Navy Personnel Research and Development Center
San Diego, CA 92152-6800 TN-90-32 August 1990

(2)

Validity of the GRIP and ASVAB Test Batteries for Job Performance of Sonar Technician (Surface) Personnel

Charles H. Cory

NPRDC-TN-90-32 August 1990

Validity of the GRIP and ASVAB Test Batteries for Job Performance of Sonar Technician (Surface) Personnel

Charles H. Cory

Reviewed by
John H. Wolfe

Approved and released by
Jules I. Borack
Director, Personnel Systems Department

Approved for public release;
distribution is unlimited.

Navy Personnel Research and Development Center
San Diego, California 92152-6800

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE August 1990	3. REPORT TYPE AND DATES COVERED Final--FY85-FY88
4. TITLE AND SUBTITLE Validity of the GRIP and ASVAB Test Batteries for Job Performance of Sonar Technician (Surface) Personnel		5. FUNDING NUMBERS Program Element 0602233N, Work Unit RM33M20.03
6. AUTHOR(S) Charles H. Cory		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Navy Personnel Research and Development Center San Diego, California 92152-6800		8. PERFORMING ORGANIZATION NPRDC-TN-90-32
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Chief of Naval Research (ONT-222) 800 North Quincy Street, Arlington, VA 22217-5000 Office of the Assistant Secretary of Defense (FM&P) (MMPP/AP) Washington, DC 20350-4000		10. SPONSORING/MONITORING
11. SUPPLEMENTARY NOTES		
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The validity of several aptitude tests for predicting the job performance of Navy Sonar Technicians (N = 138) was determined. The tests were the Armed Services Vocational Battery (ASVAB) and a set of six new computerized tests. The results indicate that while the ASVAB predicted Sonar Technician performance, better predictions were obtained when the ASVAB scores were supplemented by the Digit Span score from the computerized battery. Follow-on research with Digit Span, both to improve the test and to verify its relationship to job performance is recommended.

14. SUBJECT TERMS ASVAB, Sonar Technicians, computerized testing, personnel selection			15. NUMBER OF PAGES 35
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UNLIMITED

FOREWORD

This report reviews the development and validation of the Graphic and Interactive Processing (GRIP) computerized test battery, which is designed to measure perceptual and reasoning abilities beyond those measured by the paper-and-pencil tests in the Armed Services Vocational Aptitude Battery (ASVAB). Measurement of new abilities, which are important for job performance, could improve Navy Personnel classification and result in substantial savings. The work was conducted under the Personnel Performance Prediction (PPP) project (Program Element 0602233N and Work Unit RM33M20.03), which was sponsored by the Chief of Naval Research (ONT-222) and Office of the Assistant Secretary of Defense (FM&P) (MMPP/AP).

The findings should be of interest to Navy recruiting and anti-submarine warfare personnel and to research communities concerned with the study of mental abilities or with the application of computers to personnel classification testing.

JULES I. BORACK
Director, Personnel Systems Department

SUMMARY

Problem

The Navy is currently evaluating how well job performance is being predicted by the written tests used to select personnel for enlisted ratings. Previous research has found that selection practices for Sonar Technician (Surface) (STG) are deficient in the measurement of some perceptual abilities, which are important for job performance. Computerized tests can measure visual and reasoning abilities that cannot be tested by paper-and-pencil tests currently used for personnel classification. Therefore, computerized tests may be useful for improving selection of STG personnel.

Objective

The present research had a two-fold purpose: (1) to develop a set of experimental computerized tests designed to improve the predictability of Armed Services Vocational Aptitude Battery (ASVAB) selector composites for job performance of STG personnel and (2) to measure the predictiveness of the ASVAB subtests for job performance of STG personnel.

Approach

The Graphic and Interactive Processing (GRIP) battery of computerized tests, designed to measure perceptual and reasoning abilities important for job performance of STG personnel, was developed. The GRIP tests were administered to first term non-prior service personnel immediately prior to their entrance at STG class "A" school. Follow-up 2 to 3 years later to collect job performance evaluations was accomplished with visits to the fleet to have supervisors complete job performance rating forms. Validity coefficients for (1) the most predictive set of ASVAB subtests, (2) the STG selector composite, and (3) a measure of general mental ability were calculated for 14 performance ratings. Incremental validities of the most predictive GRIP tests were evaluated for statistical significance. The predictiveness of ASVAB and ASVAB + GRIP predictor sets was compared.

