowledge of auto mechanics, shop practices and tool functions, in verbal and pictorial items	25	11
Mathematics Knowledge (MK)	Knowledge of algebra, geometry, and fractions	25	24

Table A-2. (Concluded)

Subtest (ASVAB Order)	Description	Number of items	Test Time (mins)
Mechanical Comprehension (MC)	Understanding mechanical principles such as gears, levers, pulleys and hydraulics in verbal and pictorial items	25	19
Electronics Information (EI)	Knowledge of electronics and radio principles in verbal and pictorial items	20	9
Total		334	144

[a] Speeded subtest.

Table A-3. ASVAB Forms 1 and 2 Composites

Composites	Subtests[a]
Air Force	
AFQT	WK + AR + SP + TK
Mechanical (M)	TK + MC + SI + AI
Administrative (A)	CS + WK
General (G)	2WK + AR
Electronics (E)	AR + SP + EI
High School[b]	
General Technical (GT)	AR + WK
Clerical (CL)	WK + 1/3CS
General Mechanical (GM)	SP + 2SI
Electronics (EL)	MC + 2EI
Motor Mechanical (MM)	MC + 2AI

[a] Definitions of abbreviations are found in Table 2.

[b] Adapted from Validation of ASVAB-2 Against Civilian Vocational-Technical High School Criteria (p. 32) by H.E. Jensen and L.D. Valentine, Jr., 1976, Brooks AFB, TX: Air Force Human Resources Laboratory.

Table A-4. ASVAB Forms 6 and 7 Selector Composites Used by Military Services

Composite	Abbreviation	Description
	All Services	
WK + AR + SP	AFQT	Armed Forces Qualification Test
WK + AR	GT	General Technical Composite
	Navy Selector Aptitude Index (SAI)	
WK + MC + SI	MC	Mechanical Composite
AR + MK + GS + EI	EL	Electronics Composite
WK + AD + NO	CL	Clerical Composite
AR + NC		Selector for Torpedomen's Mate School
WK + MC		Selector for Aviation Structural Mechanic School
WK + MC + MK + EI + GS		Selector for Ocean Systems Technician School
AR + SI		Selector for Quartermaster School
WK + AR + NO + AD		Selector for Communications Technician (Interpreter) School
AR + 2MK + GS		Selector for Basic Electricity and Electronics School

Table A-4. (Concluded)

Composite	Abbreviation	Description
	Air Force	
MC + SP + AI	M	Mechanical Composite
AR + SP + EI	E	Electronics Composite
	Army	
AR + MC + AI + GS	GM	Similar to General Maintenance Composite, except that the Army uses only the 10 biological items in the GS test
AR + EI + MC + SI[a]	EL	Electronics Composite
AS + WK + AR	CL	Clerical Composite
MK + EI + SI + AI[a]	MM	Mechanical Maintenance
WK + AR + SP + MC	SC	Surveillance and Communications Composite
AD + AR + SP + SI[a]	CO	Combat Composite
GI + AR + MK + EI[a]	FA	Field Artillery Composite
GI + AR[a]	OF	Operators and Food Composite
AR + MK + GS	ST	Similar to Skilled Technician Composite, except that the Army uses only 10 biological items in the GS test

[a] These composites use a scale from the Classification Inventory (CI), which is actually subtest 13 of the ASVAB.

Table A-5. Titles and Abbreviations of Selector Composites by Service for Forms 8, 9, and 10

Title	Abbreviation	Title	Abbreviation
Army		Navy Selector Apitude Index	(SAI)
Electronics	(EL)	General Technical	(GT)
Operators/Foods	(OF)	Mechanical	(MECH)
Surveillance/ Communications	(SC)	Electronics	(ELEC)
Mechanical Maintenance	(MM)	Clerical	(CLER)
Clerical	(CL)	Aviation Structural Mechanical	(AM)
Skilled Technical	(ST)	Basic Electricity/ Electronics	(BE/E)
Combat	(CO)	Boiler Technician/ Engineman/Machinists Mate	(BT/EN/MM)
Field Artillery	(FA)		
General Technical	(GT)	Machinery Repairman	(MR)
General Maintenance	(GM)	Submarine	(SUB)
		Communications Technician	(CT)
		Hospitalman	(HM)
Marine Corps		Air Force	
Combat	(CO)	Mechanical	(M)
Field Artillery	(FA)	Administrative	(A)
Clerical	(CL)	General	(G)
Electronics Repair	(EL)	Electronics	(E)
Mechanical Maintenance	(MM)		
General Technical	(GT)		

Note. From Armed Services Vocational Aptitude Battery (ASVAB) Test Manual (p. 20) by the Department of Defense, 1984, North Chicago, IL: United States Military Entrance Processing Command.

APPENDIX B: RELIABILITY ESTIMATES FOR ASVAB FORMS 1 THROUGH 17 SUBTESTS AND COMPOSITES

Table B-1. Reliability Estimates for ASVAB Forms 1, 2, and 3, Subtests and Composites

	Internal Consistency		
	Form 1	Form 2	Form 3
	Subtests		
Word Knowledge (WK)	.87	.87	.86
Arithmetic Reasoning (AR)	.87	.87	.86
Space Perception (SP)	.84	.84	.83
Mechanical Comprehension (MC)	.83	.83	.83
Shop Information (SI)	.81	.81	.82
Automotive Information (AI)	.85	.85	.88
Electronic Information (EI)	.83	.83	.82
Tool Knowledge (TK)	.70	.78	.80
	Composites[a]		
AFQT	.92		
Mechanical (M)	.91		
Administrative (A)	.88		
General (G)	.90		
Electronics (E)	.92		

Note. From Technical Supplement to the high school counselor's guide (p. 54, 58) by Harry Wilfong, 1980, Ft. Sheridan, IL: U.S. Military Entrance Processing Command.

[a] Estimated by Wherry and Gaylord (1943)

Table B-2. Reliablity Coefficients for the Aptitude Indices of ASVAB Form 1

Aptitude Index	Reliability
Mechanical	.84
Administrative	.91
General	.86
Electronics	.91

Note. From Airman classification batteries from 1948 to 1975: A review and evaluation (p. 44), J. L. Weeks, C. J. Mullins, and B. M. Vitola, 1975, Lackland AFB, TX: Air Force Human Resources Laboratory.

