REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPRDC TR 83-27
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): RELATIONSHIP BETWEEN CORRESPONDING ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB) AND COMPUTERIZED ADAPTIVE TESTING (CAT) SUBTESTS
5. TYPE OF REPORT & PERIOD COVERED: Final Report, Oct 1980-Sep 1982
6. PERFORMING ORG. REPORT NUMBER: 12-83-11
7. AUTHOR(s):
Kathleen E. Moreno
C. Douglas Wetzel
James R. McBride
David J. Weiss
8. CONTRACT OR GRANT NUMBER(s): N00123-79-C-1273
9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Navy Personnel Research and Development Center
San Diego, California 92152
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: PE62763N, CF63-521-080-101-04.12
11. CONTROLLING OFFICE NAME AND ADDRESS: Navy Personnel Research and Development Center, San Diego, California 92152
12. REPORT DATE: August 1983
13. NUMBER OF PAGES: 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report):

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

Computerized adaptive testing (CAT)
Psychological testing
Armed Forces Qualification Test (AFQT)
Armed Services Vocational Aptitude Battery (ASVAB)
Bayesian adaptive test
Factor analysis
Personnel testing
Tailored testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The relationship between selected subtests from the Armed Services Vocational Aptitude Battery (ASVAB) and corresponding subtests administered as computerized adaptive tests (CAT) was investigated using a sample of Marine recruits. Results showed that the CAT subtest scores correlated as well with initial ASVAB scores as did ASVAB retest scores, even though the CAT subtests contained only half the number of items. Factor analysis showed the CAT subtests loaded on the same factors as did the corresponding ASVAB subtests, indicating that the same mental abilities were being

measured. The Armed Services Qualification Test (AFQT) composite was predicted equally well from either ASVAB or CAT administrations, even though the CAT contained only three of the four AFQT subtests. CAT requires fewer test items to perform the same task as the current paper-and-pencil ASVAB.

NPRDC TR 83-27

AUGUST 1983

RELATIONSHIP BETWEEN CORRESPONDING ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB) AND COMPUTERIZED ADAPTIVE TESTING (CAT) SUBTESTS

NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152

NPRDC TR 83-27 August 1983

RELATIONSHIP BETWEEN CORRESPONDING ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB) AND COMPUTERIZED ADAPTIVE TESTING (CAT) SUBTESTS

Kathleen E. Moreno
C. Douglas Wetzel
James R. McBride

Navy Personnel Research and Development Center

David J. Weiss

University of Minnesota
Minneapolis, Minnesota

Reviewed by
Martin F. Wiskoff

Released by
J. W. Renard
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

FOREWORD

A joint-service coordinated effort is in progress to develop a computerized adaptive testing (CAT) system and to evaluate its potential for use in the Military Entrance Processing Stations as a replacement for the Armed Services Vocational Aptitude Battery (ASVAB) printed tests. The Department of the Navy (Headquarters, U.S. Marine Corps) has been designated as lead service for CAT system development; and the Navy Personnel Research and Development Center, as lead laboratory.

This report describes an investigation of the relationship between selected ASVAB subtests and corresponding CAT subtests, which was conducted as part of project CF63-521-080-101-04.12 (USMC Computerized Adaptive Testing). The data were collected by the University of Minnesota, pursuant to contract N00123-79-C-1273. Results are directed toward technical, professional, and contractor personnel involved in implementing CAT.

Previous reports described CAT system functional requirements and schedules, preliminary design considerations, and the influence of fallible item parameter on adaptive testing (NPRDC Tech. Note 82-22 and Tech. Reps. 82-52 and 83-15).

J. W. RENARD
Commanding Officer

JAMES W. TWEEDDALE
Technical Director

SUMMARY

Problem

The Navy Personnel Research and Development Center is developing a computerized adaptive testing (CAT) system as a possible replacement for the paper-and-pencil Armed Services Vocational Aptitude Battery (ASVAB). An essential feature of CAT is the tailoring of aptitude test items to the individual by selecting those items whose psychometric characteristics closely match the examinee's apparent ability level. In developing CAT as a replacement for ASVAB, care is being taken to ensure that CAT tests will be as accurate as the current printed ASVAB tests. This concern raises the question as to whether CAT and ASVAB measure the same abilities. The relationship between the two types of tests has not been investigated thoroughly.

Objective

The objective of this effort was to determine (1) the relationship between selected paper-and-pencil ASVAB subtests and an experimental battery of three corresponding CAT subtests and (2) whether corresponding CAT and ASVAB subtests measure the same aptitudes.

Approach

Marine recruits were administered an initial ASVAB, an ASVAB retest, and CAT subtests corresponding to ASVAB subtests on Word Knowledge (WK), Arithmetic Reasoning (AR), and Paragraph Comprehension (PC). The CAT subtests were approximately half as long as the ASVAB subtests.

