CRM 92-36 / July 1992

Factor Structure and Incremental Validity of the Enhanced Computer-Administered Tests

Neil B. Carey

50 Years
CNA 1992
CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

Work conducted under contract N00014-91-C-0002.

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*

2. REPORT DATE
July 1992

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Factor Structure and Incremental Validity of the Enhanced Computer-Administered Tests

5. FUNDING NUMBERS
C - N00014-91-C-0002
PE - 65153M
PR - C0031

6. AUTHOR(S)
Neil B. Carey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Naval Analyses
4401 Ford Avenue
Alexandria, Virginia 22302-0268

8. PERFORMING ORGANIZATION REPORT NUMBER
CRM 92-36

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commanding General
Marine Corps Combat Development Command (WF 13F)
Studies and Analyses Branch
Quantico, Virginia 22134

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Armed Services Vocational Aptitude Battery (ASVAB), used to select and classify enlisted personnel, is highly correlated to math and verbal content areas. New computerized predictor tests that are sensitive to traits not measured by the current ASVAB subtests may be able to improve predictive validity. This research memorandum investigates the potential of one such group of tests, the Enhanced Computer-Administered tests, to predict performance in the mechanical maintenance specialties.

14. SUBJECT TERMS
Aptitude tests, ASVAB (Armed services vocational aptitude battery), CAT (Computer administered Test), Computer applications, Correlation, JPM (Job performance measurement), Performance (human), Performance tests, Predictions, Reliability, Scoring, Tables (data), Test methods, Validation

15. NUMBER OF PAGES
98

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
CPR	CPR	CPR	SAR

NSN 7540-01-280-5500

Standard Form 298, (Rev. 2-89)
Prescribed by ANSI Std. 239-18
299-01

50 Years
CNA 1992 CENTER FOR NAVAL ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268 • (703) 824-2000

25 August 1992

MEMORANDUM FOR DISTRIBUTION LIST

Subj: CNA Research Memorandum 92-36

Encl: (1) CNA Research Memorandum 92-36, *Factor Structure and Incremental Validity of the Enhanced Computer-Administered Tests*, by Neil B. Carey, Jul 1992

1. Enclosure (1) is forwarded as a matter of possible interest.

2. The Armed Services Vocational Aptitude Battery (ASVAB), used to select and classify enlisted personnel, is highly correlated to math and verbal content areas. New computerized predictor tests that are sensitive to traits not measured by the current ASVAB subtests may be able to improve predictive validity. This research memorandum investigates the potential of one such group of tests, the Enhanced Computer-Administered Tests, to predict performance in the mechanical maintenance specialties.

Donald J. Cymrot
Director
Manpower and Training Program

Distribution List:
Reverse page

Subj: Center for Naval Analyses Research Memorandum 92-36

Distribution List

SNDL

45A2	CG I MEF
45A2	CG II MEF
45A2	CG III MEF
45B	CG FIRST MARDIV
45B	CG SECOND MARDIV
A1	DASN - MANPOWER (2 copies)
A1H	ASSTSECNAV MRA
A2A	CNR
A5	BUPERS
A5	PERS-11
A5	PERS-2
A5	PERS-5
A6	CG MCRDAC - WASHINGTON
A6	HQMC AVN
A6	HQMC MPR & RA Attn: Code M Attn: Code MR Attn: Code MP Attn: Code MM Attn: Code MA (3 copies) Attn: Code MPP-54
FF38	USNA Attn: Nimitz Library
FF42	NAVPGSCOL
FF44	NAVWARCOL
FJA1	COMNAVMILPERSCOM
FJA13	NAVPERSRANDCEN Attn: Technical Director (Code 01) Attn: Technical Library Attn: Dir, Manpower Systems (Code 11) Attn: Dir, Personnel Systems (Code 12) Attn: Dir, Testing Systems (Code 13) Attn: CAT/ASVAB PMO
FJB1	COMNAVCRUITCOM
FT1	CNET
V8	CG MCRD PARRIS ISLAND
V8	CG MCRD SAN DIEGO
V12	CG MAGTEC
V12	CG MCCDC Attn: Studies and Analyses Branch Attn: Director, Warfighting Center Attn: Warfighting Center, MAGTF Proponency and Requirements Branch (2 copies)
V12	CG MCRDAC - QUANTICO
V25	MCAGCC

OTHER

Military Accession Policy Group (8 copies)
Defense Advisory Committee on Military Personnel Testing (8 copies)
Joint Service Job Performance Measurement Working Group (12)

CRM 92-36 / July 1992

Factor Structure and Incremental Validity of the Enhanced Computer-Administered Tests

Neil B. Carey

Operations and Support Division

50 Years CNA 1992

CENTER FOR NAVAL ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

ABSTRACT

The Armed Services Vocational Aptitude Battery (ASVAB), used to select and classify enlisted personnel, is highly correlated to math and verbal content areas. New computerized predictor tests that are sensitive to traits not measured by the current ASVAB subtests may be able to improve predictive validity. This research memorandum investigates the potential of one such group of tests, the Enhanced Computer-Administered Tests, to predict performance in the mechanical maintenance specialties.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The military services use the Armed Services Vocational Aptitude Battery (ASVAB) to select and classify enlisted personnel. The ASVAB contains ten subtests that measure four general abilities: verbal, mathematical, technical, and speed. This memorandum investigates the Enhanced Computer-Administered Tests (ECAT), a new battery of computer-administered tests that differs in content from the ASVAB. Each new test is judged first on the ability of that test to improve the ASVAB's ability to predict mechanical performance and, second, on whether significant gains were made upon retest.

METHOD

Examinees were 698 first-term automotive mechanics (MOS 3521) and 443 helicopter mechanics (MOS 6112, 6113, 6114, and 6115). On the first day, each examinee completed eight hours of hands-on testing; on the second day, each completed a job knowledge test, a computerized adaptive testing version of the ASVAB (CAT-ASVAB), the psychomotor portion of the General Aptitude Test Battery (GATB), and the ECAT. More than 130 examinees were retested 10 to 14 days later to determine the reliability of the tests over time.

The ECAT consists of nine subtests designed to assess areas that are not presently represented by the ASVAB. The examinees in this study took eight of the nine. A subtest called sequential memory (SM), consisting of 35 items, assessed memory. Four subtests assessed spatial abilities: spatial reasoning (SR), with 30 items; integrating details (ID), with 40 items; assembling objects (AO), with 32 items; and spatial orientation (SO), with 24 items. A sixth subtest, target identification (TI), measured perceptual speed and accuracy; it contained 39 items. And last, two subtests measured general hand-eye coordination: one-hand tracking, with 18 nonpractice items, and two-hand tracking, with 18 nonpractice items. The examinees did not take the ninth ECAT subtest, mental counters, because of time constraints; however, they did take the psychomotor portions of the GATB for comparative purposes.

RESULTS

Test-Retest Gains

Findings of significant test-retest gains are a concern because gains indicate that coaching or practice might influence the validity of the test. Test-retest gains were statistically significant for the hand-eye coordination measures (OT and TT) and for the response times for target identification. SM and SO also showed significant gains.

Incremental Validity

Incremental validity is presented for two types of correlations. Sample values refer to observed correlations; in contrast, range-corrected values adjust for the fact that sample values are restricted because the ASVAB was used to select enlisted personnel for the Marine Corps. For automotive mechanics, the spatial composite added incremental validity of 0.032 (6.5 percent) against enlistment ASVAB, using sample correlations (table I). The AO test was by far the most important contributor to this increment, at 0.036. The addition of the other three spatial tests (SR, SO, and ID) actually detracted from the incremental validity of AO alone. The incremental validity of AO was also greater than for the combined validity of the entire ECAT battery and the GATB. For the automotive mechanics, increments for AO ranged from a high of 0.036 (7.3 percent), using sample values and comparing against the paper-and-pencil ASVAB given at enlistment, to a low of 0.012 (1.6 percent), using range-corrected correlations and comparing against the concurrently administered, computer-adaptive ASVAB (CAT-ASVAB).

The incremental validities for helicopter mechanics showed a similar pattern. The incremental validity of AO was greater than the validity of the entire spatial composite. AO also had more incremental validity than did the combination of the four ECAT composites and the GATB. The incremental validity of AO for helicopter mechanics ranged from a high of 0.024 (4.2 percent) using sample correlations versus enlistment ASVAB to a low of 0.015 (2.2 percent) using range-corrected correlations versus CAT-ASVAB.

Stepwise Regressions

The CNA analyst conducted stepwise regressions to determine whether ECAT subtests would be candidates to replace some elements of the Marine Corps' current mechanical maintenance (MM) composite. The current MM composite is AS + MC + AR + EI.[1] The analyses showed that the auto shop subtest (from the current ASVAB) and AO (from the ECAT) were usually the earliest components of a predictor composite formed using stepwise regression. MC, from the ASVAB, was the strongest subtest of CAT-ASVAB for predicting performance of helicopter mechanics. These findings indicated that two present components of the MM composite, AS and MC, were highly effective at predicting mechanical hands-on performance. The findings also showed that AO would probably contribute to predictive validity of the current MM composite more than do the current ASVAB components, AR and EI.

1. AS is auto shop, MC is mechanical comprehension, AR is arithmetic reasoning, and EI is the electronics information subtest of the ASVAB.

Table I. Incremental validity for psychomotor measures and ECAT above ASVAB and time-in-service base (automotive mechanics)

	Sample values				Corrected values		
	df	SS	MSE	adj R	df	SS	adj R
Total	697	40,502.65	—	—	697	57,759.80	—
Enlistment ASVAB							
+ TIS, TIS squared	12	10,455.88	43.86	.495	10	27,713.12	.687
+ spatial (SR+AO+SO+ID)	1	1,348.59	41.96	.032***	1	1,348.52	.017
+ coordination (OT+TT)	1	190.04	43.65	.004*	1	190.07	.002
+ memory (SM)	1	367.52	43.39	.008**	1	367.55	.004
+ perceptual speed (MDECCORR)	1	105.42	43.77	.002	1	105.43	.001
+ dexterity (GATB)	1	267.47	43.54	.006*	1	267.42	.003
+ all 5 composites (above)	5	1,543.71	41.92	.033***	5	1,543.64	.017
Spatial							
+ SR	1	331.94	43.44	.007**	1	331.97	.004
+ SO	1	278.36	43.52	.006*	1	278.33	.003
+ AO	1	1,521.74	41.70	.036***	1	1,521.73	.019
+ ID	1	733.73	42.86	.017***	1	733.74	.009
Coordination							
+ OT	1	112.57	43.77	.002 N.S.	1	112.54	.001
+ TT	1	214.91	43.61	.005*	1	214.90	.002
Target identification							
TIACC	1	237.92	43.58	.005*	1	237.97	.003
GEOTOT	1	217.08	43.61	.005*	1	217.11	.002
GEODEC	1	87.38	43.80	.001	1	87.37	.001
TIMOVMED	1	428.34	43.30	.010**	1	428.36	.005
GATB finger dexterity	1	329.75	43.45	.007**	1	329.73	.004
CAT ASVAB							
+ TIS, TIS squared	12	14,039.73	38.63	.579	10	31,165.68	.730
+ spatial (SR+AO+SO+ID)	1	878.88	37.40	.018***	1	915.23	.011
+ coordination (OT+TT)	1	30.61	38.64	.000	1	34.30	.000
+ memory (SM)	1	239.94	38.34	.004*	1	244.32	.003
+ perceptual speed (MDECCORR)	1	35.45	38.64	.000	1	38.78	.000
+ dexterity (GATB)	1	68.10	38.59	.001	1	71.02	.001
+ all 5 composites (above)	5	953.75	37.51	.016***	5	988.49	.010
Spatial							
+ SR	1	249.95	38.32	.005**	1	270.20	.003
+ SO	1	158.83	38.46	.003*	1	175.17	.002
+ AO	1	1,007.63	37.22	.021***	1	1,005.64	.012
+ ID	1	469.32	38.00	.009***	1	498.77	.006
Target identification							
TIACC	1	138.36	38.49	.002	1	140 39	.002
GEOTOT	1	65.41	38.60	.001	1	70.90	.001
GEODEC	1	25.01	38.65	.000	1	27.71	.000
TIMOVMED	1	169.08	38.44	.003*	1	171.37	.002
GATB finger dexterity	1	171.22	38.44	.003*	1	182.49	.002

NOTES: Significance tests were performed only for sample values. * = $p < .05$, ** = $p < .01$, *** = $p < .001$. TIS was a constant for range correction, so base model df is 10 for corrected values. Negative adjusted r's were set to zero. Dexterity is a GATB composite; Finger Dexterity is one component of the Dexterity composite.