Table B-3. Alternate Forms Reliability Estimates For ASVABS Forms 5, 6, and 7 Subtests and Composites

	Form 5	Form 6	Form 7
	Subtests[a]		
General Information	.67	.75	.73
Numerical Operations	--[b]	.88	.86
Attention to Detail	--[b]	.82	.78
Word Knowledge	.91	.91	.91
Arithmetic Reasoning	.82	.86	.84
Space Perception	.82	.77	.80
Mathematics Knowledge	.88	.85	.80
Electronics Information	.87	.87	.84
Mechanical Comprehension	.81	.81	.79
General Science (Biology)	.63	.73	.70
General Science (Physics)	.64	.59	.66
General Science (Total)	.77	.79	.81
Shop Information	.83	.85	.82
Automotive Information	.84	.84	.86
	Composites[c]		
General	.93	.94	.93
Mechanical	.93	.93	.93
Administrative	--[a]	.94	.92
Electronics	.93	.93	.92
AFQT	.93	.94	.94

Note. From Armed Services Vocational Aptitude Battery development (ASVAB Forms 5, 6, and 7) (p. 10) by H. E. Jensen, I. H. Massey and L. D. Valentine, Jr., 1976, Lackland AFB, TX: Air Force Human Resources Laboratory.

[a] Not defined for speeded tests.

[b] Subtest reliabilities were derived using Kuder-Richardson Formula 20.

[c] Estimated from Wherry and Gaylord (1943).

Table B-4. Alternate Forms Reliability Coefficients (r) of Subtests[a] and Composites of ASVAB Form 11a with Forms 9a and 9b

Subtests[bc]	r(9a)	r(9b)	Composites[b]	r(9a)	r(9b)
GS	.84	.83	MECH[d]	.92	.92
AR	.88	.88	ADM[d]	.86	.88
WK	.89	.87	GEN[d]	.93	.93
PC	.72	.68	ELEC[d]	.93	.93
NO	.68	.70	AFQT[c]	.92	.93
CS	.75	.75			
AS	.85	.85			
MK	.86	.85			
MC	.78	.76			
EI	.72	.71			

Note. From Armed Services Vocational Aptitude Battery: Alternate forms reliability (Forms 8, 9, 10, and 11) (p. 11) by P. Palmer, D. D. Hartke, M. J. Ree, J. R. Welsh, and L. D. Valentine, Jr.,1988, Brooks AFB, TX: Air Force Human Resources Laboratory.

[a] The estimates of the reliability coefficients are correlations with Ns ranging from 4,512 to 747 in Form 9a and from 4,011 to 648 in Form 9b.

[b] Definitions for abbreviations are found in Table A-2 for the subtests and Table A-5 for the composites.

[c]Raw scores used to estimate r.

[d]Standard scores used to estimate r.

Table B-5. Alternate Forms Reliability Coefficients (r) of Subtests and Composites of ASVAB Form 11a with Forms 10a and 10b

Subtests[a]	r(10a)	r(10b)	Composites[a]	r(10a)	r(10b)
GS	.84	.83	MECH	.92	.91
AR	.87	.87	ADM	.86	.87
WK	.89	.88	GEN	.93	.93
PC	.75	.69	ELEC	.93	.94
NO	.69	.71	AFQT	.93	.93
CS	.72	.74			
AS	.83	.84			
MK	.84	.85			
MC	.79	.77			
EI	.72	.72			

Note. From Armed Services Vocational Aptitude Battery: Alternate forms reliability (Forms 8, 9, 10, and 11) (p. 11) by P. Palmer, D. D. Hartke, M. J. Ree, J. R. Welsh, and L. D. Valentine, Jr., 1988, Brooks AFB, Texas: Air Force Human Resources Laboratory.

[a] Definitions of abbreviations are found in Table A-2 for the subtests and Table A-5 for the composites.

Table B-6. Subtest Reliabilities[a] for ASVAB Forms 15, 16, and 17

	Form						
Subtest[b]	15a	15b	15c	16a	16b	17a	17b
GS	.751	.742	.729	.771	.769	.721	.735
AR	.842	.841	.843	.820	.847	.848	.837
WK	.830	.817	.809	.801	.809	.834	.826
PC	.682	.685	.634	.696	.689	.663	.667
NO[c]							
CS[c]							
AS	.824	.823	.799	.865	.864	.838	.837
MK	.846	.838	.828	.840	.843	.836	.827
MC	.780	.783	.798	.763	.772	.746	.745
EI	.699	.699	.700	.724	.731	.768	.742
N	2,774	2,752	2,504	2,678	2,712	2,501	2,540

Note. From Equating and implementation of Armed Services Vocational Aptitude Battery (ASVAB): Forms 15, 16, and 17 in the 1980 youth population metric (Appendix B) by M. J. Ree, J. R. Welsh, J. A. Earles and L. T. Curran, in press, Brooks AFB, TX: Air Force Human Resources Laboratory.

[a] KR-20 values.

[b] Definitions of abbreviations are found in Table A-2.

[c] Not estimated for speeded tests.

APPENDIX C: EFFECT SIZES, BY STUDIES WITH VALIDITY DATA USED FOR EFFECT SIZE SUMMARIES, ALL FORMS

Table C-1. Summary Validity Statistics for AQE-64 against Final School Grade

Composite[a]	Mean r[b]	Standard deviation[b]	Total N	Number r's	BESD range	
M	.181	.049	4,907	3	.409	.591
A	.069	.045	2,664	2	.466	.534
G	.129	.026	6,744	3	.436	.564
E	.162	.028	6,230	3	.419	.581

Note. Data are from Cultural subgroup differences in the relationships between Air Force aptitude composites and training criteria (AFHRL-TR-70-35, AD-715-922) by N. Guinn, E. C. Tupes, and W. E. Alley, 1970a, Lackland AFB, TX: Air Force Human Resources Laboratory.

[a] Definitions of abbreviations are found in Table A-2.

[b] Weighted by study sample size.