Findings

The three CAT subtests correlated as well or better with initial ASVAB subtests as did subtests from the ASVAB retest. Factor analysis showed that the CAT subtests loaded on the same factors as the corresponding ASVAB subtests. The Armed Forces Qualification Test (AFQT) composite was predicted equally well from either the ASVAB administration or the CAT administration, despite the fact that the CAT contained only three of the four subtests used to compute the AFQT score.

Conclusions

The results support the continued development of CAT as a replacement for the paper-and-pencil ASVAB. It appears that CAT can serve the same ability measurement purpose as ASVAB, and can do so with substantial efficiency.

Current Efforts

1. Additional research has been undertaken to extend the present findings to a full complement of ASVAB-counterpart CAT subtests.

2. The utility of CAT for predicting recruits' performance in service schools will be evaluated in a criterion-related validity assessment.

CONTENTS

LIST OF TABLES

INTRODUCTION

Problem and Background

The Department of Defense is currently developing a computerized adaptive testing (CAT) system as a potential replacement for the conventional paper-and-pencil tests used for enlisted personnel selection and classification. The existing Armed Services Vocational Aptitude Battery (ASVAB) consists of a fixed sequence of test items administered to all examinees. CAT entails automated tailoring of a sequence of test items to each examinee, contingent upon his/her responses to earlier items in the sequence. Correct responses are generally followed by more difficult items, and incorrect responses by easier items. CAT requires substantially fewer test items than does ASVAB because items that are too easy or too difficult for the examinee are not administered. Additionally, computerization offers further advantages by eliminating the clerical errors inherent in manual test administeration and by increasing test security.

In developing CAT as a replacement for ASVAB, care is being taken to ensure that CAT tests will be as accurate as the current printed ASVAB tests. Data related to this question have been presented by McBride (1979) and by McBride and Martin (1983), who found that a CAT test of verbal ability was more reliable and more valid than a conventional test. Concern for CAT's accuracy also raises the question as to whether CAT and ASVAB measure the same abilities. The relationship between CAT and the conventional tests currently employed in the military has not been investigated thoroughly (cf., Sympson, Weiss, & Ree, 1982).

Objective

The objective of this effort was to determine (1) the relationship between selected paper-and-pencil ASVAB subtests and an experimental battery of three corresponding CAT subtests and (2) whether corresponding CAT and ASVAB subtests measure the same aptitudes.

APPROACH

Subjects

Subjects were 356 male Marine Corps recruits between 17 and 26 years of age, stationed at the Marine Corps Recruit Depot (MCRD), San Diego.

Test Instruments

ASVAB

The current versions of ASVAB (forms 8, 9, and 10) consist of 10 subtests, which are listed in Table 1. Each ASVAB subtest consists of items of difficulty levels that span the range of abilities to be found in an unselected applicant population. The Armed Forces Qualification Test (AFQT) score, which is used by the military services to determine eligibility for enlistment, is computed from scores obtained by an applicant on four ASVAB subtests: Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), and Numerical Operations (NO). AFQT was computed as the sum of the AR, WK, and PC raw scores and half the NO raw score (AR + WK + PC + .5 NO). In this study, the raw ASVAB subtest scores and the raw AFQT composite were used for analysis.

Table 1

ASVAB and CAT Subtests

Subtests	Abbreviation	Number of Items	Time (Minutes)[b]
ASVAB:			
General Science	GS	25	11
Arithmetic Reasoning	AR[a]	30	36
Word Knowledge	WK[a]	35	11
Paragraph Comprehension	PC[a]	15	13
Numerical Operations	NO[a]	50	3
Coding Speed	CS	84	7
Auto and Shop Information	AS	25	11
Mathematics Knowledge	MK	25	24
Mechanical Comprehension	MC	25	19
Electronics Information	EI	20	9
CAT:			
Arithmetic Reasoning	CATAR	15	20
Word Knowledge	CATWK	15	6
Paragraph Comprehension	CATPC	8	10

[a]These subtests are used to compute the Armed Forces Qualification Test (AFQT) score (AR + WR + PC + .5 NO).

[b]Times are standard administration times for ASVAB subtests and average administration times for CAT subtests. Times do not include that needed to read instructions and perform other administrative details.

CAT

The CAT battery used in this investigation consisted of three subtests designed to measure Arithmetic Reasoning (CATAR), Word Knowledge (CATWK), and Paragraph Comprehension (CATPC). These tests, administered with a fixed number of items, are listed in Table 1.