Factor Structure of ASVAB With AO

The analyst conducted factor analyses to determine whether AO would change the factor structure of ASVAB. The current dimensions of the ASVAB are technical, verbal, mathematical, and speed. When AO was added to the CAT-ASVAB, AO and MC defined a fourth, spatial factor that was related to the ability to visualize relationships among objects. When AO was added to the enlistment ASVAB, AO defined its own factor. The CAT- and enlistment-ASVAB findings indicate that the addition of AO might add a new dimension to the structure of the ASVAB.

CONCLUSIONS

The results from the retest gains have implications for the usefulness of some ECAT subtests. Significant retest gains indicated that those who have had practice on the tests will have an unfair advantage. The fact that there were retest gains indicates that these tests are sensitive to practice and, hence, might be coachable. If the tests are coachable, they would have limited use as a tool for military selection.

The incremental validity results also have implications for ECAT. These results indicated that, for the purpose of predicting mechanical job performance, AO was the only test that showed much promise for improving the current ASVAB. AO might improve the current MM composite; AO also showed no retest gains that would impair its usefulness in military selection. Factor analyses indicate that AO might change the ASVAB by adding a spatial dimension that reflects the ability to visualize objects.

CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

TABLES

INTRODUCTION

The Department of Defense is considering methods to enhance the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB, which is used by the military services to select and classify enlisted personnel, contains ten subtests that measure four general abilities: verbal, mathematical, technical, and speed. The Marine Corps uses four aptitude composites, computed from the ten ASVAB subtests, to classify recruits into clusters of military occupational specialties (MOSs) that are most suited to their abilities.

Earlier analyses have confirmed the four general aptitudes of the ASVAB [1], although the factors tend to be somewhat correlated. The correlations suggest that the ASVAB is limited in the number of dimensions that it measures. Because military jobs are multidimensional and require a wide range of skills and abilities, the ASVAB may not be able to predict some of these necessary qualities [2, 3]. Therefore, tests that measure these new dimensions might supplement the existing ASVAB and improve the overall selection and classification system.

When looking at new predictors, one must carefully consider the performance measures against which the new tests are to be validated. Traditionally, the ASVAB has been validated against training grades. The ASVAB is good at predicting training grades because of the shared academic abilities it measures. Training grades are often based on paper-and-pencil knowledge tests, and people who do well on that type of test also perform well on the paper-and-pencil ASVAB. Given the similarities between ASVAB and training grades, new predictors are not likely to improve the ASVAB-training grade relationship across a variety of jobs or clusters.

The joint-service Job Performance Measurement (JPM) project offers an opportunity to validate new predictor tests. A primary purpose of the JPM project has been to develop objective and standardized measures of job performance that reflect the broad range of military job requirements. The expanded scope of the hands-on performance tests will measure those unique abilities that are needed in the work setting but are not necessarily required for academic success.

To derive useful conclusions regarding the Enhanced Computer-Administered Test (ECAT) battery, it should be related to a comparable version of the ASVAB. When the new predictors from the ECAT are compared to ASVAB scores of record, there are two alternative explanations to any improvement in validity from the ECAT: First, scores of record are older, so the new predictors might be better merely because they were taken concurrently. Second, unlike the ECAT scores, scores of record were derived from ASVAB tests taken with paper and pencil. To permit a more useful analysis of ECAT's potential, the analyst compared it to a computerized adaptive-testing version of the ASVAB (CAT-ASVAB) that was administered concurrently.

This memorandum investigates the potential of the ECAT to improve the prediction of mechanical performance beyond what the ASVAB is able to achieve. As part of the JPM project, the examinees took eight ECAT subtests. These included a measure of sequential memory (SM), consisting of 35 items and four measures of spatial ability; spatial reasoning (SR), with 30 items; integrating details (ID), with 40 items; assembling objects (AO), with 32 items; and spatial orientation (SO), with 24 items. A sixth subtest, target identification (TI), with 39 items measured perceptual speed and accuracy. The last two subtests measured coordination: one-hand tracking (OT), with 18 nonpractice items, and two-hand tracking (TT), with 18 nonpractice items. The examinees did not take a ninth ECAT subtest, mental counters, because of time constraints. They did, however, take the psychomotor portions of the General Aptitude Battery (GATB) for comparative purposes.

The CNA analyst compared the increments in validity due to these new tests relative to the complete battery of ASVAB subtests. He examined two sources of aptitude scores: an ASVAB administered at the time of enlistment and a concurrent, computer-adaptive test (CAT) ASVAB administered as part of the JPM project. The analyst computed reliability estimates for both the predictors and criteria in addition to the absolute incremental validities of each new predictor test; he also computed uniqueness estimates for each new predictor test.

TECHNICAL CONSIDERATIONS FOR ASSESSING INCREMENTAL VALIDITY

Sampling

The analyst computed correlation matrices of all ASVAB subtests, ECAT subtests, and the General Aptitude Test Battery (GATB) based on complete data for both enlistment and CAT aptitude scores. Because the number of women in the original testing sample was small, women were excluded from all analyses. Fewer than 50 women were in the original sample.

Correction for Range Restriction

The mechanics who participated in the study are a selected sample. To account for these selection effects, the analyst corrected the sample correlations for range restriction using a multivariate procedure based on all ten enlistment ASVAB subtests [4]. The 1980 youth population served as the reference population from which all corrections were derived [5].

Shrinkage of Multiple Correlations and Cross-Validation

Multiple correlations (MRs) are extensions of simple correlation coefficients in that the criterion is regressed, not on one predictor measure, but on several. Regression weights are assigned to each predictor to maximize the multiple correlation for the sample on which the regression is computed. If the regression weights are then applied to a different sample, the resulting MR will almost always be smaller than the MR obtained in the original sample. This is because the original optimal weights "overfit" the original sample, fitting some error of measurement and outliers [6]. This decrement in MRs is referred to as "shrinkage."

Formula methods have been derived to estimate the degree of shrinkage in MRs. These formulas make use of all observations and result in more precise estimates of the shrinkage. The analyst did not use formulas for shrinkage in this study, however, because formulas rely on the false assumption that the same regression weights computed on the samples in this study would be used to create a composite--including negative weights. Furthermore, even if that assumption were true, the sample sizes used in this study are so large that the effect of correction would be fairly small. The maximum bias introduced by failure to make the correction for shrinkage is about 0.1/N [7]. In earlier work on infantry data [2], the effect of corrections for shrinkage was very small.

Criterion Unreliability

All performance criteria are not measured with the same reliability. To the extent that the criteria are unreliable and contain measurement error, estimates of validity coefficients will also be affected. Theoretically, a test cannot correlate with another variable more highly than it correlates with its own true score (a test score measured with no error); therefore, test validity cannot exceed the square root of test reliability.

Corrections can be made to compensate for unequal measurement reliability [8]. Such corrected values are the maximum coefficients that are obtainable if all measurement error could be eliminated, i.e., perfect criterion reliability. To obtain the proper correction, an accurate estimate of criterion reliability is essential.

The primary objective of this study is to make relative comparisons among validity gains for new predictors within a criterion, not to make absolute comparisons of the magnitude of validity increments across criteria. Therefore, the analyst did not compute corrections to validity coefficients for criterion unreliability. However, the study provides enough information to allow such corrections to be calculated.

Controlling for Time in Service

Validities may be adversely affected by a time lapse between the administration of the enlistment predictors and the new predictors of interest. To account for the possible effect of time differences, the ASVAB was readministered so that all predictor information would be collected at the same time and under the same conditions.

However, the examinees also differed with respect to their length of service, ranging from 8 to 160 months for the automotive mechanics and from 9 to 128 months for the helicopter mechanics. Such differences in amount of experience may affect performance on the predictor tests and/or the performance tests simply due to on-the-job experience, training, or maturity. To control for these potential developmental effects, the analyst used time in service (TIS) and its square as covariates to correct the correlation matrices. In this manner, he statistically adjusted performance scores as if all examinees had the same number of months in service.

Factor Analysis

The analyst conducted factor analyses to determine which groups of ECAT subtests should be entered together for incremental validity analyses. Because the mechanics who participated in this study are a selected sample, the analyst used the corrected correlation matrices as the basis for all factor analyses. Consistent with [9], he used an iterative principal factors procedure with a promax rotation. He used squared multiple correlations as the initial estimates of communalities. The analyst conducted separate analyses for both the enlistment ASVAB and CAT-ASVAB scores. In addition to comparing factor loadings, he noted estimates of communality and interfactor correlations. He conducted multiple analyses with the solutions constrained to have 4 to 7 factors.

Incremental Validity Analysis

The analyst computed increments in validity as the difference between two validity coefficients (increments in sums of squares are also presented). He then computed the multiple correlation between all ASVAB subtests, time in service and its square, with the hands-on performance test for the automotive and helicopter mechanics; these correlations provided the base against which increments in validity by the ECAT and GATB measures would be judged. Time in service was required in the regression for the sample values because most of the personnel who took the tests had significant amounts of service.

To reduce capitalization on chance due to the number of significance tests to be made, the analyst aggregated the five psychomotor tests and then entered them as separate variables into the regression. Consistent with the results of the factor analysis and prior expectations, he computed a spatial variable as the sum of standardized SR, AO, SO, and ID. Memory was simply the standardized sequential memory score. Consistent with the analyses performed by Mayberry and Divgi [9], the analyst computed a coordination variable by standardizing and summing the scores for the one-hand and two-hand tests. He also aggregated the three psychomotor scores from the GATB according to documented procedures [9].

The analyst based all comparisons of multiple correlations on estimates adjusted for the number of predictors.

TEST ADMINISTRATION

Each Marine took tests for two days. The first day was devoted to hands-on testing and the second to written tests. All tests were administered by retired Marines who had received extensive training in how to administer tests in a standardized manner and accurately score and record test performance. The administrators specialized in giving either the hands-on tests or the written and computer-administered tests. To monitor the scoring consistency and accuracy of test administrators throughout the four-month testing period, several different administrators rated the performance of selected examinees.

Examinees were Marine Corps automotive and helicopter mechanics who were tested as part of the JPM project. The CNA analyst used complete data only for the correlational and incremental validity analyses; these data resulted in a sample size of 698 for automotive mechanics (MOS 3521) and 443 for helicopter mechanics. Only males were used in these analyses because fewer than 50 females were part of the original data set. Helicopter mechanics included MOSs 6112 (CH-46), 6113 (CH-53A/D), 6114 (U/AH-1), and 6115 (CH-53E). Seventy-two automotive mechanics and 63 helicopter mechanics repeated the test after an interval of 10 to 14 days.