Table C-2. Summary Validity Statistics for ASVAB Form 3 against Final School Grade

Composite[a]	Mean r[b]	Standard deviation[b]	Total N	Number r's	BESD range	
M (Blacks)	.131	.098	10,511	39	.434	.566
M (Whites)	.270	.110	32,488	43	.365	.635
M (Others)	.277	.180	655	11	.361	.639
M (Females)	.178	.107	8,373	31	.411	.589
M (Males)	.331	.098	35,570	43	.334	.666
M (Total)	.290	.102	43,985	43	.355	.645
A (Blacks)	.098	.082	10,511	39	.451	.549
A (Whites)	.201	.058	32,488	43	.400	.600
A (Others)	.212	.157	655	11	.394	.606
A (Females)	.222	.080	8,373	31	.389	.611
A (Males)	.242	.077	35,570	43	.379	.621
A (Total)	.217	.064	43,985	43	.392	.608
G (Blacks)	.173	.110	10,511	39	.414	.586
G (Whites)	.325	.085	32,488	43	.338	.662
G (Others)	.272	.153	655	11	.364	.636
G (Females)	.338	.095	8,373	31	.331	.669
G (Males)	.328	.097	35,570	43	.336	.664
G (Total)	.317	.085	43,985	43	.341	.659
E (Blacks)	.193	.084	10,511	39	.403	.597
E (Whites)	.344	.088	32,488	43	.328	.672
E (Others)	.342	.152	655	11	.329	.671
E (Females)	.293	.092	8,373	31	.353	.647
E (Males)	.365	.087	35,570	43	.318	.682
E (Total)	.347	.082	43,985	43	.327	.673
AFQT (Blacks)	.170	.080	10,511	39	.415	.585
AFQT (Whites)	.333	.068	32,488	43	.333	.667
AFQT (Others)	.357	.143	655	11	.322	.678
AFQT (Females)	.336	.089	8,373	31	.332	.668
AFQT (Males)	.349	.079	35,570	43	.325	.675
AFQT (Total)	.341	.070	43,985	43	.329	.671

Note. Data from Prediction of Air Force technical training success from ASVAB and educational background (AFHRL-TR-77-18, AD-A041 735) by L. D. Valentine, Jr., 1977, Brooks AFB, TX: Air Force Human Resources Laboratory.

[a]Definitions of abbreviations are found in Table A-5.

[b]Weighted by study sample size.

Table C-3. Summary Validity Statistics for ASVAB Forms 6 and 7 Against Job Performance Measures

Composite[a]	Mean r[b]	Standard deviation[b]	Total N	Number r's	BESD range	
M	.032	.033	1,343	3	.484	.516
M	.011	.063	1,343	3	.494	.506
M	.104	.097	1,343	3	.448	.552
M	.012	.057	1,343	3	.494	.506
M	.026	.033	1,343	3	.487	.513
M	.016	.026	1,343	3	.492	.508
A	-.007	.025	1,343	3	.504	.496
A	-.021	.039	1,343	3	.510	.490
A	.006	.026	1,343	3	.497	.503
A	-.013	.042	1,343	3	.507	.493
A	-.018	.039	1,343	3	.509	.491
A	-.003	.028	1,343	3	.502	.498
G	-.001	.019	1,343	3	.500	.500
G	-.010	.067	1,343	3	.505	.495
E	.000	.050	1,343	3	.500	.500
E	-.024	.072	1,343	3	.512	.488
E	.011	.067	1,343	3	.494	.506
E	.008	.063	1,343	3	.496	.504
E	.002	.071	1,343	3	.499	.501
E	.021	.044	1,343	3	.489	.511
E	.018	.023	1,343	3	.491	.509
E	-.003	.051	1,343	3	.502	.498
E	-.007	.047	1,343	3	.504	.496
E	-.002	.024	1,343	3	.501	.499

Note. Data are from Task level job performance criteria development (AFHRL-TR-77-75, AD-A055 694) by L. N. Wiley, and C. P. Hahn, 1977, Brooks AFB, TX : Air Force Human Resources Laboratory.

[a]Definitions of abbreviations are found in Table A-2.

[b]Weighted by study sample size.

Table C-4. Summary Validity Statistics for ASVAB Forms 5, 6, and 7 Against Final School Grade

Composite	Mean r[a]	Standard[a] deviation	Total N	Number r's	BESD range	
Electronics	.404	.092	1,559	10	.298	.702
Communications	.362	.087	1,559	10	.319	.681
General Technical	.455	.068	1,559	10	.272	.728
Motor Technical	.428	.078	1,559	10	.286	.711
General Maintenance	.326	.117	1,559	10	.337	.663
Clerical	.249	.099	1,559	10	.375	.625
Verbal	.414	.130	1,559	10	.293	.707
Analytical /Quantitative	.391	.082	1,559	10	.305	.695
Clerical	.062	.096	1,559	10	.469	.531
Mechanical	.278	.089	1,559	10	.361	.639
Trade Mechanical	.243	.124	1,559	10	.379	.621
Academic	.455	.068	1,559	10	.272	.728
HS (Electronics)	.346	.131	3,154	16	.327	.673
HS (Commun)	.302	.139	3,154	16	.349	.651
HS (Gen Tech)	.367	.109	3,154	16	.316	.684
HS (Motor Tech)	.348	.142	3,154	16	.326	.674
HS (Gen Maint)	.292	.123	3,154	16	.354	.646
HS (Clerical)	.245	.101	3,154	16	.377	.623
HS (Verbal)	.340	.119	3,154	16	.330	.670
HS (Anal/Quant)	.344	.117	3,154	16	.328	.672
HS (Clerical)	.117	.102	3,154	16	.441	.559
HS (Mech)	.241	.128	3,154	16	.380	.620
GS (Trade Mech)	.229	.123	3,154	16	.386	.614
HS (Academic)	.367	.109	3,154	16	.316	.684

Note. Data are from Validity of high school composites from Armed Services Vocational Aptitude Battery (ASVAB) (Forms 6 and 7) for Air Force technical training (Staff Report) by L. D. Valentine, Jr., and J. J. Mathews, 1977, Lackland AFB, TX: Air Force Human Resources Laboratory.

[a]Weighted by study sample size.

Table C-5. Summary Validity Statistics for ASVAB Forms 6 and 7 Against Final School Grade

Composite[a]	Mean r^b	Standard deviation[b]	Total N	Number r's	BESD range	
M	.283	.130	1,392	4	.359	.641
A	.228	.048	1,392	4	.386	.614
G	.378	.047	1,392	4	.311	.689
E	.363	.102	1,392	4	.319	.681
AFQT	.371	.051	1,392	4	.314	.686

Note. Data are from Calculation of predictor composites in the absence of a criterion (AFHRL-TR-79-53, AD-A080 921) by C. J. Mullins, J. A. Earles, and J. M. Wilbourn, 1979, Brooks AFB, TX: Air Force Human Resources Laboratory.

[a]Definitions of abbreviations are found in Table A-4.

[b]Weighted by study sample size.