Results

1. Statistically significant measures of incremental validity were found for four GRIP scores, of which one, Digit Span, a measure of short term memory for digits, may have promise for use in enlisted classification. The other scores with significant incremental validity, a measure of reasoning efficiency and two perceptual measures, are less promising, either because they are narrowly focussed or because their effects are quite small relative to the time required to collect the score.

2. ASVAB variables were found to have substantial predictiveness for STG job performance. Median estimated validity coefficients of ASVAB subtests for the 14 criteria ranged from .22 to .41. Median validity coefficients for both the STG selector composite and G were .37.

3. The ASVAB variables are most predictive for the traits of common sense, flexibility/motivation, maintenance and repair, planned maintenance system abilities and overall sonar operation, and for such basic sonar operation tasks as deriving environmental predictions and setting up equipment. They were not predictive of narrow band target detection ability or ability to set up for towed arrays.

4. GRIP tests significantly incremented the validity of the ASVAB variables for predicting 6 of the 14 criteria. Their greatest increases were for overall sonar operation and task performance abilities of deriving environmental predictions and equipment setup.

5. Median incremental validity of the GRIP variables for ASVAB subtests was 28 percent based on Wherry estimates of the shrunken $\underline{R}$s, and 16 percent based on Stein estimates of shrunken $\underline{R}$s. GRIP variables provide median increases of about 18 percent for the STG selector composite and about 12 percent for G.

6. The GRIP tests did not provide incremental validity to the ASVAB subtests for predicting narrow band target detection abilities. Part of the failure may have resulted from the relatively low reliabilities of the GRIP perceptual measures.

Conclusions

1. The current ASVAB selector composite provides effective selection for STG personnel. It effectively predicts all aspects of STG rating performance except "narrow band target detection" and "ability to set up towed arrays."

2. Although Digit Span substantially improves predictiveness for STG job performance, its low reliability makes it inappropriate for operational use in its current form.

Recommendations

Research should be continued with Digit Span to (a) improve the reliability of the test and (b) evaluate its predictiveness for job performance of personnel in other ratings.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Problem

The Armed Services Vocational Aptitude Battery (ASVAB), composed of 10 tests, is the primary classification instrument for Navy enlisted personnel. Eleven composites, each composed of two, three, or four tests, are used to assign personnel to ratings. However, misassignments can occur if the abilities required for "A" school in a rating differ from those required for job performance. Mackie, McCauley, and Ridihalgh (1975) and Mackie, Ridihalgh, and Schultz (1981) indicate that such misassignments are occurring for Sonar Technicians (STs).

Mackie et al. (1975) evaluated the trends in sonar hardware development and concluded that the selection standards for STs do not measure some abilities necessary for sonar target detection and classification. Mackie et al. (1981) found that the ASVAB selector composite for ST (Submarine) had a negative correlation ($\underline{r} = -.18$) with overall score on a battery of job sample tests. In contrast, they found that a composite formed from four experimental predictors and the ASVAB selector composite correlated positively ($\underline{R} = .47$) with the job sample criterion.

Objective

The present research had a two-fold purpose: (1) to develop a set of experimental computerized tests designed to improve the predictability of ASVAB selector composites for job performance of Sonar Technician (Surface) (STG) personnel and (2) to measure the predictiveness of the ASVAB subtests for job performance of STG personnel.

APPROACH

Background

Computerized personnel classification tests appear to offer a number of advantages over paper-and-pencil (P&P) tests. Computerized tests can:

1. Increase the types of stimuli that can be presented. They can present motion and other types of perceptual stimuli. They can present stimuli which model the information on sonar screens.

2. Measure response latencies. This capacity modifies the way that speed influences test score measurement. Instead of speeded tests being restricted to easy, interchangeable items, speed, accuracy, and the speed/accuracy trade-off can be collected and scored separately.

3. Measure recall abilities instead of being limited to recognition abilities. Thus, the scanning and word checking capabilities of computers can be used to require test takers to produce answers instead of merely recognizing correct answers from multiple choice items.