Owen's Bayesian sequential tailored testing procedure (Owen, 1969, 1975), which selects items by optimizing a mathematical function of the difference between the examinee's estimated ability and the item's difficulty, was used to choose the sequence of items administered to an examinee. All examinees start with the same test item, which is of intermediate difficulty. The difficulty of subsequent items varies according to individual responses; more difficult items follow correct responses and easier items follow incorrect responses. The Bayesian test score yielded by CAT after each item is a statistical estimate of an examinee's location on a real number scale of ability. In practice, such estimates generally range between +3 and -3, when scaled to have a theoretical mean of 0 and variance of 1. This scale was employed because item difficulties varied among examinees, so that number-correct scoring was inappropriate. The adaptive tests were administered without a time limit, while ASVAB was given with a standard timed administration. This procedural difference should be borne in mind when comparing the test times shown in Table 1.

Item banks for the three CAT subtests had previously been calibrated, using a three-parameter logistic item response model (see Lord, 1980, p. 12 and Wetzel & McBride, 1983). The three parameters provide indices of guessing, difficulty, and discriminability, as described by an item response function that describes the probability of correctly answering an item as a function of examinee ability. The guessing parameter reflects the probability of correctly answering an item by individuals of infinitely low ability; a value of zero would be obtained if an item cannot be answered by guessing. The difficulty parameter reflects the location of the item response function with respect to ability; this parameter is the ability level where the probability of a correct answer is half way between 1.0 and the guessing parameter. Finally, the discrimination parameter is proportional to the slope of the item response function at the inflection point; it represents the degree to which item response varies with ability level. The obtained average and upper and lower limits of the estimated item parameters for each CAT subtest are summarized in Table 2.

Table 2

Descriptive Statistics for CAT Item Parameters

Subtest/ Item Parameter	Lower Limit	Upper Limit	Average
CATAR:			
Discrimination	.716	2.000	1.194
Difficulty	-1.355	1.966	.582
Guessing	.029	.297	.179
CATWK:			
Discrimination	.800	2.690	1.355
Difficulty	-1.980	2.000	-.337
Guessing	.040	.260	.128
CATPC:			
Discrimination	1.000	1.000	1.000
Difficulty	-2.952	1.665	-.026
Guessing	.000	.000	.000

The CAT item banks are described below:

1. The CATAR item bank consisted of 225 items, 148 of which had been calibrated on a selected population of Air Force enlistees (Sympson, Weiss, & Ree, 1982). Since this 148-item pool was deficient in easier items, 77 additional items were calibrated from a

paper-and-pencil test administered to a sample of 4,100 Navy and Marine recruits. Item parameters were estimated using the LOGIST program (Wood, Wingersky, & Lord, 1976). Reckase's (1979) "major axis" method was used to link the new items with the original item pool.

2. The CATWK item bank consisted of 78 items--39 that had been computer-administered to 677 Marine recruits and 39 that had been calibrated from a paper-and-pencil test administered to samples of up to 1,300 Marine recruits. Item parameters were estimated using item calibration methods developed by Urry (1977, 1978).

3. The CATPC item bank consisted of 25 items that had been computer-administered to samples ranging in size from 239 to 481 Marine recruits. LOGIST was used to develop item parameter estimates. Due to the small sample size obtained for some items, the discrimination and guessing parameters were set at 1.0 and 0.0 respectively. During both item calibration and the actual CATPC test session, the paragraph and the question to be answered were presented on separated screens. Thus, unlike the ASVAB PC subtest, examinees were not allowed to refer back to the paragraph while responding to the multiple-choice item.

Procedure

Subjects were administered the initial ASVAB test by recruiters at the Military Entrance Processing Station before they enlisted in the armed forces; and the ASVAB retest (using an alternate ASVAB form), as part of a routine testing program by Marine Corps examiners at the recruit depot approximately 2 weeks after they entered active duty. The time lapse between the two ASVAB administrations varied between 2 weeks and approximately 6 months because of the availability of training programs.

The CAT tests were administered to available recruits approximately 24 hours after their arrival at the recruit depot during 3 months in 1981. They were administered by computer, on one of four cathode-ray tube terminals located in a specially designated testing room. The computer that controlled test administration, which was located at the University of Minnesota, was connected to the remote terminals by a dedicated telecommunications line using a data transmission rate of 120 characters per second on each terminal. Instructions introducing the examinees to the testing situation were given by a civilian proctor. Instructions on how to enter answers, change answers, etc. were given directly on each terminal, using interactive instruction under computer program control. In addition, each subtest was preceded by a set of instructions and one or more practice questions. To ensure that an examinee used the terminal correctly, the subtest began only after he had responded correctly to the practice questions. Scratch paper was provided for computations during the AR subtest. At the end of the testing session, the examinee's percentile rank for each subtest was displayed on the screen. Total test time for CAT was, on the average, 55 minutes, including all instructions on terminal use (see individual test times in Table 1).

Data for examinees with missing scores on any of the three tests (initial ASVAB test, ASVAB retest, or CAT) and for those who had taken obsolete forms of ASVAB on either initial testing or retest (i.e., versions other than forms 8, 9, or 10) were excluded from analysis, leaving a final sample of 270 subjects. Table 3 contains the mean and standard deviation of each subtest and AFQT composite for this sample.