Table C-6. Summary Validity Statistics for ASVAB Forms 5, 6, and 7 Against Final School Grade

Composite[a]	Mean r^b	Standard deviation[b]	Total N	Number r's	BESD range	
M (M Schools)	.341	.097	8,329	28	.330	.670
A (M Schools)	.264	.087	8,329	28	.368	.632
G (M Schools)	.411	.106	8,329	28	.294	.706
E (M Schools)	.399	.065	8,329	28	.301	.699
M (A Schools)	.254	.073	5,568	13	.373	.627
A (A Schools)	.241	.081	5,568	13	.379	.621
G (A Schools)	.430	.088	5,568	13	.285	.715
E (A Schools)	.337	.074	5,568	13	.331	.669
M (G Schools)	.271	.095	13,772	33	.365	.635
A (G Schools)	.236	.075	13,772	33	.382	.618
G (G Schools)	.389	.088	13,772	33	.305	.695
E (G Schools)	.315	.091	13,772	33	.343	.657
M (E Schools)	.305	.140	6,549	45	.348	.652
A (E Schools)	.256	.134	6,549	45	.372	.628
G (E Schools)	.394	.149	6,549	45	.303	.697
E (E Schools)	.399	.119	6,549	45	.300	.700

Note. Data are from Weighting of aptitude components based on differences in technical school difficulty (AFHRL-TR-81-19, AD-A102 045) by C. J. Mullins, J. A. Earles, and M. J. Ree, 1981, Brooks AFB, TX: Air Force Human Resources Laboratory.

[a]Definitions of abbreviations are found in Table A-5.

[b]Weighted by study sample size.

Table C-7. Summary Validity Statistics for ASVAB Forms 8, 9, and 10 Against Final Class Grade

Composite	Mean r[a]	Standard[a] deviation	Total N	Number r's	BESD range
AFQT	.420	.028	2,476,608	16	.290 .710

Note. Data are from Alternative Armed Forces Qualification Test composites (AFHRL-TP-86-27, AD-A173 027) by T. G. Wegner, and M. J. Ree, 1986, Brooks AFB, TX: Air Force Human Resources Laboratory.

[a]Weighted by study sample size.

Table C-8. Summary Validity Statistics for ASVAB Forms 8, 9, and 10 Against Job Performance Measures

Composite[a]	Mean r[b]	Standard[b] deviation	Total N	Number r's	BESD range
A	.505	.034	93,312	9	.248 .752
SAI	.421	.030	158,394	9	.289 .711
E	.467	.042	50,418	9	.266 .734
M (GM)[c]	.460	.034	23,139	9	.270 .730
M (MM)	.454	.047	63,657	9	.273 .727
SAI (OF)	.459	.040	78,336	9	.270 .730
SAI (SC)	.467	.038	33,561	9	.266 .734
SAI (ST)	.554	.029	63,549	9	.223 .777
A	.427	.028	47,448	9	.287 .713
SAI (CO)	.338	.018	25,911	9	.331 .669
E	.397	.023	23,490	9	.302 .698
SAI (FA)	.340	.034	15,831	9	.330 .670
M	.501	.040	17,496	9	.250 .750
M	.416	.041	48,834	9	.292 .708
SAI (OF)	.336	.024	41,634	9	.332 .668
SAI (SC)	.341	.014	13,167	9	.329 .671
SAI (ST)	.514	.030	28,629	9	.243 .757
A	.520	.035	72,054	9	.240 .760
SAI (CO)	.424	.031	143,730	9	.288 .712
E	.432	.039	53,640	9	.284 .716
SAI (FA)	.444	.040	62,676	9	.278 .722
M (GM)	.394	.029	11,736	9	.303 .697
M (MM)	.418	.048	38,781	9	.291 .709
SAI (OF)	.486	.046	42,516	9	.257 .743
SAI (SC)	.498	.045	41,441	9	.251 .749
SAI (ST)	.530	.031	62,235	9	.235 .765

Note. Data are from Validation of current alternative Armed Services Vocational Aptitude Battery (ASVAB) area composites, based on training and Skill Qualification Test (SQT) information in fiscal year 1981 and 1982 (ARI-TR-651, AD-A156 807) by D. H. McLaughlin, P. G. Rossmeissl, L. L. Wise, D. A. Brandt and M. Wang, 1984, Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

[a] Definitions of abbreviations are found in Table A-5.

[b] Weighted by study sample size.

[c] Definitions of abbreviations are found in Table A-3.

Table C-9. Summary Validity Statistics for ASVAB Forms 6 and 7 Against Final School Grade

Composite[a]	Mean r^b	Standard[b] deviation	Total N	Number r's	BESD range	
A	.304	.104	13,491	30	.348	.652
A	.334	.098	13,491	30	.333	.667
A	.315	.105	13,491	30	.342	.658
M	.199	.153	2,250	17	.401	.599
M	.337	.142	7,926	21	.331	.669
M	.293	.207	2,250	17	.354	.646
M	.213	.175	2,250	17	.393	.607
M	.350	.139	7,926	21	.325	.675
M	.470	.217	7,926	21	.265	.735
E	.300	.133	3,853	18	.350	.650
E	.312	.119	3,853	18	.344	.656
E	.393	.118	3,853	18	.304	.696

Note. Data are from An investigation of alternatives for setting second-to-third tour reenlistment standards (ARI-TR-690, AD-A164 694) by F. C. Grafton and D. K. Horne, 1985, Alexandria, VA: Army Research Institute.

[a] Definitions of abbreviations are found in Table A-4.

[b] Weighted by study sample size

Table C-10. Summary Validity Statistics for ASVAB Forms 6 and 7 Against Job Performance Measures

Composite	Mean r^a	Standard[a] deviation	Total N	Number r's	BESD range	
Hands-on	.310	.379	256	4	.345	.655
Hands-on	.375	.355	440	5	.312	.688

Note. Data are from A report to the House Committee on Appropriations: Fourth annual report to Congress on Joint-Service efforts to link enlistment standards to job performance, by Department of Defense, 1985, Washington, DC: Office of the Assistant Secretary of Defense.

[a] Weighted by study sample size.

Table C-11. Summary Validity Statistics for Forms 6 and 7 Against Job Performance Measures

Composite	Mean r^a	Standard[a] deviation	Total N	Number r's	BESD range	
AFQT	.364	.065	3,846	9	.318	.682
AFQT	.296	.088	3,368	8	.352	.648
AFQT	.108	.065	3,846	9	.446	.554
AFQT	.066	.066	3,846	9	.467	.533
AFQT	-.044	.045	3,846	9	.522	.478

Note. Data are from A report to the House Committee on Appropriations: Fifth annual report to Congress on Joint-Service efforts to link enlistment standards to job performance, by Department of Defense, 1986, Washington, DC: Office of the Assistant Secretary of Defense.

[a]Weighted by study sample size.