Characteristics of tasks performed by STGs suggest that computerized tests could measure better than P&P tests the visual abilities required for detection and classification of sonar contacts and some of the reasoning tasks involved in sonar operation. Previous exploratory research (Cory, 1974, 1976) found tests from the Graphic and Interactive Processing (GRIP) battery of computerized tests to add to the predictiveness of the STG selector composite for predicting job performance of STGs.

GRIP Tests

A new version of the GRIP battery was developed for evaluation for classification of STGs. The six tests in the battery, together with their rationale and background, are described below.

Pattern

A major part of the STG job involves scanning sonar sweeps to detect and classify contacts. This involves comparing successive sweeps displayed on the screen. Therefore, short-term visual memory may be important for STGs.

The Patterns test (Figure 1) displays a pattern of numbers across the screen (Frame 1). The pattern is shown briefly, then taken away. Two seconds later pattern 2 is shown (Frame 2), which is identical to pattern 1 except that one digit has moved. The task is to identify the wandering digit ("5" in the example).

Frame 1	Frame 2
1	1
2	2
3	3
4	4
5	5

Figure 1. Sample item from patterns.

Patterns vary in difficulty, depending on the distance that the changed digit moves. Initially, a pattern of medium difficulty is shown. If answered correctly, it is followed by a harder pattern; if missed, by an easier one. The alternation terminates when the test schematic calls for a fifth presentation of a pattern level.

Vigilance (VIG)

VIG has been defined by Mackworth (1957) as a "state of readiness to detect and respond to certain specified small changes occurring at random time intervals in the environment." Therefore, VIG should be important for performing a boring task such as scanning a sonar screen for long periods.

The classic VIG research has used a clock watching task, which requires the detection of jumps in a sweep second hand as it moves around the clock face. In other research, VIG tasks have involved the detection of variations in light intensities and in auditory pitches (Buckner & McGrath, 1961). All VIG tests have required long administration times. Mackworth (1957) tested her subjects for 2 hours; others such as Dobbins, Tiedemann, and Skordahl (1961) tested subjects for 7 1/2 hours. These time limits were too long to be feasible for Navy classification tests. Thus, a research objective was to produce a VIG test to reliably predict target detection abilities that could be administered in less than 30 minutes.

For VIG, a series of randomly placed upper case letters are emitted rapidly across the screen, a line at a time. If the example (Figure 2) were being displayed, each line of letters would be rapidly scrolling up. At random intervals, a lower case letter would be emitted. The task is to respond quickly when a lower case letter is detected. In Figure 2, the target letter is the small "w," center right in the distribution.

```
        F  G   E U F  M Y K  L   K A    Y
       W    R   H J   I S  S   W H   F  S W
     E A    G    R    E E S    A A w S D T T   T
   R   T    Y    Y U    U    E E D O   F  J  J  L L
     FJ     I    E     D   E    E E  I   F  F L D
  Q Z    ADR    Y    U U    I  PO  P D W S  D F    H
```

Figure 2. Replica of a screen display for VIG.

Total time for VIG is 26 minutes, during which, 18 target letters are displayed, approximately one every 87 seconds.

Scan String (SCAN)

Perceptual speed in comparing figures or symbols and in scanning to find figures or symbols is another ability that is important for STGs. The P&P tests currently used to measure perceptual speed intermix speed and accuracy components. The score is not a pure measure of perceptual speed; it is a composite of perceptual speed, accuracy, and the manual dexterity required for blackening responses on an answer sheet. In previous research (Cory, 1976), counting numbers, a P&P test of perceptual speed, was found to be statistically significant as a predictor of STG job performance. SCAN is a computerized test, which is similar to counting numbers.

Each item in SCAN consists of a target digit followed by a string of digits. The task is to count the number of times the target digit appears in the string and type the answer into the computer. In the example (Figure 3), the target (6) appears 9 times.

(6)

813267410386296724813062485l163

2501729638359206427016274016380

Figure 3. Sample item from SCAN string.

SCAN contains 16 2-line number strings. It is administered in free format--without restrictions on the testing time of an examinee.

Password

Good reasoning ability is required for STGs to perform adequately the numerous tasks associated with sonar operation and maintenance. The manual for the most recent Educational Training Service (ETS) factor-referenced battery of ability tests (Ekstrom, French, Harman, & Derman, 1976) describes tests for inductive and deductive reasoning abilities, both of which appeared to be too highly structured to have face validity for personnel classification. Ekstrom et al. (1976) also describes a third reasoning ability, "Integrative Processes," as "the ability to keep in mind simultaneously or to combine several conditions, premises, or rules in order to produce a correct response." It was desired to measure this type of ability using a technique that did not require information to be assimilated or rules to be applied as was required for Integrative Processes.