Table 3

Means, Standard Deviations, and Range for ASVAB, AFQT, and CAT Subtests (N = 270)

Variable	Mean	Std. Dev.	Range
ASVAB Initial Test:			
GS	17.53	3.97	7.0 - 25.0
AR	21.77	5.41	5.0 - 30.0
WK	28.17	4.89	16.0 - 35.0
PC	11.78	2.20	3.0 - 15.0
NO	40.60	7.22	17.0 - 50.0
CS	49.50	11.49	12.0 - 84.0
AS	18.04	4.18	5.0 - 25.0
MK	14.94	5.26	5.0 - 25.0
MC	17.26	4.15	4.0 - 25.0
EI	13.39	3.21	3.0 - 20.0
AFQT composite	82.03	11.81	51.0 - 105.0
ASVAB Retest:			
GS	17.41	4.07	8.0 - 25.0
AR	21.43	5.71	7.0 - 30.0
WK	28.06	4.86	9.0 - 35.0
PC	11.48	2.50	3.0 - 15.0
NO	42.15	8.08	11.0 - 50.0
CS	52.66	14.44	14.0 - 84.0
AS	18.15	4.27	5.0 - 25.0
MK	15.24	5.34	4.0 - 25.0
MC	17.78	4.35	5.0 - 25.0
EI	13.62	3.29	0.0 - 20.0
AFQT composite	82.05	12.64	48.5 - 105.0
CAT:			
AR	0.40	0.82	-1.57 - 2.52
WK	0.59	0.79	-1.63 - 2.54
PC	0.08	0.85	-2.52 - 1.53

Note. ASVAB and AFQT scores are in raw (number correct) score units; CAT scores are in scaled (real number) score units.

Data Analyses

1. Pearson correlation coefficients were computed between all variables. Those computed between CAT and ASVAB subtest scores were compared to those computed between the ASVAB initial test and retest subtest scores.

2. To reveal those clusters of subtests with high intercorrelations but low correlations with the remaining subtests, two factor analyses were performed on the intercorrelation matrix (see appendix), using the principal axes method. The main diagonal elements of the correlation matrix were replaced with communality estimates, with squared multiple correlations used as initial estimates of communality. Each analysis was followed by a varimax rotation to simplify the factor structure. The first analysis included only ASVAB subtests as variables, in order to establish the internal factor structure of ASVAB. The second also included the CAT variables.

3. Two multiple regression analyses were performed. The first was performed to determine whether the AFQT composite computed from initial ASVAB subtest scores could be predicted using the ASVAB retest AR, WK, PC, and NO scores; and the second, whether it could be predicted using CAT AR, WK, and PC scores.

RESULTS

Intercorrelations

Table 4, which provides correlations for ASVAB and CAT AR, WK, and PC subtests, shows that each CAT subtest correlated slightly higher with its ASVAB counterpart than did the corresponding ASVAB alternate form. This indicates that the relationship between CAT and ASVAB scores is as strong as that between ASVAB initial test and retest scores. This result was obtained even though the two ASVAB test forms are considered parallel for these three subtests, and the CAT subtests were half the length of their ASVAB counterparts. Correlations of the magnitude observed here have been reported by Sympson, Weiss, and Ree (1982) for Air Force jet engine mechanic trainees who took AR and WK subtests administered both in ASVAB and adaptive testing. The ASVAB test-retest correlations shown here were also similar to those observed in previous research on the reliability of the ASVAB (Fruchter & Ree, 1977; Ree, Mullins, Mathews, & Massey, 1982; OSD(MRA&L), 1982).

Table 4

Pearson Correlation Coefficients for ASVAB and CAT AR, WK, and PC Subtests

	ASVAB Initial Test			ASVAB Retest		
	AR	WK	PC	AR	WK	PC
ASVAB Retest:						
AR	.7673					
WK		.7705				
PC			.4636			
CAT:						
AR	.7996			.7997		
WK		.8059			.7991	
PC			.5072			.5052

Note. The full correlation matrix is in the appendix.

Factor Analyses

From the first analysis, which included only ASVAB subtests as variables, four factors were extracted, based on an eigenvalue of 1.0 or greater. These factors accounted for 62 percent of the total variance. Table 5, which presents the varimax rotated factor matrix solution, indicates that Factors 1 through 3 are of approximately equivalent strength and Factor 4 is slightly weaker. The four factors have been tentatively labeled as follows:

1. Verbal: Word knowledge and the ability to manipulate words and verbal concepts.

2. Technical-Mechanical: Mechanical comprehension or mechanical experience factor, dealing with the functions of machines or simple physical devices.

3. Mathematical-Quantitative: Ability to use numbers and mathematical concepts.

4. Speed: Ability to solve simple problems rapidly and perform clerical tasks accurately.

These factors are very similar to those identified in other factor analyses of ASVAB (Fischl, Ross, & McBride, 1979; Ree, Mullins, Mathews, & Massey, 1982).