Table C-12. Summary Validity Statistics for ASVAB Forms 8 through 13 Against Job Performance Measures

Composite	Mean r^a	Standard[a] deviation	Total N	Number r's	BESD range	
AFQT	.401	.065	3,701	9	.300	.700
AFQT	.336	.092	3,223	8	.332	.668
AFQT	.109	.063	3,701	9	.446	.554
AFQT	.080	.057	3,701	9	.460	.540
AFQT	-.077	.051	3,701	9	.538	.462

Note. Data are from Joint-Service efforts to link enlistment standards to job performance: Recruit quality and readiness, A report to the House Committee on Appropriations Department of Defense, 1989, Washington, DC: Office of the Assistant Secretary of Defense.

[a]Weighted by study sample size.

Table C-13. Summary Validity Statistics for ASVAB Form 3 Synthetic Validity

Composite	Mean r[a]	Standard[a] deviation	Total N	Number r's	BESD range	
Subtest (AI)[b]	.380	.181	2,975	9	.310	.690
Subtest (AR)	.585	.148	2,975	9	.208	.792
Subtest (CS)	.173	.068	1,340	5	.413	.587
Subtest (EI)	.510	.176	2,975	9	.245	.755
Subtest (MC)	.482	.130	2,975	9	.259	.741
Subtest (SI)	.411	.152	2,975	9	.295	.705
Subtest (SP)	.436	.102	2,975	9	.282	.718
Subtest (TK)	.274	.147	2,808	8	.363	.637
Subtest (WK)	.577	.178	2,975	9	.212	.788

Note. Data are from Development and evaluation of a method for approximating ASVAB validity data (Unpublished manuscript) by J. P. Weisen and A. I. Siegel, 1977, Wayne, PA: Applied Psychological Services, Inc.

[a]Weighted by study sample size.

[b]Definitions of abbreviations are found in Table 1.

Table C-14. Summary Validity Statistics for ASVAB Forms 11, 12, and 13 Against Final School Grade

Composite[c]	Mean[ab] r	Standard[b] deviation	Total N	Number r's	BESD range	
M Subtest (GS)	.795	.205	7,934	10	.103	.897
M Subtest (AR)	.741	.128	7,934	10	.129	.871
M Subtest (WK)	.745	.143	7,934	10	.128	.872
M Subtest (PC)	.771	.243	7,934	10	.114	.886
M Subtest (NO)	.607	.130	7,934	10	.196	.804
M Subtest (CS)	.526	.099	7,934	10	.237	.763
M Subtest (AS)	.676	.175	7,934	10	.162	.838
M Subtest (MK)	.714	.113	7,934	10	.143	.857
M Subtest (MC)	.746	.180	7,934	10	.127	.873
M Subtest (EI)	.806	.217	7,934	10	.097	.903
A Subtest (GS)	.686	.137	3,170	7	.157	.843
A Subtest (AR)	.698	.158	3,170	7	.151	.849
A Subtest (WK)	.689	.152	3,170	7	.156	.844
A Subtest (PC)	.724	.168	3,170	7	.138	.862
A Subtest (NO)	.590	.202	3,170	7	.205	.795
A Subtest (CS)	.534	.150	3,170	7	.233	.767
A Subtest (AS)	.460	.060	3,170	7	.270	.730
A Subtest (MK)	.691	.123	3,170	7	.154	.846
A Subtest (MC)	.573	.074	3,170	7	.213	.787
A Subtest (EI)	.625	.117	3,170	7	.187	.813
G Subtest (GS)	.750	.139	9,997	10	.125	.875
G Subtest (AR)	.708	.087	9,997	10	.146	.854
G Subtest (WK)	.752	.167	9,997	10	.124	.876
G Subtest (PC)	.791	.244	9,997	10	.104	.896
G Subtest (NO)	.615	.067	9,997	10	.192	.808
G Subtest (CS)	.537	.084	9,997	10	.231	.769
G Subtest (AS)	.526	.074	9,997	10	.237	.763
G Subtest (MK)	.698	.100	9,997	10	.151	.849
G Subtest (MC)	.616	.070	9,997	10	.192	.808
G Subtest (EI)	.699	.086	9,997	10	.150	.850
E Subtest (GS)	.833	.214	3,351	10	.084	.916
E Subtest (AR)	.833	.147	3,351	10	.084	.916
E Subtest (WK)	.779	.169	3,351	10	.111	.889
E Subtest (PC)	.805	.153	3,351	10	.098	.902
E Subtest (NO)	.672	.087	3,351	10	.164	.836
E Subtest (CS)	.597	.076	3,351	10	.202	.798
E Subtest (AS)	.635	.094	3,351	10	.182	.818
E Subtest (MK)	.836	.135	3,351	10	.082	.918
E Subtest (MC)	.772	.120	3,351	10	.114	.886
E Subtest (EI)	.856	.233	3,351	10	.072	.928

Note. Data are from Investigation of the efficacy of general ability versus specific abilities as predictors of occupational success by G. E. Jones, 1988, Unpublished master's thesis, St. Mary's University, San Antonio, TX.

[a]Corrected for restriction range.

[b]Weighted by study sample size.

[c]Definitions of abbreviations are found in Table A-2.

Table C-15. Summary Validity Statistics for ASVAB Forms 8, 9, and 10 Against Attrition

Composite	Mean r[a]	Standard[a] deviation	Total N	Number r's	BESD range
AFQT	.030	.011	41,146	8	.485 .515
AFQT	.149	.029	40,512	7	.425 .575
AFQT	.469	.034	41,146	8	.265 .735
AFQT	.414	.039	41,146	8	.293 .707
AFQT	.033	.028	41,146	8	.483 .517
AFQT	.168	.033	40,512	7	.416 .584
AFQT	.444	.034	41,146	8	.278 .722
AFQT	.409	.036	41,146	8	.296 .704
AFQT	.251	.135	41,146	8	.374 .62
AFQT	.094	.040	8,170	3	.453 .547
AFQT	-.318	.131	3,608	3	.659 .341
AFQT	.419	.117	27,590	3	.291 .70
AFQT	.264	.134	41,146	8	.368 .632
AFQT	.143	.029	8,170	3	.429 .571
AFQT	-.321	.104	3,608	3	.661 .339
AFQT	.327	.080	27,590	3	.336 .664

Note. Data are from Attrition of nonprior-service reservists in the Army National Guard and Army Reserve (R-3267-RA, AD-A161 639) by D. W. Grissmer and S. N. Kirby, 1985, Santa Monica, CA: The Rand Corporation.

[a]Weighted by study sample size.