Password (PAS) is a computerized adaptation of "Password," the popular parlor game. For each item in PAS, one-word hints, which suggest another word, are displayed. The task is to identify the target word and type it into the computer. In the example (Figure 4), Frame 1 shows the two initial hints: "metal" and "finger." If the subject does not type the correct answer, "ring," he is given another hint, "circle." Presentation of hints and responding alternate until the fourth frame has been completed or the object has been identified, whichever comes first. PAS is not a measure of vocabulary. All of the items in the test are common objects. The test measures how efficiently a person can reason from hints to identify objects. The scoring algorithm for PAS ignores a wide range of spelling errors to produce accurate right/wrong scores.

Frame 1	Frame 2	Frame 3	Frame 4
Metal Finger	Circle	Shiny	Wedding

Figure 4. Sample item from password.

PAS contains 20 items. It is administered without time restrictions, but the test instructions encourage examinees to work as fast as they can.

Digit Span (DSpan)

An individually administered digit span test included in the Wechsler Adult Intelligence Scale (WAIS) has a venerable record (Wechsler, 1981) as a measure of mental ability and of short-term memory. DSpan is a computerized test, which has a format similar to the digit span test of the WAIS.

DSpan displays a string of digits one at a time. Then, after a pause, the task is to recall and type the digit string into the computer. In the example (Figure 5), Frame 1 shows the span presented, and Frame 2 shows the screen set up to receive the answer. DSpan uses a step-up, step-down format in which a correct answer is followed by a span one number longer, and an incorrect answer by a span one number shorter. The alternation terminates when the schematic calls for a fifth presentation of any span size.

Frame 1	Frame 2
5 9 2 3	- - - -

Figure 5. Sample item from DSpan.

Memory Probe

Memory probe is based on the work of Sternberg (1975) and uses a paradigm that has been extensively employed to measure speed of mental processes.

First a string of digits is displayed (Figure 6), then six target digits are displayed. After each target digit, the subject responds (Yes/No) to indicate whether or not the digit was in the original string. The test contains three 6-digit strings, each having a set of six target digits (3 x 6 = 18 items).

Frame 1	Frame 2
	WAS THIS NUMBER IN THE STRING?
6 3 7 4 8 9	2?
	3?
	7?
	5?
	4?
	1?

Figure 6. Sample item from memory probe.

The GRIP battery was programmed in PASCAL and administered on IBM personal computers. Administration time is about 1 1/2 hours.

Method

From 14 March 1985 to 17 July 1986, 556 personnel were administered the GRIP battery prior to entering STG "A" school. A sample of convenience was used. Test-takers consisted of first term non-prior service personnel who were waiting to class up for STG "A" school and were available for testing. ASVAB scores were collected from records at the Fleet Anti-Submarine Warfare Training Center, Pacific (FLEASWTRACENPAC) and The Navy Personnel Research and Development Center (NAVPERSRANDCEN), and scores for the STG selector composites were computed by summing the appropriate ASVAB subtests. From 4 May to 30 July 1988, visits to the fleet were conducted to collect supervisor evaluations of job performance of 138 personnel previously tested. Because of the expense of collecting supervisory ratings by fleet visits, the following restrictions were adopted in order to increase data collection efficiency. Personnel selected for follow-up were: (1) located on ships home ported in one of six large Navy bases, (2) either on frigates, destroyers, or cruisers, the major users of STGs, and (3) on ships having two or more personnel from the original sample (some exceptions were made to rule 3).

Job Performance Rating

The job performance rating (JPR) form was based upon extensive input from the Fleet. Prior to its development, visits were made to the USS LYNDE MCCORMICK, USS ALBERT DAVID, and USS FANNING for familiarization with the duties and requirements of the STG rating. The enthusiastic cooperation of personnel in the anti-submarine warfare (ASW) divisions of these commands was of great assistance in the development of the criterion rating instruments. The JPR and an information recording form were drafted on the basis of information from the visits, and a trial administration of the forms was conducted with instructors at FLEASWTRACENPAC. Subsequently, final versions of the forms were developed.