Table 6 presents the varimax rotated factor solution to the second analysis, which was performed with the CAT variables added to the data matrix. As shown, CATWK and CATPC loaded substantially on the verbal factor; and CATAR, on the mathematical factor. While the amount of total variance accounted for by each of the factors changed by adding the CAT variables, the structure of the four factors remained essentially the same. The verbal factor was still the strongest, accounting for 20.2 percent of the total variance instead of 17.3 percent. This increase in explained variance could be expected with the addition of CATWK and CATPC, which test verbal and reading skills. With the addition of CATAR, the mathematical factor became stronger than the technical factor. The total variance explained by these two factors was 17.2 and 15.8 percent respectively, compared to 16.5 and 17.3 when only ASVAB variables were included in the analyses.

CATWK loaded higher (.83) than any other variable on the verbal factor, which accounted for 68 percent of the variance in CATWK. This indicates that CATWK is mainly a measure of verbal ability. While CATPC loaded higher (.54) on the verbal factor than on any other factor, the verbal factor accounted for only 29 percent of the variance in CATPC. The four factors together accounted for 43 percent of the variance in CATPC, as shown by the final communality estimate. These results suggest that much of the CATPC variance is unique or unreliable. The latter seems more likely since the CATPC test was short, the small item bank had been calibrated with only a one-parameter model, and the corresponding ASVAB PC subtest had the lowest test-retest reliability obtained. The fact that factor loadings for CATPC were comparable to those for ASVAB PC, both initial test and retest, indicates that CATPC measures reading comprehension as well as its ASVAB counterparts, despite its shorter length.

CATAR loaded higher (.76) than any other variable on the mathematical factor, which accounted for 58 percent of the variance in CATAR. The four factors together accounted for 78 percent of the variance in CATAR, with the verbal factor explaining 12 percent of the variance. Thus, while CATAR is primarily a measure of mathematical ability, verbal ability is also involved in understanding and solving these word problems. This is true for the majority of the ASVAB subtests, with the possible exception of computational tests such as NO.

Table 5

Varimax Rotated Factor Matrix for Analysis Using Only ASVAB Variables

Variable	Factor 1 (Verbal)	Factor 2 (Technical)	Factor 3 (Math)	Factor 4 (Speed)	Final Communality Estimates
ASVAB Initial Test:					
GS	.60	.44	.31	.07	.66
AR	.31	.21	.72	.17	.68
WK	.82	.16	.23	.08	.76
PC	.55	.09	.33	.10	.43
NO	.04	.12	.22	.68	.53
CS	.12	-.00	.06	.72	.54
AS	.10	.83	.04	-.02	.71
MK	.28	.17	.77	.26	.77
MC	.33	.48	.44	.11	.55
EI	.34	.56	.25	.02	.48
ASVAB Retest:					
GS	.57	.48	.36	.07	.70
AR	.33	.26	.70	.25	.73
WK	.82	.23	.19	.09	.76
PC	.52	.17	.30	.27	.46
NO	.10	-.08	.20	.56	.37
CS	.05	.04	.06	.73	.54
AS	.07	.85	.05	.02	.73
MK	.35	.16	.73	.27	.75
MC	.26	.61	.44	.13	.65
EI	.33	.62	.32	-.00	.59
Factor Contribution	3.46	3.45	3.31	2.20	12.41
Common Variance	27.85%	27.79%	26.63%	17.74%	
Cumulative Variance[a]	27.85%	55.64%	82.27%	100.00%	
Total Variance	17.29%	17.25%	16.53%	11.01%	
Cumulative Variance[a]	17.29%	34.53%	51.06%	62.07%	

Note. Factor loadings greater than .35 are underlined.

[a]Cumulative values do not always total due to rounding.

Table 6

Varimax Rotated Factor Matrix for
Analysis Using Both ASVAB and CAT Variables

Variable	Factor 1 (Verbal)	Factor 2 (Math)	Factor 3 (Technical)	Factor 4 (Speed)	Final Communality Estimates
ASVAB Initial Test:					
GS	.62	.27	.45	.07	.66
AR	.31	.75	.21	.15	.73
WK	.82	.22	.16	.07	.75
PC	.56	.34	.08	.08	.45
NO	.04	.24	.12	.68	.53
CS	.13	.06	.00	.72	.54
AS	.12	.05	.81	-.02	.68
MK	.31	.73	.19	.26	.72
MC	.35	.41	.49	.11	.55
EI	.34	.23	.56	.02	.49
ASVAB Retest:					
GS	.58	.33	.49	.07	.69
AR	.34	.72	.27	.23	.76
WK	.82	.17	.23	.08	.76
PC	.54	.30	.17	.26	.48
NO	.10	.21	-.08	.56	.37
CS	.06	.07	.04	.73	.54
AS	.08	.05	.84	.02	.72
MK	.37	.72	.18	.26	.75
MC	.25	.41	.63	.13	.64
EI	.31	.30	.63	-.01	.59
CAT:					
AR	.35	.76	.20	.21	.78
WK	.83	.25	.26	.13	.83
PC	.54	.33	.12	.10	.43
Factor Contribution	4.64	3.94	3.62	2.22	14.44
Common Variance	32.16%	27.32%	25.09%	15.43%	
Cumulative Variance[a]	32.16%	59.48%	84.57%	100.00%	
Total Variance	20.18%	17.15%	15.75%	9.69%	
Cumulative Variance[a]	20.18%	37.33%	53.08%	62.77%	