Table C-16. Summary Validity Statistics for ASVAB Forms 6 and 7 Against Final School Grade

Composite[a]	Mean r^b	Standard[b] deviation	Total N	Number r's	BESD range	
AFQT (MC/Army)	.397	.158	14,115	38	.301	.699
AFQT (Air Force)	.336	.147	14,115	38	.332	.668
AFQT (Navy)	.380	.146	14,115	38	.310	.690
AFQT	.400	.157	14,115	38	.300	.700
Subtest (GI)	.216	.093	14,115	38	.392	.608
Subtest (NO)	.202	.093	14,115	38	.399	.601
Subtest (AD)	.094	.076	14,115	38	.453	.547
Subtest (WK)	.235	.135	14,115	38	.382	.618
Subtest (AR)	.289	.149	14,115	38	.356	.644
Subtest (SP)	.174	.081	14,115	38	.413	.587
Subtest (MK)	.316	.149	14,115	38	.342	.658
Subtest (EI)	.243	.090	14,115	38	.379	.621
Subtest (MC)	.266	.135	14,115	38	.367	.633
Subtest (GS)	.284	.126	14,115	38	.358	.642
Subtest (SI)	.205	.140	14,115	38	.397	.603
Subtest (AI)	.206	.132	14,115	38	.397	.603

Note. Data are from An application of factor analysis to the construction of improved classification composites from the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 (CNA-78-309) by W.H. Sims, 1978, Arlington, VA: Center for Naval Analyses.

[a]Definitions of abbreviations are found in Table A-2.

[b]Weighted by study sample size.

Table C-17. Summary Validity Statistics for ASVAB Forms 6 and 7 Against Final School Grade

Composite[a]	Mean r^b	Standard[b] deviation	Total N	Number r's	BESD range	
SAI (CO)	.335	.102	20,640	38	.333	.667
SAI (FA)	.420	.112	20,640	38	.290	.710
E (All)	.436	.131	20,640	38	.282	.718
SAI (OF)	.300	.108	20,640	38	.350	.650
M (All)	.411	.159	20,640	38	.295	.705
SAI (MM)	.350	.162	20,640	38	.325	.675
A (All)	.331	.106	20,640	38	.335	.665
SAI (ST)	.424	.132	20,640	38	.288	.712
G (All)	.363	.111	20,640	38	.319	.681
SAI (SC)	.393	.114	20,640	38	.303	.697
SAI (GCT)	.371	.107	20,640	38	.315	.685
E (Army)	.371	.136	20,640	38	.314	.686
M (AF)	.317	.180	20,640	38	.342	.658
A (AF)	.292	.095	20,640	38	.354	.646
G (AF)	.350	.114	20,640	38	.325	.675
E (AF)	.366	.118	20,640	38	.317	.683
M (Navy)	.359	.136	20,640	38	.320	.680
A (Navy)	.269	.090	20,640	38	.365	.635
G (Navy)	.368	.121	20,640	38	.316	.684
E (Navy)	.445	.128	20,640	38	.278	.722
Subtest (GI)	.251	.097	20,640	38	.374	.626
Subtest (NO)	.229	.096	20,640	38	.385	.615
Subtest (AD)	.099	.079	20,640	38	.451	.549
Subtest (WK)	.274	.091	20,640	38	.363	.637
Subtest (AR)	.327	.128	20,640	38	.337	.663
Subtest (SP)	.202	.073	20,640	38	.399	.601
Subtest (MK)	.359	.126	20,640	38	.321	.679
Subtest (EI)	.292	.103	20,640	38	.354	.646
Subtest (MC)	.309	.121	20,640	38	.346	.654
Subtest (GS)	.334	.089	20,640	38	.333	.667
Subtest (SI)	.241	.156	20,640	38	.380	.620
Subtest (AI)	.253	.153	20,640	38	.374	.626

Note. Data are from Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 with applications to Forms 8, 9, and 10 (CNSA-1160) by W. H. Sims, and C. M. Hiatt, 1981, Alexandria, VA: Center for Naval Analyses.

[a]Definitions of abbreviations are found in Table A-2 for the subtests and Table A-4 for the composites.

[b]Weighted by study sample size.

Table C-18. Summary Validity Statistics for Forms 8, 9, and 10 Against Final School Grade

Composite[a]	Mean r[b]	Standard[a] deviation	Total N	Number r's	BESD range	
SAI (CO)	.580	.213	14,692	34	.210	.790
SAI (FA)	.595	.233	14,692	34	.202	.798
A	.526	.180	14,692	34	.237	.763
E	.601	.242	14,692	34	.199	.801
M	.560	.211	14,692	34	.220	.780
G	.594	.236	14,692	34	.203	.797
AFQT	.548	.196	14,692	34	.226	.774
HS Verbal	.563	.197	14,692	34	.218	.782
HS Quant.	.577	.227	14,692	34	.211	.789
HS Tech.	.511	.189	14,692	34	.245	.755
HS Speed	.451	.140	14,692	34	.275	.725
HS Mech.	.560	.208	14,692	34	.220	.780
HS Off & Sup	.577	.219	14,692	34	.212	.788
HS Elect	.601	.237	14,692	34	.199	.801
HS Skill Serv	.600	.227	14,692	34	.200	.800

Note. Data are from Validity of ASVAB Forms 8, 9, and 10 for Marine Corps training courses: Subtests and current composites (Memorandum No. 83-3109) by M. H. Maier and A. R. Truss, 1983, Alexandria, VA: Center for Naval Analyses.

[a]Definitions of abbreviations are found in Table C-4.

[b]Weighted by study sample size.

Table C-19. Summary Validity Statistics for ASVAB Forms 8, 9, and 10 Against Final School Grade

Composite[a]	Mean r^{b}	Standard deviation[b]	Total N	Number r's	BESD range	
M (Total)	.432	.108	9,285	20	.284	.716
A (Total)	.216	.101	3,170	7	.392	.608
G (Total)	.446	.078	11,038	17	.277	.723
E (Total)	.483	.090	6,166	26	.258	.742
AFQT (Total)	.408	.097	29,619	70	.296	.704
M (Male)	.440	.107	8,609	20	.280	.720
A (Male)	.256	.086	2,221	7	.372	.628
G (Male)	.448	.077	9,806	17	.276	.724
E (Male)	.486	.085	5,623	26	.257	.743
AFQT (Male)	.419	.096	26,259	70	.291	.709
M (Female)	.251	.148	534	7	.375	.625
A (Female)	.111	.144	949	7	.445	.555
G (Female)	.419	.100	1,133	13	.290	.710
E (Female)	.330	.226	309	9	.335	.665
AFQT (Female)	.374	.121	2,925	36	.313	.687
M (White)	.437	.123	7,986	20	.282	.718
A (White)	.255	.104	2,076	7	.373	.627
G (White)	.450	.066	8,618	17	.275	.725
E (White)	.483	.092	5,575	26	.259	.741
AFQT (White)	.407	.096	24,256	70	.297	.703
M (Black)	.225	.114	917	10	.387	.613
A (Black)	.070	.119	977	7	.465	.535
G (Black)	.292	.099	2,007	11	.354	.646
E (Black)	.383	.124	217	6	.309	.691
AFQT (Black)	.288	.140	4,118	34	.356	.644

Note. Data are from Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to the Air Force technical school final grades (AFHRL-TP-84-8, AD-A114 213) by J. M. Wilbourn, L. D. Valentine, Jr., and M. J. Ree, 1984, Brooks AFB, TX: Air Force Human Resources Laboratory.