Criterion Measure

Hands-on performance tests (HOPTs) were developed for the automotive mechanics and each of the four helicopter specialties. The domain of job requirements was specified based on official Marine Corps publications, training materials, and extensive task analyses by the job experts [10]. Tasks were organized by function (i.e., inspect, service, troubleshoot, remove and replace, test, align/adjust) and by system. For automotive mechanics, systems were defined by the vehicle (LVS, M998, M1008, M813, or M923) and the subsystem (brake, electrical, drive train, diesel engine, steering, or hydraulic system). For helicopter mechanics, the systems were flight control systems, rotor systems, power train, and power plant systems [10]. Although separate tests were developed for each of the four helicopter mechanic MOSs, there was considerable overlap in the skills, functions, and systems assessed by the four tests. Tasks were sampled from the most representative systems so that hands-on test scores would generalize to the full range of mechanic duties [11].

Predictor Tests

The CNA analyst used eight ECAT subtests for this research. The subtests and their editing rules are described below.

Sequential memory [12]

Sequential memory was designed as a test of memory capacity or efficiency. In the first section, the computer presents either three or four dots in the center of the screen. Each dot represents a particular single digit number, shown below the dot. For example, in figure 1,[1] the dots represent 4, 7, and 8, respectively.

Left	Center	Right
.	.	.
4	7	8

Figure 1. Example screen for sequential memory test

Each Marine was given three seconds to memorize the numbers, after which the numbers disappeared from the screen. Then, in random order, the dots were briefly changed into an "X." The appearance of the "X" called the number previously below the dot, and the Marine was asked to recall the sequence of calls. Each item involved five calls and a unique set of numbers. For example, using figure 1, if the five Xs replaced the dots in this order, center-right-right-left-center, the correct answer would be 7-8-8-4-7. Therefore, this task required the examinee to be able to (a) remember the numbers assigned to the dots and (b) remember the sequence in which the numbers were called. The subject was required to type in the correct sequence of five digits.

The second part of the test presented an additional challenge. After the numbers had been assigned to the dots and the sequence of calls had been completed, the Marines were told to convert the numbers-in-memory to specified new values, while maintaining the correct order.

There were 35 sequential memory items. The proportion of digits correct across all nonpractice items was used as a summary score. Practice problems, including feedback on accuracy, preceded each part of the test. (The first ten items in each section were practice problems.) The Marines received no feedback during the actual test. In editing, sequential memory was set to missing if fewer than 75 percent of items were attempted or if none of the items on the first two of the four performance levels were answered correctly.[2] This editing rule was

1. A figure is presented for sequential memory, and not for other ECAT subtests, because sequential memory is the most difficult to understand without seeing a picture of a hypothetical item.
2. Editing rules were varied by subtest because previous investigators had found that differing rules were warranted on the basis of their research findings. Furthermore, differing editing rules also reflected the fact that different institutions were responsible for developing different subtests in the ECAT. For example, the Army Research Institute was primarily responsible for one-hand and two-hand tracking, whereas Navy Personnel Research and Development Center (NPRDC) was primarily responsible for sequential memory and integrating detail.

established on the basis of recommendations from Navy Personnel Research and Development Center (NPRDC) [13], which developed the sequential memory test and used it in earlier validation research.

Spatial reasoning

The spatial reasoning test contained 30 items. For each item, Marines were shown a series of four figures. The task was to identify the pattern or relationship among the figures and then to identify from among the five possible answers the one figure that appeared next in the series. In editing, the spatial reasoning score was deleted if the total score was zero or if fewer than 75 percent of items were attempted. This rule was adopted in order to be consistent with the rule for sequential memory.

Integrating details

This 40-item measure [14] was a systematic test of spatial ability. First, puzzle elements were presented alone. The examinee was allowed as much time as needed to correctly connect the matching elements to form a complete object and store the object in memory. When the Marine had formed and remembered the object, he was to press any key and the puzzle pieces would be replaced by the alternative. He then had as long as necessary to decide whether the object in memory matched the one presented, and then to select either a *same* or *different* response. Therefore, each item produced three dependent measures: time spent synthesizing the puzzle elements (integrate time), time spent deciding whether the presented alternative was correct (decision time), and response accuracy (accuracy). Of these three, accuracy was the recommended operational score. This integrating detail test was set to missing if fewer than 75 percent of the items were attempted, if fewer than 40 percent of the items attempted were correct, or if the examinee had to be warned six or more times that he was not spending enough time on items. This editing rule was adopted on the recommendation of researchers at NPRDC; these researchers developed the test and have had experience administering it.

Assembling objects

This 32-item test measured the ability to visualize how an object would look when its parts were put back together again. There were two types of problems in the test. In one part, the item showed a picture of labeled parts. By matching the letters used to label parts of the elements, the examinee could "see" where the parts should touch when the object was put together correctly. The second type of problem did not label any of the parts. The parts fit together like the pieces of a puzzle. In each section, the examinee was shown four possible figures and asked to pick the correct one. Scores on the assembling objects test were set to missing if the Marine attempted fewer than 75 percent or if he answered no items correctly. This rule was used to be consistent with the editing rule for sequential memory (see above).

Spatial orientation

This 24-item test measured the ability to mentally rotate figures. Each item contained a picture within a circular or rectangular frame. The bottom of the frame had a circle with a dot inside it. The picture or scene was not in an upright position. The task was to mentally rotate the frame so that the bottom of the frame was positioned at the bottom of the picture. The examinee then had to decide where the dot would appear in the circle. Scores on the spatial orientation test were set to missing if fewer than 75 percent of the items were attempted or if no items were answered correctly. This rule was used in order to be consistent with the rule used for sequential memory (see above).

Target identification

This 36-item test was a measure of perceptual speed and accuracy. The Marine was shown a target object and three stimulus objects. The objects were pictures of military vehicles or aircraft (e.g., tanks, planes, helicopters). The target object, which was the same as one of the stimulus objects, could be rotated in relation to its stimulus counterpart. The Marine had to determine which of the three stimulus objects was the same as the target object and then press the button on the response pedestal that corresponded to that choice. The primary dependent variable was the Marine's average decision time across all trials in which he made a correct response that was consistent with earlier investigators [15,16]. Nevertheless, the analyst investigated several alternative measures as part of this research:

- TIACC = Percentage correct, standardized to mean 50 and standard deviation 10 in the sample of examinees.
- MDECCORR = Mean decision time for correct items.
- GEOTOT = Geometric mean of total time (includes decision and movement time).
- GEODEC = Geometric mean of decision time.
- TIMOVMED = Median of movement times.
- TIMEAN = Clipped mean of decision times. For this variable, the items were separated into "hard" and "easy" items. The highest and lowest decision times within the hard and easy items were set to missing. Separate means then were computed for the hard and easy items. The mean of these two values was the TIMEAN.

All these measures showed that those who were faster did somewhat better on the hands-on criterion. Initial analyses showed that three measures--GEOTOT, TIMOVMED, and MDECCORR--had slightly higher correlations with the total hands-on score than did the other measures. Target identification scores were set to missing if less than one-third of the items were answered correctly. The analyst used this editing rule to be consistent with investigators at the Army Research Institute who had worked with these test scores previously.

One- and two-hand tracking

One- and two-hand tracking assessed general hand-eye coordination. The test required a special response pedestal with a joy stick for the one-hand tracking and two sliding knobs (one moving vertically, the other horizontally) for two-hand tracking. For both tests, a cursor moved in a random pattern across the screen. The object of the test was to use the joy stick or the sliding knobs to move a crosshair so as to minimize the distance between the moving cursor and the cross-hair. Each test consisted of 24 trials, the first three of which were practice trials. In this recent test, the computer designated the next three trials as "real," but previous investigators had considered those three to be practice as well. Thus, to be consistent with the work of previous investigators, only the last 18 trials were used for the present analyses. Because of the controlled presentation of the tests, no time limits were necessary. Scores of the tests were the average root mean square distance of the crosshair from cursor over all trials standardized to a mean of 50 and a standard deviation of 10 in the sample of examinees. One-hand and two-hand tracking scores were set to missing if fewer than 90 percent of the items were attempted. The analyst used this editing rule to be consistent with investigators for Army Project A who had worked with these scores previously.

The editing rules for the ECAT resulted in small percentages of scores being set to missing (table 1). In addition, five subtests from the GATB were administered. The five subtests were combined into three aptitude composites, which were then used as the unit of analysis for this study [9]. The three aptitude scores and the constructs they measured were:

- Motor coordination: the ability to make precise movements with hands and fingers

- Finger dexterity: the ability to move fingers and manipulate small objects with the fingers rapidly and accurately

- Manual dexterity: the ability to move the hands easily and skillfully.

Table 1. Numbers and percentages of ECAT scores set to missing

	Total scores	Total set to missing	Percentage set to missing
Sequential memory	1,585	53	3.34
Spatial reasoning	1,588	32	2.02
Integrating details	1,579	42	2.66
Assembling objects	1,586	49	3.09
Spatial orientation	1,581	1	0.06
Target identification	1,586	1	0.06
One-hand tracking	1,588	5	0.31
Two-hand tracking	1,586	2	0.13

NOTE: The total number varies because some respondents did not finish all ECAT subtests. Furthermore, a very small number (fewer than 15 for any subtest) of subtest scores were unreadable, and, hence, were not considered part of the total number available.

RESULTS

Reliability Estimates

Appendix A and table 2 present the reliability estimates for the criterion measures and the ECAT new predictor tests. Where possible, the analyst computed the following reliability estimates:

- Test-retest: Alternate forms of the hands-on test and the same form for the new predictors were readministered to about 20 percent of the sample group after a period of 10 to 14 days.
- Alpha coefficient: A measure of internal consistency of test items (or tasks) that reflects the degree to which item responses are homogeneous.
- Scorer agreement: The percentage of agreement between two test administrators as they observe and score the step-level performance of one examinee. (This was applicable to the HOPT only).

Table 2. Reliability estimates for ECAT subtests

Reliability measure	ECAT subtest measure (number of examinees for retest)												
	SM (131)	SR (134)	ID (123)	AO (133)	SO (135)	OT (134)	TT (134)	TIACC[a] (135)	GEOTOT[b] (135)	GEODEC[c] (134)	MDECCORR[d] (135)	TIMEAN[e] (135)	TIMOVMED[f] (135)
Test-retest	.74	.80	.79	.75	.73	.76	.73	.52	.80	.78	.80	.78	.67
Split-halves	.84	.88	.75	.84	.86	.95	.96	.64	.96	.96	.84	.93	.98
Alpha coefficient	.90	.93	.85	.91	.92	.97	.97	.73	.97	.98	.97	.98	.94

NOTES: The number of examines for the split-halves and alpha coefficient estimates is 1,141. All estimates have been corrected for restriction of range. TIACC, GEOTOT, GEODEC, MDECCORR, TIMEAN, and TIMOVMED all refer to the target identification task.

a. Average accuracy score.
b. Geometric mean of total time.
c. Geometric mean of decision time.
d. Mean decision time for items answered correctly.
e. Clipped mean decision time. Split-halves reliability for TIMEAN is the correlation of the easy items with the hard items.
f. Median of the movement time.

- Split-halves: The correlation of the odd-numbered items (or tasks) with the even-numbered items.

The hands-on tests were found to be very reliable [11]. Test-retest reliability was 0.79 for automotive mechanics and 0.88 for MOS 6112 (CH-46 mechanics) [11]. Because of practical constraints [11], test-retest data were available for MOS 6112 only. Alpha coefficients were consistently high for all MOSs. Test administrators also agreed on the scoring of the performance that they observed.