Note. Factor loadings greater than .35 are underlined.

[a]Cumulative values do not always total due to rounding.

In sum, the factor loadings for the three CAT subtests were quite similar to those for their ASVAB counterparts. Therefore, it appears that the CAT and ASVAB subtests measure the same aptitude factors.

AFQT Regressions

Table 7 presents a summary of the multiple regressions used to evaluate the predictability of the AFQT composite computed from the initial ASVAB subtest scores. As shown, the regression of the initial AFQT composite on the best linear composite of ASVAB retest scores resulted in a multiple correlation of .85. The regression of the initial AFQT composite on CAT AR, WK, and PC subtests was .87, with CAT WK and AR subtests contributing significantly to predicting the variance in AFQT. The beta weights for CAT AR, WK, and PC subtests were .53, .43, and .03 respectively. Overall, the three CAT subtests explained 75 percent of the variance in AFQT initial test scores, compared to 73 percent explained by the four ASVAB retest subtests.

Table 7

Summary of Multiple Regression Analyses Performed to Predict AFQT Composite Computed from Initial ASVAB Test

Variable	Multiple R	B Weights (Unstdzd.)	Std. Error of B	Beta Weights (Stdzd.)	F
		From ASVAB Retest Subtests			
AR	.764	1.004	.084	.485	142.420*
WK	.832	.814	.099	.335	67.824*
NO	.849	.243	.050	.166	23.845*
PC	.853	.551	.202	.117	7.447*
(Constant)		21.095			
		From CAT Subtests			
AR	.788	7.568	.567	.527	178.151*
WK	.865	6.382	.605	.428	111.464*
PC	.866	.369	.540	.027	.467
(Constant)		75.210			

Note. Multiple Rs reflect values obtained using a stepwise procedure; all others are final values obtained after all variables had been entered into the equation.

$* < .01$.

DISCUSSION AND CONCLUSIONS

The results of this research support the continued development of CAT as a replacement for the paper-and-pencil ASVAB. These results are notable in that military examinees were used to calibrate the test items and to determine the relationship between CAT and ASVAB.

CAT was clearly found to be as valid a measure of the abilities tested as were the corresponding ASVAB subtests, as noted below:

1. CAT subtest scores correlated as highly with ASVAB initial test scores as did the ASVAB retest scores.

2. Factor analysis showed that ability estimates from CAT subtests loaded on the same factors as did their counterpart ASVAB subtests, with the factor loadings for the CAT subtests being comparable in value to those for the ASVAB subtests.

3. The AFQT composite score was predicted equally well by either the ASVAB retest scores or the CAT subtest scores, despite the fact that the CAT subtests were substantially shorter and represented only three of the four AFQT component subtests.

The psychometric quality of ASVAB may be achieved by CAT with about half the number of test items. With ASVAB, all examinees answer exactly the same items, which vary considerably in difficulty. Thus, examinees with more extreme abilities must take items that are either too easy or too difficult. With CAT, each examinee receives a unique sequence of items that are tailored in difficulty to that examinee, based on his or her prior pattern of responses. The CAT technique can achieve the same quality of test scores with fewer items because many items that the examinee would most likely have answered correctly or incorrectly are not administered. This feature of CAT means that fewer items need be administered to achieve the same measurement precision as a conventional test.

CURRENT EFFORTS

1. While the present results are favorable for the implementation of CAT, this effort must be extended to include a CAT battery that spans all the ASVAB subtests. Such a battery has been developed. Work is in progress to administer it to selected groups of military personnel prior to entry level technical training. This research will yield data similar to those reported here, as well as validities with respect to a school performance criterion.

2. The utility of CAT for predicting recruits' performance in service schools will be evaluated in a criterion-related validity assessment.

REFERENCES

Croll, P. R. Computerized adaptive testing system design: Preliminary design considerations (NPRDC Tech. Rep. 82-52). San Diego: Navy Personnel Research and Development Center, July 1982. (AD-A118 495)

Fischl, M. A., Ross, R. M., & McBride, J. R. Development of factorially based ASVAB high school composites (Tech. Paper 360). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, 1979.

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

Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum, 1980.

