

# 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

The relationships between selected subtests from the Armed Services Vocational Aptitude Battery (ASVAB) and corresponding subtests administered as computerized adaptive tests (CAT) were investigated using Marine recruits as subjects. Three adaptive subtests were shown to correlate as well with ASVAB as did a second administration of ASVAB, even though the CAT subtests contained only half the number of items. Factor analysis showed the CAT subtests to load on the same factors as the corresponding ASVAB subtests, indicating that the same abilities were being measured. The preenlistment Armed Forces Qualification Test (AFQT) composite scores were predicted as well from the CAT subtest scores as from the retest ASVAB subtest scores, even though the CAT contained only three of the four AFQT subtests. It is concluded that CAT can achieve the same measurement precision as a conventional test, with half the number of items.

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 ex-

aminee, contingent upon his/her responses to earlier items in the sequence (e.g., Weiss, 1982, 1983). Correct responses are generally followed by more difficult items and incorrect responses by easier items. CAT requires substantially fewer test items than a conventional test 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 administration 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 (1980) and by McBride and Martin (1983), who found that an adaptive test of verbal ability was more reliable and more valid than a conventional test. Concern for CAT's accuracy also raises the question of whether CAT and ASVAB measure the same abilities. A limited amount of research has been done on the relationship between CAT and the conventional tests currently employed in the military (Sympson, Weiss, & Ree, 1982). Further investigation is needed.

The objective of this study was to determine (1) the relationship between scores on 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.

---

*APPLIED PSYCHOLOGICAL MEASUREMENT*  
Vol. 8, No. 2, Spring 1984, pp. 155-163  
© Copyright 1984 Applied Psychological Measurement Inc.  
0146-6216/84/020155-09\$1.70

## Method

### 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, CA.

### Test Instruments

**ASVAB.** The current versions of ASVAB (Forms 8, 9, & 10) consist of 10 subtests, listed in Table 1. Each ASVAB subtest consists of items with 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 raw scores obtained by an applicant on four ASVAB subtests: Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), and Numerical Operations (NO). In this study, the

raw ASVAB subtest scores and the raw AFQT composite scores were used for analysis.

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

For each of these subtests, Owen's Bayesian sequential tailored testing procedure (Owen, 1969, 1975) was used to select items and to estimate an examinee's ability level. 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 (Lord, 1980). The three parameters provide indices of guessing, difficulty, and discrimination, for an item response function describing the probability of correctly answering

Table 1  
ASVAB and CAT Subtests

Subtest	Abbreviation	Number of Items	Time in Minutes <sup>b</sup>
<b>ASVAB</b>			
General Science	GS	25	11
Arithmetic Reasoning	AR <sup>a</sup>	30	36
Word Knowledge	WK <sup>a</sup>	35	11
Paragraph Comprehension	PC <sup>a</sup>	15	13
Numerical Operations	NO <sup>a</sup>	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
<b>CAT</b>			
Arithmetic Reasoning	CAT AR	15	20
Word Knowledge	CAT WK	15	6
Paragraph Comprehension	CAT PC	8	10

<sup>a</sup> These subtests are used to compute the Armed Forces Qualification Test (AFQT) score (AR + WK + PC + .5 NO) .

<sup>b</sup> 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.

an item as a function of examinee ability. The average, maximum, and minimum values of the estimated item parameters for each CAT subtest are summarized in Table 2.

The Arithmetic Reasoning item bank consisted of 225 items, 148 of which had been calibrated on a selected population of Air Force enlistees (Sympson et al., 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.

The Word Knowledge 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).

The CAT PC item bank consisted of 25 items that had been computer-administered to samples ranging from 239 to 481 Marine recruits. LOGIST was used to obtain 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 CAT PC test session, the paragraph to be read and the question to be answered were presented on separate screens. Thus, unlike the ASVAB PC subtest, examinees were not allowed to refer back to the paragraph while responding to the multiple-choice question.

### Procedure

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

During three months in 1981, the CAT tests were administered to available recruits approximately 24 hours after their arrival at the recruit depot. The tests 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 was located at the University of Minnesota and 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 intro-

Table 2  
Descriptive Statistics for CAT Item Parameters

Subtest and Item Parameter	Lower Limit	Upper Limit	Average
<b>CAT AR</b>			
Discrimination	.716	2.000	1.194
Difficulty	-1.355	1.966	.582
Guessing	.029	.297	.179
<b>CAT WK</b>			
Discrimination	.800	2.690	1.355
Difficulty	-1.980	2.000	-.337
Guessing	.040	.260	.128
<b>CAT PC</b>			
Discrimination	1.000	1.000	1.000
Difficulty	-2.952	1.665	-.026
Guessing	.000	.000	.000

ducing 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 sample mean and standard deviation for each subtest and the AFQT composite.