The information recording form served as the basis for a structured interview and was used to record amplifying information on the individual's job assignments, training, achievement, and supervisory evaluations. Ratings, usually by the work center supervisor, but occasionally by the division chief or the ASW Officer, were recorded onto the JPR.

Visits were made to commands home ported in the San Diego, Long Beach, and Pearl Harbor Naval Bases on the West Coast, and in Mayport, Charleston, and Norfolk bases on the East Coast. Sixty-two percent of the personnel were from commands home ported on the West Coast and 38 percent from commands home ported on the East Coast.

Statistical Calculations

Predictors were 10 ASVAB scores, 15 scores from the GRIP battery, the ASVAB selector composite score, and an ASVAB composite, which Schmidt, Hunter, and Dunn (in review) recommend as a measure of "G," general mental ability. Criteria were 14 ratings of sonar performance included in or derived from the JPR.

Principal component analyses were made of the supervisory ratings and factor scores of three rotated principal components were computed. Zero order validity coefficients for ASVAB and GRIP variables were computed for the job performance ratings. The following predictor sets of ASVAB variables were used for multiple regression runs:

1. All ASVAB tests that had positive correlations with the criterion (usually the ASVAB predictor set contained from 4 to 7 tests).[1]

2. The STG selector composite score (SELC).

3. G, where G = GS + AR + WK + MK + MC + EI.

Hypothesis Testing

Stepwise multiple regression runs were used to determine the most predictive ASVAB subsets for the 14 criteria. For each run, the inclusion of additional ASVAB subtests were stopped when the increase in the multiple correlation coefficient was not significant at $p < .05$, based on an F ratio test. The ASVAB subset for a criterion was then used in a forced entry multiple regression mode with the GRIP variable with the highest zero order validity coefficient that was statistically significant for the criterion. And finally, the GRIP variables were used in separate forced entry multiple regressions with SELC and G. For the 11 criteria for which GRIP variables

---

[1]Multiple regression runs were also performed using all ASVAB tests with positive correlations with the criterion as the predictor set. With one exception, Wherry estimates of the shrunken validity coefficients were .02 to .04 greater for this subset versus the smaller subset. The increment for the one exception was .09.

were found to provide significant incremental validations over the ASVAB variables, tests of the statistical significance of the increments were made using a general F test for an increment (model 1 error) recommended by Cohen and Cohen (1983, p.145).

Point Estimates of the Magnitude of Improvement

Several statistical procedures were performed on the raw validity coefficients in order to counteract effects that distort them. These produced estimates of the predictive relationships for the Navy enlisted population, which were more accurate than the original coefficients. The following adjustments were made to the raw validity coefficients.

1. Correction for restriction of range recommended by Thorndike (1982, p. 208) was used (case A for SELC and case C for the most predictive ASVAB composite and for G).

2. Attenuation in both predictors and criteria was corrected using a method recommended by Hunter, Schmidt, and Jackson (1982).

3. Finally, estimates of correction for shrinkage were made using both the Wherry (1931) correction, one of the most commonly used estimates of an R in the population, and the Stein (1960), a very conservative correction, which has been found by Kennedy (1988) to be more accurate than the Wherry correction for estimating cross validity coefficients.

Tests to determine the statistical significance of the corrected coefficients were not performed because such tests are appropriate only for raw (not corrected) coefficients.

Predictors

ASVAB

The ASVAB 11/12/13 subtests are listed and described in Table 1. They are scaled in t-scores, which set the mean of Navy enlisted applicants at 50 and the standard deviation at 10.

The STG selector composite is the sum of ASVAB subtest scores MK + EI + GS (passing = 156 or higher) + AR (passing = 218 or higher). The first STG pass point requires an average at the 58th percentile on the first three tests. Persons scoring at the first pass point would need to score at the 88th percentile on AR in order to meet the second pass point. Shortfalls on AR must be compensated for with score increases on one or more of the other tests. The second passing score is set at an average on the four tests at the 67th percentile. The SELC was selected to represent the operational selector for the statistical analyses.

Schmidt, Hunter, and Dunn (in review) recommend the following composite of ASVAB scores as a measure of general mental ability: G = AR + MK + WK + GS + MC + EI. The predictiveness of G was evaluated in comparison with that of ASVAB subtests and SELC.