McBride, J. R. Adaptive verbal ability testing in a military setting. Weiss, D. J. (Ed.), Proceedings of the 1979 Computerized Adaptive Testing Conference. Minneapolis: Computerized Adaptive Testing Laboratory, Department of Psychology, University of Minnesota, September 1980.

McBride, J. R. Computerized adaptive testing project: Objectives and requirements (NPRDC Tech. Note 82-22). San Diego: Navy Personnel Research and Development Center, July 1982. (AD-A118 447)

McBride, J. R., & Martin, J. T. Reliability and validity of adaptive ability tests in a military setting. In Weiss, D. J. (Ed.), New horizons in testing: Latent trait theory and computerized adaptive testing. New York: Academic Press, 1983.

Office of the Assistant Secretary of Defense: Manpower, Reserve Affairs, and Logistics (OSD:MRA&L). Armed Services Vocational Aptitude Battery (ASVAB) forms 8, 9, and 10: Technical manual. Author, 1982.

Owen, R. J. A Bayesian approach to tailored testing (Res. Bul. 69-92). Princeton, NJ: Educational Testing Service, 1969.

Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Reckase, M. D. Item pool construction for use with latent trait models. Paper presented at The Annual Meeting of the American Educational Research Association, San Francisco, April 1979.

Ree, M. J., Mullins, C. J., Mathews, J. J., & Massey, R. H. Armed Services Vocational Aptitude Battery: Item and factor analysis of forms 8, 9, and 10 (Tech. Rep. 81-55). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, March 1982.

Sympson, J. B., Weiss, D. J., & Ree, M. J. Predictive validity of conventional and adaptive tests in an Air Force training environment (Tech. Rep. 81-40). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, August 1982.

Urry, V. OGIVIA: Item parameter estimation program with normal ogive and logistic three-parameter model options. Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center, 1977.

Urry, V. ANCILLES: Item parameter estimation program with normal ogive and logistic three-parameter model options. Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center, 1978.

Wetzel, C. D., & McBride, J. R. Influence of fallible item parameters on test information during adaptive testing (NPRDC Tech. Rep. 83-15). San Diego: Navy Personnel Research and Development Center, April 1983.

Wood, R. L., Wingersky, M. S., & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (RM-76-6). Princeton, NJ: Educational Testing Service, 1976.

APPENDIX

PEARSON CORRELATION COEFFICIENTS FOR ASVAB, AFQT, AND CAT VARIABLES

Pearson Correlation Coefficients
for ASVAB, AFQT, and CAT Variables
(N = 270)

	GS1	AR1	WK1	PC1	NO1	CS1	AS1	MK1	MC1	EI1
GS1	----									
AR1	.5163	----								
WK1	.6572	.4811	----							
PC1	.4727	.4567	.5728	----						
NO1	.2284	.3045	.1402	.1750	----					
CS1	.1401	.1905	.1709	.1423	.4855	----				
AS1	.4708	.2607	.2496	.1404	.1131	-.0132	----			
MK1	.5316	.7145	.4426	.4297	.3741	.3104	.1797	----		
MC1	.5564	.5132	.4609	.4072	.1912	.1858	.4742	.5960	----	
EI1	.5514	.4137	.3964	.3343	.1557	.0992	.5009	.4164	.5829	----
GS2	.7626	.5296	.6055	.4520	.2167	.1335	.4371	.5763	.5438	.5540
AR2	.5049	.7673	.4947	.4955	.3763	.2276	.3165	.6847	.5660	.3896
WK2	.6223	.4169	.7705	.5227	.1614	.1882	.2601	.4311	.4803	.4398
PC2	.4724	.4913	.5158	.4636	.3040	.2684	.2047	.4422	.4084	.3333
NO2	.1565	.2664	.1838	.1800	.5177	.3790	-.0653	.2900	.1043	.0520
CS2	.0980	.1910	.1288	.1171	.4766	.5912	.0341	.2520	.1766	.0579
AS2	.3974	.2425	.2097	.1402	.1370	.0167	.7535	.1834	.4061	.4684
MK2	.5177	.6865	.5089	.4475	.3866	.2898	.1796	.8046	.5199	.3480
MC2	.5419	.4949	.4151	.3440	.2481	.1624	.5000	.5424	.6954	.5030
EI2	.5600	.4663	.4353	.3326	.1861	.0146	.5416	.4310	.4655	.5811
AFQT1	.6669	.8357	.7845	.6867	.5361	.3331	.2836	.7053	.5605	.4638
AFQT2	.6109	.6892	.6807	.5741	.4575	.3494	.2626	.6552	.5545	.4277
CATAR	.4887	.7996	.5011	.5083	.3907	.2266	.2829	.7142	.5109	.3931
CATWK	.7087	.5296	.8059	.5538	.2213	.1998	.3432	.5164	.5098	.4745
CATPC	.4820	.4286	.4859	.5072	.1632	.1639	.2239	.4407	.4440	.3522

Note. Initial ASVAB subtests are followed by a 1; retest subtests by a 2.