Given that the ECAT subtests were somewhat shorter in length, their reliabilities tended to be slightly lower than those of the criterion measures (hands-on performance or job knowledge test). Table 2 shows that test-retest estimates were moderate for all measures except for the accuracy of target identification (TIACC), which was lowered because of a severe ceiling effect, with over 90 percent of the responses correct. Appendix B gives the raw reliabilities. The large percentage of accurate responses was a desirable characteristic for this measure (TIACC) because mean decision time for correct responses (MDECCORR) was the criterion of interest. Otherwise, MDECCORR would be based on a smaller sample of items.

An important consideration in evaluating the ECAT subtests is whether there are significant performance gains upon retest. Table 3 shows the magnitude of standardized gains when the examinees were retested tested 10 to 14 days after the initial test (appendix C shows the magnitude of raw gains). Retest gains were statistically significant for SM, SO, GEOTOT, GEODEC, OT, and TT (table 4). Such gains over 10 to 14 days may reflect the positive effect of practice on the performance of these ability tests or simply a better understanding of the testing procedures. The size of these gains can be compared with those of the CAT-ASVAB. For this sample, the average change was a loss of about 0.42 standard score points for the eight power CAT-ASVAB subtests. There were no significant gains upon retest for the eight power subtests, although there were significant gains for the speeded tests NO and CS.

Estimates of New Predictor Uniqueness

A necessary, but not a sufficient, condition for new predictors to demonstrate increments in validity is that the new tests need to measure aptitudes that are somewhat unique relative to the ASVAB. Predictors that have high correlations with ASVAB can improve validity only by enhancing test reliability, which is unlikely given the already high ASVAB reliabilities. New tests that measure unique aptitudes have greater potential for incremental validity.

Table 3. Test-retest changes in standardized ECAT subtest scores (both automotive and helicopter mechanics)

Construct	Subtest measure	Initial test n	Initial test Mean	Initial test Std. dev.	Retest n	Retest Mean	Retest Std. dev.	Gain (decrement) vs. initial std. dev. (percent)	Test-retest reliability
Memory	SM[a]	132	50.3	9.5	133	51.7	10.0	14.7	.75
Spatial	SR[b]	134	50.6	9.8	135	51.9	9.1	13.3	.80
	ID[c]	131	49.6	10.3	127	49.9	10.0	2.9	.79
	AO[d]	133	50.6	10.1	134	49.9	11.3	(6.9)	.75
	SO[e]	135	50.4	10.5	135	52.7	11.1	21.9	.73
Perceptual accuracy/speed	TIACC[f]	135	50.0	9.0	134	51.2	10.2	13.3	.52
	GEOTOT[g]	135	47.5	9.0	135	45.6	9.5	21.1	.80
	GEODEC[h]	135	47.8	9.2	135	45.9	9.8	20.7	.79
Coordination	OT[i]	134	49.4	10.6	134	46.6	10.4	26.4	.75
	TT[j]	134	48.2	11.0	134	45.8	11.6	21.8	.74

NOTE: Retest reliabilities were corrected for range restriction. All variables were standardized to mean 50, standard deviation 10 based on the sample scores for all examinees (including both the examinees that were retested and those that were not).

a. SM = sequential memory (35 items), percentage correct, standardized to mean 50, standard deviation 10.
b. SR = spatial reasoning (30 items).
c. ID = integrating details (40 items).
d. AO = assembling objects (32 items).
e. SO = spatial orientiation (24 items).
f. TIACC = percentage of target identification items answered correctly (36 items).
g. GEOTOT = geometric mean of total time for target identification items.
h. GEODEC = geometric mean of decision time for target identification items.
i. OT = mean log distance from the target for one-hand tracking (18 items).
j. TT = mean log distance from the target for two-hand tracking (18 items).

Table 4. Paired T-test for test-retest differences

Subtest measure	n	Mean difference	Standard error	T ratio
SM	131	1.52	.69	2.19*
SR	134	1.20	.65	1.84
ID	123	.38	.72	.52
AO	133	-.65	.77	-.85
SO	135	2.31	.81	2.87**
TIACC	134	1.29	.83	1.55
GEOTOT	135	-1.90	.57	-3.34***
GEODEC	135	-1.96	.57	-3.41***
OT	134	-2.79	.68	-4.14***
TT	134	-2.42	.79	-3.08**

NOTES: * = $p < .05$, two-tailed; ** = $p < .01$; and *** = $p < .001$. Mean differences are negative for GEOTOT, GEODEC, OT and TT because examinees became faster and had less error on retest. T-ratios are shown for informational purposes only.

The uniqueness (U) of a new test is defined as the reliable variance of the test that is not related to ASVAB:

$$U = Rel(NP) - R^2(NP, ASVAB) ,$$

where Rel(NP) is the reliability of the new predictor test (NP) and R^2(NP, ASVAB) is the squared multiple correlation for the regression of the new predictor test on all ASVAB subtests adjusted for shrinkage.

Table 5 shows the estimates of uniqueness for each new predictor test. The analyst computed these estimates based on both enlistment and concurrent ASVAB using retest or split-halves as the measure of reliability. Note that uniqueness estimates appear larger when split-halves reliability is used.

There is little difference in the uniqueness estimates based on enlistment and concurrent aptitude. The OT, TT, GEOTOT, GEODEC, MDECCORR, and TIMEAN showed the highest uniqueness estimates because these measures had high reliability and little relationship with the ASVAB subtests. The other ECAT subtests showed moderate levels of uniqueness. From the uniqueness perspective, the coordination tasks (OT and TT) and the response latencies (GEOTOT, GEODEC, MDECCORR, and TIMEAN) were the best candidates for possible use in improving the validity of ASVAB against mechanical job performance.

Table 5. Uniqueness estimates[a] for new predictor tests relative to enlistment and concurrent aptitude scores

New predictor test	Aptitude scores: Enlistment Retest	Split-half	Concurrent Retest	Split-half
SM	.38	.48	.34	.44
SR	.27	.35	.25	.33
ID	.27	.23	.28	.29
AO	.34	.43	.29	.38
SO	.29	.42	.30	.43
TIACC	.47	.59	.47	.59
OT	.55	.74	.52	.71
TT	.49	.72	.48	.71
GEOTOT	.55	.71	.54	.70
GEODEC	.58	.76	.57	.75
MDECCORR	.59	.63	.57	.61
TIMEAN	.58	.73	.57	.72

NOTES: Multiple, adjusted R^2 of these tests regressed on all ASVAB subtests, used for computing uniqueness estimates, were as follows: SM (.36 for enlistment ASVAB, .40 for CAT); SR (.53 for enlistment, .55 for CAT); ID (.52 for enlistment, .51 for CAT); AO (.41 for enlistment, .46 for CAT); SO (.44 for enlistment, .43 for CAT); TIACC (.05 for enlistment, .05 for CAT); GEOTOT (.25 for enlistment, .26 for CAT); GEODEC (.20 for enlistment, .21 for CAT); OT (.21 for enlistment, .24 for CAT); TT (.24 for enlistment, .25 for CAT), MDECCORR (.21 for enlistment, .23 for CAT); TIMEAN (.20 for enlistment, .21 for CAT). Reliabilities and multiple correlations were corrected for range restriction.

a. Uniqueness is the reliability minus the squared multiple correlation. Two estimates are used, one based on the test reliability and one based on split-half reliability.

Intercorrelations and First-Order Validities

The analyst examined the intercorrelations among the new predictors and ASVAB to determine the degree to which the tests measured the same concept. Tables 6 and 7 show the corrected correlations for the entire sample for enlistment ASVAB and CAT-ASVAB, respectively. Note that the

Table 6. Corrected correlation matrix of enlistment ASVAB, ECAT, and GATB subtests (n = 1,141)

																						TI	
	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	DEX	FING	MOT	AFQT	SM	SR	ID	AO	SO	OT	TT	CORR	TIMOV
GS	1.00	.72	.80	.69	.52	.45	.64	.70	.70	.76	.31	.27	.21	.83	.41	.60	.60	.53	.53	.35	.40	.41	.17
AR	.72	1.00	.71	.67	.63	.52	.53	.83	.68	.66	.34	.29	.27	.90	.56	.69	.64	.57	.57	.38	.39	.36	.16
WK	.80	.71	1.00	.80	.62	.55	.53	.67	.59	.68	.33	.25	.28	.91	.43	.59	.53	.46	.50	.35	.36	.37	.18
PC	.69	.67	.80	1.00	.61	.56	.42	.64	.52	.57	.34	.23	.31	.85	.46	.56	.47	.42	.42	.32	.31	.31	.16
NO	.52	.63	.62	.61	1.00	.70	.31	.62	.41	.42	.44	.29	.41	.69	.44	.47	.40	.36	.34	.36	.34	.34	.15
CS	.45	.52	.55	.56	.70	1.00	.23	.52	.34	.34	.48	.34	.41	.60	.41	.43	.39	.38	.32	.37	.36	.33	.14
AS	.64	.53	.53	.42	.31	.23	1.00	.42	.74	.75	.20	.24	.02	.54	.25	.44	.50	.44	.52	.20	.30	.31	.20
MK	.70	.83	.67	.64	.62	.52	.42	1.00	.60	.59	.35	.28	.30	.88	.54	.67	.62	.56	.54	.37	.37	.37	.16
MC	.70	.68	.59	.52	.41	.34	.74	.60	1.00	.74	.29	.29	.14	.68	.40	.59	.64	.57	.61	.36	.42	.38	.21
EI	.76	.66	.68	.57	.42	.34	.75	.59	.74	1.00	.27	.27	.13	.71	.34	.54	.56	.51	.53	.28	.35	.36	.17
DEX	.31	.34	.33	.34	.44	.48	.20	.35	.29	.27	1.00	.78	.68	.32	.31	.33	.29	.31	.28	.32	.32	.31	.20
FING	.27	.29	.25	.23	.29	.34	.24	.28	.28	.29	.27	1.00	.26	.29	.25	.31	.31	.30	.27	.25	.28	.27	.16
MOT	.21	.27	.28	.31	.41	.41	.02	.30	.14	.13	.68	.26	1.00	.32	.25	.23	.15	.16	.17	.24	.21	.19	.08
AFQT	.83	.90	.91	.85	.69	.60	.54	.88	.68	.71	.32	.29	.32	1.00	.55	.70	.64	.56	.58	.40	.41	.41	.19
SM	.41	.56	.43	.46	.44	.41	.25	.54	.40	.34	.31	.25	.25	.55	1.00	.59	.54	.53	.48	.39	.38	.29	.12
SR	.60	.69	.59	.56	.47	.43	.44	.67	.59	.54	.33	.31	.23	.70	.59	1.00	.64	.64	.62	.42	.43	.36	.14
ID	.60	.64	.53	.47	.40	.39	.50	.62	.64	.56	.29	.31	.15	.64	.54	.64	1.00	.66	.60	.39	.43	.36	.16
AO	.53	.57	.46	.42	.36	.38	.44	.56	.57	.51	.31	.30	.16	.56	.53	.64	.66	1.00	.62	.43	.47	.39	.17
SO	.53	.57	.50	.42	.34	.32	.52	.54	.61	.53	.28	.27	.17	.58	.48	.62	.60	.62	1.00	.39	.43	.31	.15
OT	.35	.38	.35	.32	.36	.37	.20	.37	.36	.28	.32	.25	.24	.40	.39	.42	.39	.43	.39	1.00	.76	.39	.16
TT	.40	.39	.36	.31	.34	.36	.30	.37	.42	.35	.32	.28	.21	.41	.38	.43	.43	.47	.43	.76	1.00	.40	.18
TIACC	.18	.21	.18	.22	.14	.14	.15	.17	.20	.17	.10	.10	.07	.21	.21	.25	.24	.26	.24	.11	.12	.22	-.05
GEOTOT	.44	.39	.40	.34	.36	.36	.35	.40	.42	.38	.34	.29	.20	.44	.32	.38	.39	.42	.33	.41	.42	.97	.32
GEODEC	.39	.34	.36	.30	.33	.32	.30	.36	.37	.34	.31	.26	.19	.39	.28	.34	.35	.38	.30	.38	.38	1.00	.09
CORR	.41	.36	.37	.31	.34	.33	.31	.37	.38	.36	.31	.27	.19	.41	.29	.36	.36	.39	.31	.39	.40	1.00	.10
TIMEAN	.40	.35	.36	.30	.33	.32	.30	.36	.37	.34	.30	.26	.18	.39	.28	.34	.36	.38	.30	.38	.39	1.00	.10
TIMOV	.17	.16	.18	.16	.15	.14	.20	.16	.21	.17	.20	.16	.08	.19	.12	.14	.16	.16	.15	.16	.18	.10	1.00

NOTE: CORR is the same variable as MDECCORR from earlier tables; TIMOV is the same as TIMOVMED from earlier tables. DEX, FING, and MOT are the manual dexterity, finger dexterity, and motor coordination composites of the GATB.