Table 1

Subtests[a] in ASVAB Forms 11, 12, and 13

Subtest	Abbreviation	Description
General Science	GS	A 25-item test of knowledge of the physical (13 items) and biological (12 items) sciences--11 minutes
Arithmetic Reasoning	AR	A 30-item test of ability to solve arithmetic word problems--36 minutes
Word Knowledge	WK	A 35-item test of knowledge of vocabulary, using words embedded in sentences (11 items) and synonyms (24 items)--11 minutes
Paragraph Comprehension	PC	A 15-item test of reading comprehension--13 minutes
Numerical Operations	NO	A 50-item speeded test of ability to add, subtract, multiply, and divide one- and two-digit numbers--3 minutes
Coding Speed	CS	An 84-item speeded test of ability to recognize numbers associated with words from a table--7 minutes
Auto and Shop Information	AS	A 25-item test of knowledge of automobiles, shop practices, and use of tools--11 minutes
Mathematics Knowledge	MK	A 25-item test of knowledge of algebra, geometry, fractions, decimals, and exponents--24 minutes
Mechanical Comprehension	MC	A 25-item test of knowledge of mechanical and physical principles--19 minutes
Electronics Information	EI	A 20-item test of knowledge of electronics, radio, and electrical principles and information--9 minutes

[a]Scores are reported as t-scores, having a mean of 50 and a standard deviation of 10 for a representative sample of 19-23 year-old American youth.

GRIP

Table 2 describes the 15 GRIP variables that were evaluated. The scores included four each from the VIG, SCAN, and Probe subtests and one each from DSpan, Patterns, and Password.

Table 2

GRIP Scores Evaluated as Predictors

Variable	Description	Subtest
DG_RIGHTS	Total number of digits recalled.	DSpan
VG_FALSP	Total number of false positive responses.	VIG
VG_LC	Average latency from target onset to response.	
VG_LLC	Average of the natural logarithms of latencies from target onset to response.	
VG_TR	Total number of correct responses.	
PT_SUMR	Sum of the right answers weighted by difficulty.	Patterns
PS_HINTS	Total number of hints received.	Password
SC_DIFS	Sum of the absolute differences between the answer of the examinee and the keyed answer.	SCAN
SC_CPM	Average number of correct answers per minute	
SC_LC	Average response latency per correct answer.	
SC_LLC	Average of the natural logs of the response latencies for items answered correctly.	
PR_RIGHTS	Total number of correct answers.	Probe
PR_CPM	Average number of correct answers per minute.	
PR_LC	Average latency to correct answer.	
PR_LLC	Average of the natural logarithms of latencies to correct answers.	

Criteria

Criteria consisted of scores for the 11 scales on the supervisor rating from plus three factors scores, which summarized scores for the most commonly used scales. The 14 criteria are described in Table 3.

Because of the variety in types of STG assignments, the Ns varied substantially from criterion to criterion. For instance, maintenance and repair abilities could only be rated for the 39 technicians included in the sample, and variables 13 and 14 (PMS and 3 traits + PMS) were only applicable for gram analysts (28% of the sample). Others (e.g. set up for towed arrays) applied only to passive operators on frigates.

Table 3

Description of Criteria

	Criterion	Description
1.	Common sense	Ability to set priorities and concentrate on important aspects of the job.
2.	Flexibility/ motivation	Flexible in working in different job assignment and working conditions. Maintains a positive attitude.
3.	Sonar Op overall	Overall evaluation of performance in operating sonar equipment.
4.	Factor score 3-traits	Factor score for the first principal component of variables Tables 1, 2, & 3. This is a summary score for the three most common traits.
5.	Make environ. measurements	Performance of tasks involved in equipment set-up: determination of CZ or BB conditions, interpret XBT trace, compute range predictions, etc.
6.	Set up equipment	Ability to select equipment settings that are appropriate for environmental and topographic conditions.
7.	Detect contracts	Performance in the use of active sonar equipment to detect and report valid contacts and in classifying contacts.
8.	Factor score 3-trt, 3-tsk	Factor score for the first principal component for variables 1, 2, 3, 5, 6, and 7. This is a summary score for the three most common tasks of active sonar.
9.	Set up for T. arrays	Performance in setting up equipment for towed arrays, including determination of appropriate target tonals, operating modes, trade-offs, and array depths.
10.	Target detect., broadband	Performance in passive target detection, broadband, including resolving for bearing ambiguity, calculating turncounts, classifying targets, etc.
11.	Target detect., narrow band	Performance in passive target detection, narrow band including interpretation of narrow band displays, identification of targets, etc.
12.	Maintenance & repair	Performance of a technician in PMS and in trouble-shooting and repairing equipment.
13.	PMS	Performance of a non-technician in PMS.
14.	Fact. score, three traits + PMS	Factor score for the first principal component of variables 1, 2, 3, and 13. This is a summary score that takes into account performance on the two major types of STG tasks, operation, and maintenance.