### Data Analyses

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.

Two factor analyses were performed on the intercorrelation matrix, 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 analysis also included the CAT variables.

Two multiple regression analyses were performed. The first was performed to determine how well the AFQT composite computed from initial

ASVAB subtest scores could be predicted using ASVAB retest AR, WK, PC, and NO scores; and the second, to determine whether it could be predicted as well 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 Sympton et al. (1982) for Air Force jet engine mechanic trainees who took AR and WK subtests administered both in ASVAB and in 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; OSD: MRA&L, 1982; Ree, Mullins, Mathews, & Massey, 1982).

### 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% 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. Based on an inspection of the factor loadings, the four factors were labeled as follows: verbal, technical-mechanical, mathematical-quantitative, and speed. These factors are very similar to those identified in other factor analyses of ASVAB (Fischl, Ross, & McBride, 1979; Ree et al., 1982).

Table 6 presents the varimax rotated factor so-

Table 3  
Means, Standard Deviations, and Range for ASVAB,  
AFQT, and CAT Subtests

Variable	Mean	Std Dev	Lower Limit	Upper Limit
<b>ASVAB Initial Test</b>				
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
<b>ASVAB Retest</b>				
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
<b>CAT</b>				
CAT AR	0.40	0.82	-1.57	2.52
CAT WK	0.59	0.79	-1.63	2.54
CAT 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.

lution for the second analysis, which was performed with the CAT variables added to the data matrix. As shown, CAT WK and CAT PC loaded substantially on the verbal factor, and CAT AR on the mathematical factor. While the structure of the four factors remained essentially the same, the total variance accounted for by each of the factors changed, as could be expected with the addition of the CAT variables. In particular, the mathematical factor became stronger than the technical factor with the addition of CAT AR.

CAT WK loaded higher (.83) than any other

variable on the verbal factor. While CAT PC loaded higher (.54) on the verbal factor than on any other factor, the verbal factor accounted for only 29% of its variance. The final communality estimate showed that the four factors together accounted for 43% of the variance in CAT PC. This suggests that much of the CAT PC variance is unique or unreliable. The latter seems more likely since the CAT PC 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 factor load-

Table 4  
Pearson Correlation Coefficients for  
ASVAB and CAT AR, WK, and PC Subtests

Variable	ASVAB Initial Test			ASVAB Retest		
	AR	WK	PC	AR	WK	PC
ASVAB Retest						
AR	.767					
WK		.771				
PC			.464			
CAT						
AR	.800			.800		
WK		.806			.799	
PC			.507			.505

Note. The full correlation matrix is available from the authors.

Table 5  
Varimax Rotated Factor Matrix for Analysis Using ASVAB Variables

Subtest	Factor 1 (Verbal)	Factor 2 (Technical)	Factor 3 (Math)	Factor 4 (Speed)	Final Communality Estimate
Initial ASVAB					
GS	<u>.60</u>	<u>.44</u>	.31	.07	.66
AR	.31	.21	<u>.72</u>	.17	.68
WK	<u>.82</u>	.16	.23	.08	.76
PC	<u>.55</u>	.09	.33	.10	.43
NO	.04	.12	.22	<u>.68</u>	.53
CS	.12	.00	.06	<u>.72</u>	.54
AS	.10	<u>.83</u>	.04	-.02	.71
MK	.28	<u>.17</u>	<u>.77</u>	.26	.77
MC	.33	<u>.48</u>	<u>.44</u>	.11	.55
EI	.34	<u>.56</u>	.25	.02	.48
ASVAB Retest					
GS	<u>.57</u>	<u>.48</u>	<u>.36</u>	.07	.70
AR	.33	.26	<u>.70</u>	.25	.73
WK	<u>.82</u>	.23	.19	.09	.76
PC	<u>.52</u>	.17	.30	.27	.46
NO	.10	-.08	.20	<u>.56</u>	.37
CS	.05	.04	.06	<u>.73</u>	.54
AS	.07	<u>.85</u>	.05	.02	.73
MK	.35	.16	<u>.73</u>	.27	.75
MC	.26	<u>.61</u>	<u>.44</u>	.13	.65
EI	.33	<u>.62</u>	.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	27.85%	55.64%	82.27%	100.00%	
Total Variance	17.29%	17.25%	16.53%	11.01%	
Cumulative	17.29%	34.53%	51.06%	62.07%	

Note. Factor loadings greater than .35 are underlined.

Table 6  
Varimax Rotated Factor Matrix for Analysis Using ASVAB and CAT Variables

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

Note. Factor loadings greater than .35 are underlined.

ings for CAT PC were comparable to those for initial and retest ASVAB PC, indicating that CAT PC measures reading comprehension as well as its ASVAB counterparts, despite its shorter length.