Table 7. Corrected correlation matrix of CAT-ASVAB, ECAT, and GATB subtests (n = 1,141)

	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	DEX	FING	MOT	AFQT	SM	SR	ID	AO	SO	OT	TT	TI	
																						CORR	TIMOV
GS	1.00	.73	.86	.78	.54	.49	.47	.67	.67	.72	.32	.26	.25	.86	.43	.62	.59	.53	.54	.35	.38	.39	.19
AR	.73	1.00	.68	.71	.62	.52	.40	.79	.68	.58	.33	.28	.24	.90	.57	.68	.64	.57	.57	.38	.42	.36	.16
WK	.86	.68	1.00	.79	.53	.49	.38	.63	.60	.70	.32	.25	.26	.88	.41	.59	.53	.47	.51	.34	.37	.37	.19
PC	.78	.71	.79	1.00	.53	.51	.35	.65	.62	.64	.31	.25	.25	.86	.46	.60	.55	.51	.51	.35	.35	.34	.16
NO	.54	.62	.53	.53	1.00	.70	.07	.69	.32	.39	.40	.22	.44	.68	.48	.49	.39	.34	.32	.35	.32	.29	.11
CS	.49	.52	.49	.51	.70	1.00	.05	.59	.31	.36	.46	.29	.45	.60	.50	.47	.40	.39	.34	.38	.35	.35	.13
AS	.47	.40	.38	.35	.07	.05	1.00	.20	.64	.57	.74	.17	.08	.37	.14	.29	.39	.34	.40	.12	.22	.22	.16
MK	.67	.79	.63	.65	.69	.59	.20	1.00	.55	.54	.36	.25	.32	.88	.55	.65	.59	.54	.49	.36	.35	.35	.14
MC	.67	.68	.60	.62	.32	.31	.64	.55	1.00	.67	.31	.33	.11	.69	.42	.61	.64	.63	.62	.39	.44	.41	.22
EI	.72	.58	.70	.64	.39	.36	.57	.54	.67	1.00	.25	.25	.15	.69	.32	.53	.52	.47	.47	.26	.30	.32	.20
DEX	.32	.33	.32	.31	.40	.46	.09	.36	.31	.25	1.00	.78	.68	.37	.31	.33	.29	.31	.28	.32	.32	.31	.20
FING	.26	.28	.25	.25	.22	.29	.17	.25	.33	.25	.78	1.00	.26	.29	.25	.31	.31	.30	.27	.25	.28	.27	.16
MOT	.25	.24	.26	.25	.44	.45	.08	.32	.11	.15	.68	.26	1.00	.31	.25	.23	.15	.16	.17	.24	.21	.19	.08
AFQT	.86	.90	.88	.86	.68	.60	.37	.88	.69	.69	.37	.29	.31	1.00	.57	.71	.66	.59	.59	.41	.42	.40	.19
SM	.43	.57	.41	.46	.48	.50	.14	.55	.42	.32	.31	.25	.25	.57	1.00	.59	.54	.53	.48	.39	.38	.29	.12
SR	.62	.68	.59	.60	.49	.47	.29	.65	.61	.53	.33	.31	.23	.71	.59	1.00	.64	.64	.62	.42	.43	.36	.14
ID	.59	.64	.53	.55	.39	.40	.39	.59	.64	.52	.29	.31	.15	.66	.54	.64	1.00	.66	.60	.39	.43	.36	.16
AO	.53	.57	.47	.51	.34	.39	.34	.54	.63	.47	.31	.30	.16	.59	.53	.64	.66	1.00	.62	.43	.47	.39	.17
SO	.54	.57	.51	.51	.32	.34	.40	.49	.62	.47	.28	.27	.17	.59	.48	.62	.60	.62	1.00	.39	.43	.31	.15
OT	.35	.38	.34	.35	.35	.38	.12	.36	.39	.26	.32	.25	.24	.41	.39	.42	.39	.43	.39	1.00	.76	.39	.16
TT	.38	.42	.37	.35	.32	.35	.22	.35	.44	.30	.32	.28	.21	.42	.38	.43	.43	.47	.43	.76	1.00	.40	.18
TIACC	.20	.18	.21	.22	.15	.15	.12	.16	.20	.16	.10	.10	.07	.21	.21	.25	.24	.26	.24	.11	.12	.22	.05
GEOTOT	.42	.38	.40	.37	.31	.36	.25	.37	.45	.35	.34	.29	.20	.43	.32	.38	.39	.42	.33	.41	.42	.97	.32
GEODEC	.38	.34	.36	.32	.28	.34	.21	.33	.40	.30	.31	.26	.19	.39	.28	.34	.35	.38	.30	.38	.38	1.00	.09
CORR	.39	.36	.37	.34	.29	.35	.22	.35	.41	.32	.31	.27	.19	.40	.29	.36	.36	.39	.31	.39	.40	1.00	.10
TIMEAN	.38	.34	.36	.33	.28	.34	.21	.34	.40	.30	.30	.26	.18	.39	.28	.34	.36	.38	.30	.38	.39	1.00	.10
TIMOV	.19	.16	.19	.16	.11	.13	.16	.14	.22	.20	.20	.16	-.08	-.19	.12	.14	.16	.17	.15	.16	.18	.10	1.00

NOTE: CORR is the same variable as MDECCORR from earlier tables; TIMOV is the same as TIMOVMED from earlier tables. DEX, FING, and MOT are the manual dexterity, finger dexterity, and motor coordination composites of the GATB.

different target identification latencies, GEOTOT, GEODEC, MDECCORR, and TIMEAN, had correlations that rounded to 1.00. This makes sense because total time (GEOTOT) was largely determined by decision time (GEODEC). MDECCORR and TIMEAN were just alternative measures of decision time. Therefore, the analyst chose MDECCORR as the response latency of choice to be congruent with the work of previous investigators [16]. MDECCORR balances the criteria of speed and accuracy because it is the mean decision time for correctly answered items. Appendix D reports the descriptive statistics for the entire sample of CAT-ASVAB and new predictor scores.

Factor Structure of the ECAT Subtests

The analyst used factor analysis to determine subsets of tests that seem to be measuring similar underlying abilities. In earlier factor analyses of the enlistment and CAT-ASVAB [9], investigators found that CAT-ASVAB factors were less correlated, and, therefore, may provide more differential validity than the paper-and-pencil version of ASVAB. They found the most striking difference in the factor correlations between the speed and technical factors; this difference was much smaller for CAT-ASVAB than for the paper-and-pencil version.

In [9], Mayberry and Divgi performed successive analyses with the number of factors ranging from 4 to 7. They found that the dexterity tests of the GATB and the coordination tests of the ECAT defined distinct and separate factors; these tests never loaded highly on the same dimension. Therefore, in the present study, the analyst entered the GATB and the coordination tests of the ECAT separately for incremental validity analyses, even though both tests measure psychomotor skills.

Factor analyses performed for this research included the ASVAB and ECAT subtests, without the GATB psychomotor measures (table 8 for enlistment ASVAB and table 9 for CAT-ASVAB, based on corrected correlations). The GATB psychomotor measures were left out because they are individually administered hands-on tests that are not likely to be added to the ASVAB. In contrast, the ECAT subtests are computer administered, and were designed specifically to add new dimensions to the CAT-ASVAB.

After examining each of the factor outcomes for interpretability and simple structure, the analyst determined the 6-factor solution to be the best. Tables 8 and 9 show this outcome for enlistment and CAT subtests, respectively, based on the corrected correlations.

Table 8 relates the ECAT to the enlistment ASVAB. It shows that all factors are correlated, which indicates a general factor underlying success in all the tests. AS, MC, and EI loaded on the first technical factor. The five power ECAT subtests (SM, SR, ID, AO, and SO) loaded most highly on the second spatial factor. Congruent with the findings of [9], OT and TT loaded on their own coordination factor. MDECCORR also had a small loading on the coordination factor. The communality for MDECCORR was low, reflecting its low correlation with other variables.

Table 8. Factor analysis of enlistment ASVAB subtests and ECAT subtests

	Factor pattern matrix						Communality estimate
	Tech	Spatial	Speed	Coord	Verbal	Math	
ENLGS	.40				.45		.80
ENLAR						.60	.87
ENLWK					.82		.91
ENLPC					.64		.73
ENLNO			.77				.76
ENLCS			.77				.68
ENLAS	.96						.78
ENLMK						.58	.80
ENLMC	.70						.78
ENLEI	.71						.78
SM		.66					.53
SR		.65					.69
ID		.62					.66
AO		.76					.67
SO		.60					.58
OT				.87			.77
TT				.86			.77
MDECCORR[a]				.27[b]			.28

NOTE: Only loadings .35 or greater are shown.

a. MDECCORR is mean decision time for correctly answered items in target identification.

b. The .27 loading for MDECCORR is shown because that is the highest loading for that test.

Interfactor correlations

	Tech	Spatial	Speed	Coord	Verbal	Math
Tech	1.00	.59	.33	.38	.59	.55
Spatial	.59	1.00	.49	.57	.55	.63
Speed	.33	.49	1.00	.43	.62	.58
Coord[a]	.38	.57	.43	1.00	.37	.34
Verbal	.59	.55	.62	.37	1.00	.69
Math	.55	.63	.58	.34	.69	1.00

NOTE: The correlation matrix was corrected for range restriction.

a. Coord represents a coordination factor.

Table 9. Factor analysis of CAT-ASVAB subtests and ECAT subtests

	Factor pattern matrix						
	Spatial	Verbal	Speed	Coord	Tech	Math	Communality estimate
CATGS		.76					.85
CATAR						.53	.86
CATWK		1.02					.90
CATPC		.68					.73
CATNO			.74				.75
CATCS			.84				.74
CATAS					.89		.75
CATMK						.39	.78
CATMC	.35				.45		.79
CATEI		.51			.38		.67
SM	.62						.53
SR	.65						.68
ID	.68						.65
AO	.84						.69
SO	.64						.57
OT				.86			.76
TT				.86			.77
MDECCORR				.26[a]			.28

NOTE: Only loadings .35 or greater are shown.

a. The .26 for MDECCORR is its highest loading.