[a]Definitions of abbreviations are found in Table A-4.

[b]Weighted by sample size.

Table C-20. Summary Validity Statistics for ASVAB Forms 8, 9, and 10 Against Time to Completion

Composite[a]	Mean r^b	Standard[b] deviation	Total N	Number r's	BESD range	
E	-.388	.143	5,572	4	.694	.306
SAI	-.281	.035	5,941	3	.641	.359
Subtest (GS)	-.173	.054	11,513	7	.587	.413
Subtest (AR)	-.284	.086	11,513	7	.642	.358
Subtest (WK)	-.173	.064	11,513	7	.587	.413
Subtest (PC)	-.170	.060	11,513	7	.585	.415
Subtest (NO)	-.184	.042	11,513	7	.592	.408
Subtest (CS)	-.211	.058	11,513	7	.606	.394
Subtest (AS)	-.179	.053	11,513	7	.589	.411
Subtest (MK)	-.289	.136	11,513	7	.645	.355
Subtest (MC)	-.226	.066	11,513	7	.613	.387
Subtest (EI)	-.212	.058	11,513	7	.606	.394

Note. Data are from An empirical comparison of the accuracy of univariate and multivariate corrections for range restrictions (NPRDC-TR-85-19, AD-A153 0171) by S. Booth-Kewley, 1985, San Diego, CA: Navy Personnel Research and Development Center.

[a]Definitions of abbreviations are found in Table A-2.

[b]Weighted by study sample size.

Table C-22. Summary Validity Statistics for ASVAB Forms 6 and 7 Against Time-to-Completion

Composite[a]	Mean r[b]	Standard[b] deviation	Total N	Number r's	BESD range	
M	-.143	.043	7,754	9	.571	.429
A	-.219	.048	7,754	9	.610	.390
G	-.181	.061	7,754	9	.591	.409
E	-.293	.095	7,754	9	.646	.354
SAI	-.244	.079	7,754	9	.622	.378
SAI	-.135	.057	7,754	9	.568	.432
SAI	-.266	.052	7,754	9	.633	.367
SAI	-.281	.089	7,754	9	.640	.360
SAI	-.149	.043	7,754	9	.575	.425
SAI	-.208	.159	7,754	9	.604	.396
Subtest (GI)	-.051	.030	7,754	9	.526	.474
Subtest (NO)	-.224	.037	7,754	9	.612	.388
Subtest (AD)	-.141	.037	7,754	9	.570	.430
Subtest (WK)	-.062	.033	7,754	9	.531	.469
Subtest (AR)	-.235	.061	7,754	9	.618	.382
Subtest (SP)	-.147	.062	7,754	9	.573	.427
Subtest (MK)	-.253	.072	7,754	9	.626	.374
Subtest (EI)	-.145	.060	7,754	9	.573	.427
Subtest (MC)	-.147	.053	7,754	9	.573	.427
Subtest (GS)	-.059	.037	7,754	9	.530	.470
Subtest (SI)	-.095	.035	7,754	9	.548	.452
Subtest (AI)	-.117	.031	7,754	9	.558	.442

Note. Data are from Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 with applications to Forms 8, 9, and 10 (CNSA-1160) by W. H. Sims, and C. M. Hiatt, 1981, Alexandria, VA: Center for Naval Analyses.

[a]Definitions of abbreviations are found in Table A-1 for the subtests and Table A-4 for the composites.

[b]Weighted by study sample size.

APPENDIX D: EFFECT SIZE SUMMARIES AND AUTHOR TABLES FOR ASVAB FORMS 1, 2, 3, 5, 6, AND 7: COMPOSITES FOR PREDICTION OF FSG

Table D-1. Effect Size Summary (Validities) for ASVAB Forms 1, 2, and 3: Composites for Prediction of Final School Grade

Composites[a]	Mean r[b]	SD[b]	Total N	Number r's	BESD range	
M	.20	.08	9,165	13	.40	.60
A	.31	.03	1,657	6	.35	.65
G	.26	.10	3,863	11	.37	.63
E	.23	.14	2,186	10	.39	.61

[a]Definitions of abbreviations are found in Table A-3.
[b]Weighted by sample size

Table D-2. Author Table for Effect Size Summary for ASVAB Forms 1, 2, and 3: Composites for Prediction of FSG, Table D-1.

Composites	No. of Entries	Authors
All	9	Jensen, H. E., Valentine, L. D., Jr. (1976)
	5	Wilbourn, J. M., Guinn, N., and Leisey, S. A. (1976)
	4	Hawley, J. K., Mullins, C. J., and Weeks, J. L. (1977)

Table D-3. ASVAB Forms 6 and 7 Subtests and Composites Effect Sizes (Validities) for Prediction of FSG

	Mean r[a]	SD[a]	Total N	Number r's	BESD range	
			Subtests[b]			
GI	.30	.080	37,006	7	.35	.65
NO	.21	.030	37,006	7	.39	.61
AD	.09	.020	37,006	7	.45	.55
WK	.26	.030	37,006	7	.37	.63
AR	.31	.040	37,006	7	.35	.65
SP	.19	.020	37,006	7	.41	.59
MK	.34	.040	37,006	7	.33	.67
EI	.27	.040	37,006	7	.37	.63
MC	.29	.040	37,006	7	.36	.64
GS	.31	.050	37,006	7	.35	.65
SI	.22	.030	37,006	7	.39	.61
AI	.23	.040	37,006	7	.38	.62
			Composites[b]			
AFQT	.25	.030	42,717	3	.37	.63
M	.38	.070	32,352	9	.31	.69
A	.30	.060	29,591	7	.35	.65
G	.37	.020	36,603	7	.31	.69
E	.42	.030	29,120	6	.29	.71
SAI	.41	.020	35,015	3	.29	.71

[a]Weighted by sample size.
[b]Definitions for abbreviations are found in Table A-1 for the subtests and Table A-4 for the composites.