CAT AR loaded higher (.76) than any other variable on the mathematical factor. The four factors together accounted for 78% of the variance in CAT AR, with the verbal factor explaining 12% of the variance. Thus, while CAT AR is primarily a measure of mathematical ability, verbal ability is also involved in understanding and solving these word

problems. This is true for the ASVAB AR subtest as well.

In sum, the factor loadings for the three CAT subtests were quite similar to those of their ASVAB counterparts. Therefore, the CAT and ASVAB subtests appear to be measuring the same aptitude factors.

### AFQT Regressions

Table 7 presents a summary of the multiple

Table 7  
Summary of Multiple Regression Analyses Performed to Predict  
AFQT Composite Computed from Initial ASVAB Test

Variable	Multiple R	B Weights	Std. Error of B	Beta Weights	F
<b>From ASVAB Retest Subtests</b>					
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			
<b>From CAT Subtests</b>					
CAT AR	.788	7.568	.567	.527	178.151*
CAT WK	.865	6.382	.605	.428	111.464*
CAT 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.

\* $p < .01$ .

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

### Discussion and Conclusions

The preceding analyses are encouraging for the development of CAT as a replacement for the paper-and-pencil ASVAB. These data are notable in that military examinees were used to calibrate the test items and a similar population was sampled for CAT and ASVAB test administrations. CAT was clearly found to be as valid a measure of the abilities tested as were the corresponding ASVAB subtests. First, CAT subtests correlated as highly with initial ASVAB test scores as did the ASVAB

retest scores. Second, 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 ASVAB subtests. Finally, the AFQT composite score was predicted equally well from 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 potentially unique sequence of items that is tailored in difficulty to that examinee, based on the examinee's 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. This finding supports earlier re-

search showing that short adaptive tests had higher reliability coefficients than conventional tests of the same length, with the differences in reliability diminishing as the test length increased (McBride & Martin, 1983).

Although the present results are favorable for the implementation of CAT, this study 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 correlations with training performance criteria.

### References

- Fischl, M. A., Ross, R. M., & McBride, J. R. (1979). *Development of factorially based ASVAB high school composites* (Technical Paper 360). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Fruchter, D. A., & Ree, M. J. (1977). *Development of the Armed Services Vocational Aptitude Battery: forms 8, 9, and 10* (Technical Report AFHRL-TR-77-19). Lackland Air Force Base TX: Personnel Research Division, Air Force Human Resources Laboratory.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- McBride, J. R. (1980). Adaptive verbal ability testing in a military setting. In D. J. Weiss (Ed.), *Proceedings of the 1979 Computerized Adaptive Testing Conference* (pp. 4–15). Minneapolis: Computerized Adaptive Testing Laboratory, Department of Psychology, University of Minnesota.
- McBride, J. R., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 223–236). New York: Academic Press.
- Office of the Assistant Secretary of Defense: Manpower, Reserve Affairs and Logistics (OSD: MRA&L). (1982). *Armed Services Vocational Aptitude Battery (ASVAB) forms 8, 9, and 10: Technical manual*. Washington DC: Author.
- Owen, R. J. (1969). *A Bayesian approach to tailored testing* (Research Bulletin 69-92). Princeton NJ: Educational Testing Service.
- Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70, 351–356.
- Reckase, M.D. (1979, April). *Item pool construction for use with latent trait models*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Ree, M. J., Mullins, C. J., Mathews, J. J., & Massey, R. H. (1982). *Armed Services Vocational Aptitude Battery: Item and factor analysis of forms 8, 9, and 10* (Technical Report 81-55). Brooks Air Force Base TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Sympson, J. B., Weiss, D. J., & Ree, M. J. (1982). *Predictive validity of conventional and adaptive tests in an Air Force training environment* (Technical Report 81-40). Brooks Air Force Base TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Urry, V. (1977). *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.
- Urry, V. (1978). *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.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6, 473–492.
- Weiss, D. J. (Ed.). (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Wood, R. L., Wingersky, M. S., & Lord, F. M. (1976). *LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters* (RM-76-6). Princeton NJ: Educational Testing Service.

### Acknowledgments

*The opinions expressed in this article are those of the authors, are not official, and do not reflect the views of the Department of the Navy. The authors thank J. B. Sympson for his comments on the data analyses and on earlier versions of this manuscript. Technical reports cited may be obtained from the U.S. Department of Commerce, National Technical Information Service, 5285 Port Royal Road, Springfield VA 22161, U.S.A.*

### Author's Address

Send requests for reprints or further information to Kathleen E. Moreno or C. Douglas Wetzel, Navy Personnel Research and Development Center, San Diego CA 92152, U.S.A.