Interfactor correlations

	Spatial	Verbal	Speed	Coord	Tech	Math
Spatial	1.00	.64	.51	.57	.50	.58
Verbal	.64	1.00	.56	.40	.55	.60
Speed	.51	.56	1.00	.40	.04	.49
Coord[a]	.57	.40	.40	1.00	.27	.28
Tech	.50	.55	.04	.27	1.00	.33
Math	.58	.60	.49	.28	.33	1.00

NOTE: The correlation matrix was corrected for range restriction.

a. Coord represents a coordination factor.

Table 9 relates the ECAT to the CAT-ASVAB. In this analysis, the ECAT power subtests, especially AO, load most highly on the first spatial factor. OT and TT defined a coordination factor that was unique from the ASVAB; MDECCORR had a moderate loading on coordination.

These analyses indicate that ECAT subtests measure a general intelligence dimension and a coordination dimension. SM is theoretically, but not empirically, different from the spatial tests. These analyses indicate that the two coordination tests should be entered together for incremental validity analyses. Similarly, the four ECAT spatial subtests should be entered together; AO had the highest loading on this spatial factor.

Descriptive Statistics for Examinees

Tables 10 and 11 show descriptive statistics for the automotive mechanics and helicopter mechanics, respectively. These tables show that there was a wide variation of time in service among examinees. Time in service, measured in months, ranged from a low of 8 months to a high of 160 months for the automotive mechanics, and from 9 to 128 months for the helicopter mechanics. Such differences in amount of experience may affect performance on the predictor tests and/or the performance tests simply because of on-the-job experience, training, or maturity. TIS correlated 0.23 with hands-on performance for automotive mechanics, and 0.40 for helicopter mechanics. To control for these potential developmental effects, the analyst used TIS and its square as covariates to correct the correlation matrices. In this manner, the analyst statistically adjusted performance scores as if all examinees had the same number of months in service.

Tables 10 and 11 also show that mechanical aptitude was fairly high among examinees, with very few scores of less than 100 on the MM composite. This is not surprising, given the required MM cutoff for admission to these MOSs.

Tables 12 and 13 show sample and range-corrected correlations of ECAT composites and subtests with hands-on performance scores. All composites were based on scores standardized to mean 50 and standard deviation 10 in the sample. Spatial was defined as SR + ID + AO + SO standard scores. Coordination was defined as OT + TT. Memory was actually the score for one test--it was simply the standardized score for SM. SM loaded on the same factors as the other power subtests, but it measured a theoretically distinct ability. Perceptual speed was defined as the mean decision time for target identification items that were answered correctly (MDECCORR), which follows the practice of some earlier investigators [16]. The tables also show alternative measures of target identification speed and accuracy.

Table 10. Descriptive statistics for automotive mechanics (n = 698)

Construct	Measure	Mean	Std. dev.
Job performance	Hands-on total score	77.87	7.62
Experience	Time in service (in months)	39.07	22.74
ECAT subtests			
Memory	SM[a]	42.25	18.22
Spatial ability	SR[b]	61.78	21.50
	ID[c]	72.80	12.48
	AO[d]	59.88	19.68
	SO[e]	46.43	25.12
Coordination	OT[f]	7.94	.12
	TT[g]	8.21	.12
Perceptual accuracy	TIACC[h]	92.93	6.86
Perceptual speed	GEOTOT[i]	3.21	.56
	GEODEC[j]	2.81	.54
Enlistment ASVAB			
Composites	ENLMM	111.66	10.52
	ENLGT	108.22	9.52
	ENLEL	105.73	10.28
	ENLCL	103.76	9.29
CAT-ASVAB			
Composites	CATMM	116.94	9.98
	CATGT	110.67	10.65
	CATEL	106.82	9.89
	CATCL	103.16	9.48

NOTE: Marine Corps mechanical maintenance composite, MM = AS + MC + AR + EI; Marine Corps general technical composite, GT = VE + AR + MC; Marine Corps electronics composite, EL = GS + AR + MK + EI; Marine Corps clerical/administrative composite, CL = VE + MK + CS. The composites are standardized to mean 10, standard deviation 10.

a. Sequential memory.
b. Spatial reasoning.
c. Integrating detail.
d. Assembling objects.
e. Spatial orientation.
f. One-hand tracking mean log of root mean square error distance over all nonpractice trials.
g. Two-hand tracking mean log of root mean square error distance over all nonpractice trials.
h. Accuracy on target identification. SM, SR, ID, AO, SO, and TIACC are percent correct scores.
i. Geometric mean total response time for target identification.
j. Geometric mean total decision time for target identification.

Table 11. Descriptive statistics for helicopter mechanics (n = 443)

Construct	Measure	Mean	Std. dev.
Job performance	Hands-on total score	79.74	7.76
Experience	Time in service (in months)	40.67	22.78
ECAT subtests			
Memory	SM[a]	46.16	17.16
Spatial ability	SR[b]	67.43	19.00
	ID[c]	76.71	12.12
	AO[d]	64.15	19.84
	SO[e]	53.96	24.92
Coordination	OT[f]	7.90	.11
	TT[g]	8.16	.12
Perceptual accuracy	TIACC[h]	92.93	6.58
Perceptual speed	GEOTOT[i]	2.98	.48
	GEODEC[j]	2.62	.47
Enlistment ASVAB			
Composites	ENLMM	116.75	8.11
	ENLGT	113.54	8.84
	ENLEL	112.21	9.41
	ENLCL	109.05	10.14
CAT-ASVAB			
Composites	CATMM	119.13	9.29
	CATGT	115.72	9.84
	CATEL	112.09	9.91
	CATCL	109.04	10.18

NOTE: Marine Corps mechanical maintenance composite, MM = AS + MC + AR + EI; Marine Corps general technical composite, GT = VE + AR + MC; Marine Corps electronics composite, EL = GS + AR + MK + EI; Marine Corps clerical/administrative composite, CL = VE + MK + CS. The composites are standardized to mean 10, standard deviation 10.

a. Sequential memory.
b. Spatial reasoning.
c. Integrating detail.
d. Assembling objects.
e. Spatial orientation.
f. One-hand tracking mean log of root mean square error distance over all nonpractice trials.
g. Two-hand tracking mean log of root mean square error distance over all nonpractice trials.
h. Accuracy on target identification. SM, SR, ID, AO, SO, and TIACC are percent correct scores.
i. Geometric mean total response time for target identification.
j. Geometric mean total decision time for target identification.

Table 12. Correlation of ASVAB, ECAT, and GATB with mechanical hands-on performance test (automotive mechanics, n = 698)

	Sample values		Corrected values	
	Enlistment	CAT	Enlistment	CAT
ASVAB subtest				
GS	.20	.25	.55	.56
AR	.18	.22	.54	.54
WK	.21	.19	.50	.48
PC	.16	.24	.45	.50
NO	-.04	-.01	.36	.31
CS	.04	.09	.31	.32
AS	.38	.51	.63	.66
MK	.18	.09	.48	.42
MC	.31	.43	.62	.66
EI	.29	.34	.60	.58
GATB subtest				
Motor	-.01		.05	
Manual	.06		.13	
Finger	.18		.36	
Dexterity	.11		.25	
ECAT subtest composites				
Spatial	.35		.60	
Coordination	.09		.30	
Memory	.13		.35	
Perceptual speed	.10		.29	
ECAT subtest measures				
Spatial				
SR	.22		.49	
ID	.27		.53	
AO	.32		.51	
SO	.25		.49	
Memory				
SM	.13		.35	
Coordination				
OT	.04		.23	
TT	.13		.33	
Target identification				
TIACC	.12		.20	
GEOTOT	.13		.32	
GEODEC	.09		.27	
MDECORR[a]	.10		.29	
TIMOVMED[b]	.15		.20	
TIMEAN[c]	.08		.27	

Table 12. (Continued)

NOTE: The correlations for GEOTOT, GEODEC MDECORR, TIMOVMED, and TIMEAN indicate that those who answer faster performed better on the hands-on test. They are measures of perceptual speed based on target identification.

a. MDECCORR = mean decision time for items answered correctly.
b. TIMOVMED = median of movement time for target identification items.
c. TIMEAN = clipped mean of decision time for target identification items. To clip the mean, items were divided into "hard" and "easy" items. The highest and lowest decision times for each set of items were discarded. The mean of the two sets of remaining decision times was the clipped mean decision time (TIMEAN).

Table 13. Correlation of ASVAB, ECAT, and GATB with mechanical hands-on performance test (helicopter mechanics, n = 443)

	Sample values		Corrected values	
	Enlistment	CAT	Enlistment	CAT
ASVAB subtest				
GS	.05	.11	.50	.46
AR	.18	.28	.52	.55
WK	-.01	.13	.39	.41
PC	.03	.15	.35	.42
NO	.04	.07	.35	.31
CS	.13	.16	.35	.28
AS	.25	.36	.58	.57
MK	.19	.15	.49	.45
MC	.23	.40	.59	.61
EI	.17	.23	.53	.48
GATB subtest				
Motor	.02		.07	
Manual	.15		.15	
Finger	.18		.18	
Dexterity	.16		.18	
ECAT subtest composites				
Spatial	.34		.58	
Coordination	.12		.31	
Memory	.13		.35	
Perceptual speed	.11		.26	
ECAT subtest measures				
Spatial				
SR	.27		.49	
ID	.23		.50	
AO	.34		.57	
SO	.21		.42	
Memory				
SM	.13		.35	
Coordination				
OT	.13		.32	
TT	.10		.27	
Target identification				
TIACC	.06		.19	
GEOTOT	.14		.28	
GEODEC	.10		.24	
MDECORR	.11		.26	
TIMOVMED	.11		.12	
TIMEAN	.10		.25	

NOTE: The correlations for GEOTOT, GEODEC, MDECORR, TIMOVMED, and TIMEAN indicate that those who answer faster have higher hands-on total scores. Correlations for OT and TT indicate that examinees with less tracking error have higher hands-on total scores.

Tables 12 and 13 show that, among the ECAT composites, the spatial composite had the highest correlation with hands-on performance. Among spatial subtests, ID and AO had the highest correlations with mechanical hands-on performance. Note that the correlations for the GATB subtests were fairly low, except for finger dexterity, which had a moderate correlation for the automotive mechanics. Coordination (OT and TT) and some alternative measures of target identification (GEOTOT, GEODEC, MDECCORR, and TIMEAN) had moderate correlations with hands-on performance.

Incremental Validity of ECAT and GATB

The analyst conducted incremental validity analyses using the procedures followed by Mayberry and Divgi [9]. These results (tables 14 and 15) show incremental validity above the ten ASVAB subtests using four different bases:

- Sample correlations and enlistment ASVAB
- Range-corrected correlations and enlistment ASVAB
- Sample correlations and CAT-ASVAB
- Range-corrected correlations and CAT-ASVAB.

Table 14, for automotive mechanics, shows the spatial composite to increment validity 0.032 (6.5 percent) against enlistment ASVAB using sample correlations. Assembling objects (AO), at .036, was by far the most important contributor to this increment. The addition of the other three spatial tests actually detract from the incremental validity of AO alone. The incremental validity of AO was also greater than for the combined validity of the four ECAT composites and the GATB. For the automotive mechanics, increments for AO ranged from a high of 0.036 (7.3 percent) using sample values and comparing against the enlistment ASVAB to a low of 0.012 (1.6 percent) using range-corrected correlations and comparing against the CAT-ASVAB.