Table D-4. Author Table for ASVAB Forms 6 and 7 Subtests and Composites Effect Sizes (Validities) for Prediction of FSG

	No. of entries	Authors
Subtests		
All	1	Department of the Navy (1981)
	3	Department of the Navy (1982)
	1	Maier, M. H., and Hiatt, C. M. (1984)
	1	Sims, W. H., (1978)
	1	Sims, W. H., and Hiatt, C. M. (1981)
Composites		
AFQT	1	Mullins, C. J., Earles, J. A., and Wilbourn, J. M. (1979)
	1	DoD (1987)
	1	Wagner, M. P., Dirmeyer, R. P., Means, B., and Davidson, M. K. (1982)
M, A	1	Mullins, C. J., Earles, J. A., and Ree, M. J. (1981)
	1	Mullins, C. J., Earles, J. A,. and Wilbourn, J. M. (1979)
	1	Sims, W. H. and Hiatt, C. M. (1981)
	1	Department of the Navy (1981)
	3	Department of the Navy (1982)
G	1	Maier, M. H. and Hiatt, C. M. (1984)
	1	Mullins, C. J., Earles, J. A., and Ree, M. J. (1981)
	1	Mullins, C. J., Earles, J. A. ,and Wilbourn, J. M. (1979)
	1	Sims, W. H. and Hiatt, C. M. (1981)
	3	Department of the Navy (1982)
E	1	Mullins, C. J., Earles, J. A., and Ree, M. J.(1981)
	1	Mullins, C. J., Earles, J. A., and Wilbourn, J. M. (1979)
	1	Sims, W. H. and Hiatt, C. M. (1981)
	3	Department of the Navy (1982)

APPENDIX E: EFFECT SIZE SUMMARIES, OPERATIONAL COMPOSITES ASVAB FORMS 5, 6, AND 7 COMPOSITES FOR PREDICTION OF TIME-TO-COMPLETION (TTC)

Table E-1. ASVAB Forms 6 and 7 Subtests and Composites Effect Sizes (Validities) for Prediction of TTC (days)

	Mean r^a	SD^a	Total N	Number r's	BESD range	
			Subtests[b]			
GI	-.10	.07	26,357	10	.55	.45
NO	-.26	.04	27,022	11	.63	.37
AD	-.14	.03	15,509	10	.57	.43
WK	-.06	.02	15,509	10	.53	.47
AR	-.24	.04	15,509	10	.62	.38
SP	-.15	.04	15,509	10	.57	.43
MK	-.24	.04	15,509	10	.62	.38
EI	-.15	.04	15,509	10	.57	.43
MC	-.15	.04	15,509	10	.57	.43
GS	-.06	.03	15,509	11	.53	.47
SI	-.10	.02	15,509	10	.55	.45
AI	-.12	.02	15,509	10	.56	.44
			Composites[b]			
AFQT	-.07	.05	358	2	.63	.34
M	-.14	.03	15,058	9	.57	.43
A	-.22	.04	15,509	10	.61	.39
G	-.18	.04	15,509	10	.59	.41
E	-.28	.06	15,054	9	.64	.36
SAI	-.21	.06	7,754	6	.61	.39

[a]Weighted by sample size.
[b]Definitions of abbreviations are found in Tables A-1 for the subtests and A-4 for the composites.

Table E-2. Author Table for ASVAB Forms 6 and 7 Subtests and Composites Effect Sizes (Validities) for Prediction of TTC (days).

	No. of Entries	Authors
Subtests		
All (except GI,NO, and GS)	1	Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981)
	9	Department of the Navy (1981)
GI	1	Booth-Kewley, S. (1985)
	1	Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981)
	8	Department of the Navy (1981)
NO	1	Booth-Kewley, S. (1985)
	1	Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981)
	9	Department of the Navy (1981)
GS	1	Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981)
	10	Department of the Navy (1981)
Composites		
M, E	1	Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981)
	8	Department of the Navy (1981)
G	1	Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981)
	9	Department of the Navy (1981)
SAI	6	Brown, C. J., Kincaid, P. J., and McMorrow, H. (1981)

APPENDIX F: EFFECT SIZE SUMMARIES, OPERATION COMPOSITES--ALL FORMS' COMPOSITES FOR PREDICTION OF JOB PERFORMANCE MEASURES (JPM)

Table F-1. ASVAB Forms 6 and 7 Subtests and Composites Effect Sizes (Validities) for Prediction of JPM

	Mean r[a]	SD[a]	Total N	Number r's	BESD range	
			Subtests[b]			
GI	.38	.09	16,005	2	.31	.69
NO	.34	.03	16,005	2	.33	.67
AD	.21	.02	16,005	2	.40	.60
WK	.43	.09	18,980	3	.28	.72
AR	.46	.09	18,980	3	.27	.73
SP	.32	.06	18,980	3	.34	.66
MK	.42	.02	16,005	2	.29	.71
EI	.44	.04	18,980	3	.28	.72
MC	.42	.03	18,980	5	.29	.71
GS	.40	.08	17,345	3	.30	.70
SI	.37	.02	18,980	3	.32	.68
AI	.35	.01	18,980	3	.32	.68
			Composites[b]			
AFQT	.38	.06	68,845	6	.31	.69
M	.30	.26	23,689	21	.35	.65
A	.19	.18	23,224	11	.40	.60
G	.25	.16	38,005	16	.37	.63
E	.17	.25	13,347	11	.42	.58
SAI's	.41	.02	110,193	8	.29	.71

[a]Weighted by sample size.
[b]Definitions of abbreviations are found in Table A-1 for the subtests and A-4 for the composites.

Table F-2. Author Table for ASVAB Forms 6 and 7 Subtests and Composites Effect Sizes (Validities) for Prediction of JPM.

	No. of Entries	Authors
Composites		
M	1	Grafton, F. C., and Horne, D. K. (1985)
	12	Maier, M. H., and Hiatt, C. M. (1984)
	1	DoD (1982a)
	1	Park, R. K., Mathews, J. J. and Ree, M. J. (1985)
	6	Wiley, L. N., and Hahn, C. P. (1977)
A	1	Grafton, F. C., and Horne, D. K. (1985)
	4	DoD (1982a)
	6	Wiley, L. N., and Hahn, C. P. (1977)
G	4	Grafton, F. C., and Horne, D. K. (1985)
	3	Maier, M. H., and Hiatt, C. M. (1984)
	3	DoD (1982a)
	6	Wiley, L. N., and Hahn, C. P. (1977)
E	1	Grafton, F. C., and Horne, D. K. (1985)
	3	Maier, M. H., and Hiatt, C. M. (1984)
	1	DoD (1982a)
	6	Wiley, L. N., and Hahn, C. P. (1977)

* U. S. GOVERNMENT PRINTING OFFICE: 1990--761-051/20083