Pearson Correlation Coefficients
for ASVAB, AFQT, and CAT Variables
(N = 270)

	GS2	AR2	WK2	PC2	NO2	CS2	AS2	MK2	MC2	EI2
GS1										
AR1										
WK1										
PC1										
NO1										
CS1										
AS1										
MK1										
MC1										
EI1										
GS2	----									
AR2	.5528	----								
WK2	.6510	.4826	----							
PC2	.5093	.5481	.5805	----						
NO2	.1271	.3186	.1205	.2464	----					
CS2	.1334	.2599	.1200	.2498	.3991	----				
AS2	.4664	.3093	.2847	.2409	-.0441	.0469	----			
MK2	.5843	.7375	.5185	.4845	.2940	.2676	.2305	----		
MC2	.6145	.5917	.4685	.3700	.1410	.1655	.5877	.5719	----	
EI2	.6331	.4784	.4626	.3739	.0641	.0618	.5795	.4618	.6088	----
AFQT1	.6441	.7640	.6572	.6182	.3901	.3085	.2661	.7271	.5388	.5130
AFQT2	.6416	.8475	.7561	.7475	.5585	.3405	.2829	.7223	.5657	.4885
CATAR	.5324	.7997	.4904	.4963	.2903	.2267	.2725	.7590	.5107	.4587
CATWK	.6896	.5642	.7991	.5984	.1902	.1811	.3111	.5645	.4568	.4751
CATPC	.4644	.4951	.5328	.5052	.1547	.1472	.1629	.4735	.3143	.2776

Note. Initial ASVAB subtests are followed by a 1; retest subtests by a 2.

Pearson Correlation Coefficients
for ASVAB, AFQT, and CAT Variables
(N = 270)

	AFQT	AFQT2	CATAR	CATWK	CATPC
GS1					
AR1					
WK1					
PC1					
NO1					
CS1					
AS1					
MK1					
MC1					
EI1					
GS2					
AR2					
WK2					
PC2					
NO2					
CS2					
AS2					
MK2					
MC2					
EI2					
AFQT1	----				
AFQT2	.8448	----			
CATAR	.7882	.7408	----		
CATWK	.7476	.7415	.5792	----	
CATPC	.5422	.5780	.5249	.5595	----

Note. Initial ASVAB subtests are followed by a 1; retest subtests by a 2.

DISTRIBUTION LIST

Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics)
Deputy Under Secretary of Defense for Research and Engineering (Research and Advanced Technology)
Military Assistant for Training and Personnel Technology (ODUSD(R&AT))
Assistant Secretary of the Navy (Manpower and Reserve Affairs)
Principal Deputy Assistant Secretary of the Navy (Manpower and Reserve Affairs) (OASN(M&RA))
Deputy Assistant Secretary of the Navy (Manpower) (OASN(M&RA))
Director of Manpower Analysis (ODASN(M))
Chief of Naval Operations (OP-01), (OP-11), (OP-12) (2), (OP-13), (OP-14), (OP-15), (OP-115) (2), (OP-140F2), (OP-987H)
Chief of Naval Material (NMAT 05), (NMAT 0722)
Chief of Naval Research (Code 200), (Code 440) (3), (Code 442), (Code 442PT)
Chief of Naval Education and Training (02), (N-2), (N-5)
Commandant of the Marine Corps (MPI-20)
Commander in Chief U.S. Atlantic Fleet
Commander in Chief U.S. Pacific Fleet
Commander Naval Military Personnel Command (NMPC-013C)
Commanding Officer, Naval Aerospace Medical Institute (Library Code 12) (2)
Commanding Officer, Naval Regional Medical Center, Portsmouth (ATTN: Medical Library)
Commanding Officer, Office of Naval Research Branch Office, Chicago (Coordinator for Psychological Sciences)
Director, Naval Civilian Personnel Command
Director, Naval Education and Training Program Development Center Detachment, Memphis
Superintendent, Naval Postgraduate School
Commander, Army Research Institute for the Behavioral and Social Sciences, Alexandria (PERI-ASL)
Headquarters Commandant, Military Enlistment Processing Command, Fort Sheridan
Chief, Army Research Institute Field Unit, Fort Harrison
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Scientific and Technical Information Office)
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base (Technical Training Branch)
Commander, Air Force Human Resources Laboratory, Williams Air Force Base (AFHRL/OT)
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base (AFHRL/LR)
Superintendent, U.S. Coast Guard Academy
Defense Technical Information Center (DDA) (12)

DEPARTMENT OF THE NAVY

NAVY PERSONNEL RESEARCH AND
DEVELOPMENT CENTER
SAN DIEGO, CALIFORNIA 92152

SUPERINTENDENT
NAVAL POSTGRADUATE SCHOOL
MONTEREY CA
93940

0142