Table 15 shows a similar pattern for the helicopter mechanics. The incremental validity of AO was greater than that for the entire spatial composite. The incremental validity of AO ranged from a high of 0.024 (4.2 percent) using sample correlations versus the enlistment ASVAB to a low of 0.015 (2.2 percent) using range-corrected correlations versus the CAT-ASVAB.

Stepwise Regressions for ECAT versus MM Composite

The CNA analyst conducted analyses to determine whether ECAT subtests would be candidates to replace some elements of the Marine Corps' current ASVAB mechanical maintenance (MM) composite. The current MM composite is AS + MC + AR + EI. To do these analyses, he performed stepwise regression using all ten subtests of the ASVAB and the eight ECAT subtests. For target identification, he used the mean decision time for items answered correctly (MDECCORR).

Table 14. Incremental validity for psychomotor measures and ECAT above ASVAB and time-in-service base (automotive mechanics)

	Sample values				Corrected values		
	df	SS	MSE	adj R	df	SS	adj R
Total	697	40,502.65			697	57,759.80	
Enlistment ASVAB							
+ TIS, TIS squared	12	10,455.88	43.86	.495	10	27,713.12	.687
+ spatial (SR+AO+SO+ID)	1	1,348.59	41.96	.032***	1	1,348.52	.017
+ coordination (OT+TT)	1	190.04	43.65	.004*	1	190.07	.002
+ memory (SM)	1	367.52	43.39	.008**	1	367.55	.004
+ perceptual speed (MDECCORR)	1	105.42	43.77	.002	1	105.43	.001
+ dexterity (GATB)	1	267.47	43.54	.006*	1	267.42	.003
+ all 5 composites (above)	5	1,543.71	41.92	.033***	5	1,543.64	.017
Spatial							
+ SR	1	331.94	43.44	.007**	1	331.97	.004
+ SO	1	278.36	43.52	.006*	1	278.33	.003
+ AO	1	1,521.74	41.70	.036***	1	1,521.73	.019
+ ID	1	733.73	42.86	.017***	1	733.74	.009
Coordination							
+ OT	1	112.57	43.77	.002	1	112.54	.001
+ TT	1	214.91	43.61	.005*	1	214.90	.002
Target identification							
TIACC	1	237.92	43.58	.005*	1	237.97	.003
GEOTOT	1	217.08	43.61	.005*	1	217.11	.002
GEODEC	1	87.38	43.80	.001	1	87.37	.001
TIMOVMED	1	428.34	43.30	.010**	1	428.36	.005
GATB finger dexterity	1	329.75	43.45	.007**	1	329.73	.004
CAT ASVAB							
+ TIS, TIS squared	12	14,039.73	38.63	.579	10	31,165.68	.730
+ spatial (SR+AO+SO+ID)	1	878.88	37.40	.018***	1	915.23	.011
+ coordination (OT+TT)	1	30.61	38.64	.000	1	34.30	.000
+ memory (SM)	1	239.94	38.34	.004*	1	244.32	.003
+ perceptual speed (MDECCORR)	1	35.45	38.64	.000	1	38.78	.000
+ dexterity (GATB)	1	68.10	38.59	.001	1	71.02	.001
+ all 5 composites (above)	5	953.75	37.51	.016***	5	988.49	.010
Spatial							
+ SR	1	249.95	38.32	.005**	1	270.20	.003
+ SO	1	158.83	38.46	.003*	1	175.17	.002
+ AO	1	1,007.63	37.22	.021***	1	1,005.64	.012
+ ID	1	469.32	38.00	.009***	1	498.77	.006
Target identification							
TIACC	1	138.36	38.49	.002	1	140.39	.002
GEOTOT	1	65.41	38.60	.001	1	70.90	.001
GEODEC	1	25.01	38.65	.000	1	27.71	.000
TIMOVMED	1	169.08	38.44	.003*	1	171.37	.002
GATB finger dexterity	1	171.22	38.44	.003*	1	182.49	.002

NOTES: Significance tests were performed only for sample values. * = p < .05, ** = p< .01, *** = p < .001. TIS was a constant for range correction, so base model df is 10 for corrected values. Negative adjusted r's were set to zero. Dexterity is a GATB composite; Finger Dexterity is one component of the Dexterity composite.

Table 15. Incremental validity for psychomotor measures and ECAT above ASVAB and time-in-service base (helicopter mechanics)

	Sample values				Corrected values		
	df	SS	MSE	adj R	df	SS	adj R
Total	442	26,592.72	—	—	442	32,017.36	—
Enlistment ASVAB							
TIS + TIS squared	12	9,098.46	40.68	.569	10	14,523.18	.664
+ spatial	1	514.26	39.58	.016***	1	514.25	.011
+ coordination	1	297.04	40.09	.009**	1	297.02	.006
+ memory	1	90.77	40.57	.002	1	90.77	.001
+ perceptual speed	1	98.99	40.55	.002	1	98.99	.001
+ dexterity	1	51.39	40.66	.000	1	51.38	.000
+ all 5 composites (above)	5	674.46	39.58	.016**	5	674.44	.011
Spatial							
+ SR	1	340.77	39.98	.010**	1	340.80	.007
+ SO	1	51.69	40.66	.000	1	51.68	.000
+ AO	1	749.89	39.03	.024***	1	749.92	.017
+ ID	1	171.95	40.38	.004*	1	171.89	.003
Coordination							
+ OT	1	397.25	39.85	.012**	1	397.18	.009
+ TT	1	158.97	40.41	.004*	1	158.94	.003
Target identification							
TIACC	1	22.39	40.73	.000	1	22.87	.000
GEOTOT	1	128.21	40.48	.003	1	128.20	.002
GEODEC	1	91.28	40.56	.002	1	91.30	.001
TIMOVMED	1	33.73	40.70	.000	1	33.73	.000
GATB finger dexterity	1	75.14	40.60	.001	1	75.14	.001
CAT ASVAB							
+ TIS +TIS squared	12	9,755.90	39.16	.591	10	14,992.67	.675
+ spatial	1	389.25	38.33	.011*	1	444.66	.010
+ coordination	1	114.79	39.00	002	1	138.90	.003
+ memory	1	17.87	39.21	.000	1	20.24	.000
+ perceptual speed	1	25.39	39.19	.000	1	32.53	.000
+ dexterity	1	55.04	39.12	.000	1	65.77	.001
+ all 5 composites (above)	5	487.27	38.47	.010**	5	563.57	.009
Spatial							
+ SR	1	211.74	38.75	.006*	1	216.99	.004
+ SO	1	43.23	39.15	.000	1	59.53	.001
+ AO	1	614.04	37.82	.019***	1	678.49	.015
+ ID	1	96.87	39.02	.002	1	124.69	.002
Target identification							
TIACC	1	53.83	39.12	.000	1	56.70	.001
GEOTOT	1	47.53	39.14	.000	1	57.70	.001
GEODEC	1	21.12	39.20	.000	1	27.44	.000
TIMOVMED	1	54.93	39.12	.000	1	57.70	.001
GATB finger dexterity	1	55.06	39.12	.000	1	70.07	.001

NOTES: Significance tests were performed for sample values only. *** = p < .001, ** = p < .01, * = p < .05. TIS was a constant for range restriction, so base model df is 10 for corrected models. Negative adjusted r's were set to zero.

Table 16 shows the results of the stepwise regression. For these analyses, the base was TIS and TIS squared, and only sample values were used. Table 16 shows that AS (from the ASVAB) and AO (from the ECAT) were usually among the earliest components of a predictor formed using stepwise regression. MC (from ASVAB) was the strongest CAT-ASVAB predictor of helicopter performance. These findings indicate that two present components of the MM composite, AS and MC, were highly effective at predicting mechanical hands-on performance. These results also show that AO would probably enhance the predictive validity of the current MM composite more than do the current ASVAB components AR and EI.

In table 16, the predictor that enters the equation first will always have the highest incremental validity, and predictors that come later will have less incremental validity. If AO is forced into the equation first, the incremental validity of AO is greater. If AO is forced into the equation first for automotive mechanics, AO's incremental validity over TIS and TIS squared is 0.153 for automotive mechanics. This shows that AO was a strong predictor of automotive mechanical performance although it was not as strong a predictor as AS, which had incremental validity of 0.215 when added first. For helicopter mechanics, AO was the first entered subtest when compared with the enlistment ASVAB. When compared against the CAT-ASVAB for helicopter mechanics, AO's incremental validity of 0.081 when added first was nearly the same as the value for CAT-MC, which was 0.090.

Factor Structure of ASVAB with AO

The analyst conducted factor analyses to determine whether AO would change the factor structure of ASVAB. The current dimensions of the ASVAB are technical, verbal, mathematical, and speed. When AO was added to the CAT-ASVAB (appendix E), AO and MC defined a fourth, spatial factor that was related to ability to visualize relationships among objects. When AO was added to the enlistment ASVAB (appendix F), AO defined its own factor. The CAT- and enlistment-ASVAB findings indicate that the addition of AO might add a new dimension to the structure of the ASVAB.

CONCLUSIONS

Reliability and uniqueness are necessary, but not sufficient, conditions to improve the predictive validity of the ASVAB. The ECAT subtests demonstrate reliabilities sufficiently high to have potential to improve the predictive validity of the ASVAB. The tests with the most uniqueness relative to ASVAB were the coordination tasks (OT and TT), and the response latencies for target identification. Uniquenesses for the spatial ECAT subtests (SR, SO, ID, AO) were moderate.

In general, the incremental validities for most ECAT subtests were low, but AO was an exception. Of the ECAT subtests, AO added the most to the validity of the ASVAB. AO had more incremental validity than did the entire battery of ECAT subtests and GATB. These improvements ranged

Table 16. Stepwise regressions for ECAT and ASVAB subtests (automotive and helicopter mechanics)

	Sample values			
	df	SS	MSE	r
Automotive mechanics (n = 698)				
Total	697	40,502.65		
TIS + TIS squared	2	2,220.26	55.08	.234***
Versus enlistment ASVAB				
+ AS (ENL)	1	5,942.41	46.60	.215***
+ AO (ECAT)	1	2,873.66	42.52	.073***
+ MC (ENL)	1	504.82	41.85	.012***
+ ID (ECAT)	1	276.04	41.51	.006*
+ EI (ENL)	1	218.91	41.26	.005*
Versus CAT ASVAB				
+ AS (CAT)	1	8,876.23	42.37	.289***
+ AO (ECAT)	1	2,572.98	38.72	.058***
+ MC (CAT)	1	672.93	37.80	.014***
+ CS (CAT)	1	312.84	37.41	.007**
+ ID (ECAT)	1	121.46	37.28	.002 N.S.
+ EI (CAT)	1	97.13	37.20	.002 N.S.
+ WK (CAT) (neg)	1	191.18	36.97	.004*
Helicopter mechanics (n = 443)				
Total	442	26,592.72		
TIS + TIS squared	2	5,440.03	48.07	.452
Versus enlistment ASVAB				
+ AO (ECAT)	1	2,112.82	43.37	.081***
+ AS (ENL)	1	1,007.10	41.17	.034***
+ MK (ENL)	1	563.70	39.97	.019***
+ OT (ECAT)	1	339.93	39.29	.011**
+ WK (ENL) (neg)	1	236.77	38.83	.007*
+ CS (ENL)	1	212.20	38.43	.007*
+ MC (ENL)	1	101.99	38.29	.003 N.S.
Versus CAT ASVAB				
+ MC (CAT)	1	2,361.79	42.80	.090***
+ AO (ECAT)	1	696.00	41.31	.023***
+ AS (CAT)	1	813.58	39.55	.027***
+ AR (CAT)	1	355.58	38.82	.011**
+ GS (CAT) (neg)	1	376.28	38.04	.012**
+ CS (CAT)	1	212.41	37.64	.006*
+ OT (ECAT)	1	113.40	37.47	.003 N.S.

NOTES: The current Marine Corps' current mechanical maintenance (MM) composite = AS + MC + AR + EI. Significance tests were performed only for sample values. * = $p < .05$, ** = $p < .01$, *** = $p < .001$, N. S. = not significant. WK and GS had negative regression coefficients for both automotive and helicopter mechanics.

from a high of 0.036 (7.3 percent) for automotive mechanics against the enlistment ASVAB using sample correlations, to a low of 0.012 (1.6 percent) against the CAT-ASVAB using corrected correlations. For helicopter mechanics, improvements ranged from a high of 0.024 (4.2 percent) using sample correlations and the enlistment ASVAB to a low of 0.015 (2.2 percent) using range-corrected correlations and the CAT-ASVAB. Stepwise regressions indicated that AO is likely to increase the predictive validity of the Marine Corps' current MM composite.

AO had the highest loading of all ECAT subtests on the spatial factor. Further analyses indicated that the addition of AO might change the factor structure of the ASVAB by adding a spatial component that measures the ability to visualize relationships among objects. However, these results should not be overinterpreted. Previous investigations have pointed out that factor analyses of selected samples, even with corrections for range restriction, pose difficult issues for interpreting factor outcomes [9]. Unknown effects are associated with comparing the same samples on aptitude measures administered at different times with substantial intervening events (e.g., training). This study was not designed to fully address these issues.

The results from the retest gains are also important. Those ECAT subtests that were found to be most unique compared to the ASVAB (OT, TT, and response latencies for target identification) had significant test-retest gains. SM and SO had significant test-retest gains as well. A significant retest gain indicates that the test could provide an advantage to applicants who had had practice on similar tasks. Retest gains also indicate that these tests might be coachable, and if that were the case, the tests would have limited use for military selection.

REFERENCES

[1] CNA, MCOAG Memorandum 83-3135, *A Factor Analysis of ASVAB Form 8a in the 1980 DOD Reference Population*, by Peter H. Stoloff, 25 Aug 1983

[2] CNA Research Memorandum 90-110, *Incremental Validity of New Tests in the Prediction of Infantry Performance*, by Paul W. Mayberry and Catherine M. Hiatt, Jul 1990

[3] Office of the Assistant Secretary of Defense (Force Management and Personnel), *Biennial Report of the Defense Advisory Committee on Military Personnel Testing*, Nov 1988

[4] H. Gulliksen. *Theory of Mental Tests*. New York: John Wiley & Sons, 1950

[5] CNA Report 116, *The ASVAB Score Scales: 1980 and World War II*, by Milton H. Maier and William H. Sims, Jul 1986

[6] Carey, Neil B., "Does Choice of a Criterion Matter?" *Military Psychology* 4 (1992): 103-117

[7] P. Cattin. "Estimation of the Predictive Power of a Regression Model." *Journal of Applied Psychology* 65 (1980): 407-414

[8] F. M. Lord and M. R. Novick. *Statistical Models of Mental Test Scores*. Reading, MA: Addison-Wesley Publishing Co., 1968

[9] CNA Research Memorandum 91-224, *Uniqueness of a Psychomotor Construct to ASVAB*, by Paul W. Mayberry and D.R. Divgi, Jan 1992

[10] CNA Research Memorandum 91-242, *Development and Scoring of Hands-on Performance Tests for Mechanical Maintenance Specialties*, by Neil B. Carey and Paul W. Mayberry, Mar 1992

[11] CNA Research Memorandum 91-246, Reliability of Mechanical Maintenance Performance Measures, by Paul W. Mayberry and William H. Wright, Feb 1992

[12] G. E. Larson. *Mental Counters and Sequential Memory: Proposals for Joint-Service Validation*. (Manuscript submitted to the Joint-Service Technical Advisory Selection Panel), 1989

[13] G. E. Larson. *Scoring of ECAT Working Memory Tests*, 20 Mar 1991

[14] D. L. Alderton. *Integrating Details: Proposal of Joint Services Validation*. (Draft manuscript submitted to the Technical Advisory Selection Panel Meeting, 4 Apr 1989), 1989

[15] Army Research Institute, Descriptive Information About Six Predictor Measures Developed for the U.S. Army Research Institute, Mar 1991

[16] N. G. Peterson et al. "Project A: Specification of the Predictor Domain and Development of New Selection/Classification Tests." *Personnel Psychology* 43 (1990): 247-275

APPENDIX A

RELIABILITY ESTIMATES FOR HANDS-ON PERFORMANCE TESTS

Table A-1. Reliability estimates[a] for hands-on performance tests

	MOS				
Reliability measure	3521	6112	6113	6114	6115
Hands-on performance test					
Test-retest	.79	.88	--	--	--
Split-halves	.80	.87	.91	.82	.87
Alpha coefficient	.81	.88	.88	.78	.81
Scorer agreement	.96	.98	.96	.96	.94

SOURCE: [11].

a. Reliability estimates corrected for restrictions in range, using the formula shown in Mayberry and Wright [11]. There were no retest examinees for MOSs 6113, 6114, or 6115 because of the small number in those MOSs.

APPENDIX B

UNCORRECTED RELIABILITIES OF ECAT SUBTESTS

Table B-1. Uncorrected reliability estimates for ECAT subtests (n = 1,141)

Reliability measure	ECAT subtest measure												
	SM	SR	ID	AO	SO	OT	TT	TIACC[a]	GEOTOT[b]	GEODEC[c]	MDECCORR[d]	TIMEAN[e]	TIMOVMED[f]
Test-retest	.67	.68	.68	.66	.62	.72	.68	.50	.75	.75	.76	.74	.67
Split-halves	.79	.81	.61	.78	.80	.94	.95	.63	.95	.96	.81	.92	.85
Alpha coefficient	.87	.89	.76	.87	.88	.96	.97	.72	.96	.98	.96	.98	.94

NOTES: Split-half reliability for TIMEAN is the correlation of easy and difficulty item sets. TIACC, GEOTOT, GEODEC, MDECORR, TIMEAN, and TIMOVMED all refer to the target identification task.

a. Average accuracy score.
b. Geometric mean of total time.
c. Geometric mean of decision time.
d. Mean decision time for items answered correctly.
e. Clipped mean decision time. Split-halves reliability for TIMEAN is the correlation of the easy items with the hard items.
f. Median of the movement time.

APPENDIX C

TEST-RETEST CHANGES IN RAW ECAT SUBTEST SCORES

Table C-1. Test-Retest changes in raw ECAT subtest scores (automotive and helicopter mechanics)

Subtest measure	Initial test n	Initial test Mean	Initial test Std. dev.	Retest n	Retest Mean	Retest Std. dev.	Performance gain (decrement) vs. initial std. dev. (percent)
SM	132	42.86	17.30	133	45.32	18.18	14.22%
SR	134	63.46	20.89	135	66.11	19.39	12.69
ID	131	72.89	13.37	127	75.35	12.87	3.44
AO	133	61.47	20.43	134	60.09	22.85	(6.75)
SO	135	48.53	26.38	135	54.33	27.73	21.99
TIACC	135	92.80	6.26	134	93.68	7.08	14.06
GEOTOT	135	3.04	0.53	135	2.93	0.55	20.75
GEODEC	135	2.67	0.51	135	2.56	0.55	21.57
OT	134	7.94	0.13	134	7.90	0.13	30.77
TT	134	8.18	0.14	134	8.15	0.15	21.43

NOTES: Scores for GEOTOT and GEODEC show that examinees answer more quickly at retest, so these retest scores are smaller than at initial test. Similiarly, OT and TT error decreases on retest. Two decimal places are used because of the metric used for OT and TT scores.

APPENDIX D

DESCRIPTIVE STATISTICS FOR ENTIRE SAMPLE

Table D-1. Descriptive statistics for entire sample (n = 1,141)

Measure	Mean	Standard deviation
ENLGS	53.04	6.41
ENLAR	53.67	6.12
ENLWK	52.47	5.60
ENLPC	53.44	5.56
ENLNO	54.41	6.20
ENLCS	52.03	6.34
ENLAS	57.93	6.84
ENLMK	52.56	6.97
ENLMC	57.10	6.36
ENLEI	55.21	6.57
TIS	39.70	22.82
CATGS	53.92	6.15
CATAR	54.01	6.24
CATWK	54.31	4.78
CATPC	53.47	6.10
CATNO	52.73	6.86
CATCS	52.46	6.54
CATAS	61.31	4.98
CATMK	50.31	7.02
CATMC	58.50	7.14
CATEI	57.35	5.85
Dexterity	300.18	42.54
Finger	97.10	19.52
SM	50.80	9.82
SR	50.88	9.71
ID	50.69	9.68
AO	50.60	9.82
SO	50.70	10.15
OT	-48.59	9.48
TT	-48.88	9.89
TIACC	50.14	9.68
GEOTOT	-48.94	9.24
GEODEC	-48.99	9.30
MDECCORR	-2.81	0.54
TIMEAN	-2.77	0.54
TIMOVMED	-0.34	0.12

NOTES: The signs for OT and TT were reversed because smaller distances to the target correspond with higher accuracy. GEOTOT, GEODEC, MDECCORR, TIMEAN, and TIMOVMED were reversed because these are time scores--smaller absolute values indicate higher speed of response.

APPENDIX E

FACTOR ANALYSIS OF CAT-ASVAB SUBTESTS AND AO

Table E-1. Factor analysis of CAT-ASVAB subtests and AO (constrained to five factors)

	Factor pattern matrix					Communality estimate
	Verbal	Speed	Tech	Spatial	Math	
CATGS	.76					.85
CATAR					.68	.87
CATWK	1.02					.90
CATPC	.68					.73
CATNO		.86				.83
CATCS		.75				.65
CATAS			.94			.84
CATMK					.48	.78
CATMC				.46		.80
CATEI	.52					.66
AO				.75		.62

NOTE: Only loadings of .35 or greater are shown.

Interfactor correlations

	Verbal	Speed	Tech	Spatial	Math
Verbal	1.00	.57	.51	.64	.68
Speed	.57	1.00	.03	.41	.59
Tech	.51	.03	1.00	.51	.41
Spatial	.64	.41	.51	1.00	.66
Math	.68	.59	.41	.66	1.00

NOTE: The correlation matrix was corrected for range restriction.

APPENDIX F

FACTOR ANALYSIS OF ENLISTMENT ASVAB SUBTESTS AND AO

Table F-1. Factor analysis of enlistment ASVAB subtests and AO (constrained to five factors)

	Factor pattern matrix					
	Tech	Speed	Verbal	Math	Assembling objects	Communality estimate
ENLGS	.55		.50			.79
ENLAR	.42	.38	.31	.69		.84
ENLWK			.87			.92
ENLPC			.63			.71
ENLNO		.74				.77
ENLCS		.80				.72
ENLAS	.99					.82
ENLMK				.81		.84
ENLMC	.69					.77
ENLEI	.65					.77
AO					.42	.54

NOTE: Only loadings of .35 or greater are shown.

Interfactor correlations

	Tech	Speed	Verbal	Math	AO
Tech	1.00	0.32	0.63	0.61	0.41
Speed	0.32	1.00	0.60	0.60	0.27
Verbal	0.63	0.60	1.00	0.70	0.34
Math	0.61	0.60	0.70	1.00	0.45
AO	0.41	0.27	0.34	0.45	1.00

NOTE: The correlation matrix was corrected for range restriction.