FINDINGS

The relatively high ability level of the sample is shown by the high mean scores and small standard deviations on ASVAB subtests (Table 4). Mean scores expressed as a deviation score ranged from .295 (CS) to .908(GS)SD above the mean of a Navy applicant sample.

Table 4

Descriptive Statistics for ASVAB and GRIP Variables

Variable	M	SD	Kurtosis	Skewness	N
		ASVAB Variables			
GS	59.08	5.08	1.07	-.83	130
AR	57.87	4.83	-.24	-.42	130
WK	55.95	4.05	2.02	-1.10	130
PC	55.15	4.70	.05	-.68	130
NO	54.24	6.31	.03	-.60	130
CS	52.95	6.63	-.39	.27	130
AS	56.19	7.26	-.29	-.55	130
MK	59.03	5.21	.01	-.61	130
MC	57.28	5.84	-.12	-.01	130
EI	58.29	5.94	.88	-.62	130
SELC	234.27	12.06	-.35	.18	130
G	347.49	17.86	-.38	.17	130
		GRIP Variables			
DG_RIGHTS	57.24	20.32	-.43	.48	128
VG_FALSP	2.44	2.08	.35	.91	128
VG_LC	3.27	2.02	5.21	1.89	127
VG_LLC	3.07	1.88	8.30	2.47	119
VG_TR	10.91	3.46	-.60	-.27	128
PT_SUMR	20.57	17.79	1.72	1.46	126
PS_HINTS	34.39	9.41	3.85	1.85	128
SC_DIFS	5.09	3.30	1.72	1.12	126
SC_CPM	4.37	1.05	-.47	.18	122
SC_LC	10.56	2.44	1.54	1.03	122
SC_LLC	10.38	2.33	.68	.84	122
PR-RIGHTS	16.61	2.15	.52	-2.65	127
PR_CPM	51.73	10.57	-.06	-.12	127
PR_LC	1.22	.29	3.98	1.60	127
PR_LLC	1.21	.21	1.25	1.32	127

Reliability coefficients for the GRIP tests were calculated on a separate sample of 170 recruits at the Recruit Training Center, San Diego. The battery was administered twice, with the retest being given 3 weeks and 1 day after the first administration. Internal consistency (coefficient alpha), test-retest, and alternate forms reliability coefficients for the GRIP tests (Table 5) were generally low.

Table 5

Reliability Coefficients for GRIP Scores

	r_{xx}		
Criterion	Coefficient Alpha	Test-Retest	Alternate Form
DG_RIGHTS			.60
VG_FALSP	.80	.63	
VG_LC	.59	.28	
VG_LLC	.71	.29	
VG_TR	.67	.49	
PT_SUMR		.44	
PS_HINTS	.83	.65	
SC_DIFS	.86	.37	
SC_CPM	.91	.62	
SC_LC	.92	.52	
SC_LLC	.93	.60	
PR_RIGHTS	.63	.35	
PR_CPM	.82	.59	
PR_LC	.82	.51	
PR_LLC	.80	.57	

Supervisors on the average were well satisfied with the job performance characteristics of STGs (Table 6). Mean scores for the 11 rating categories on the JPR form ranged from 6.02 to 7.17 on a behaviorally-anchored scale from 1 to 9 (ineffective = 1, effective = 5, very effective = 9). Supervisors gave the highest performance marks to maintenance and repair abilities of technicians. The two sonar operations categories with the highest scores were make environmental measurements and target detection, broadband. The rating categories that received the lowest average marks were PMS (for non-technicians), flexibility/motivation, and set up for towed arrays, a category that applies only to passive operators on frigates.

Table 6

Descriptive Statistics for Supervisors' JPR Responses

Ratings	M	SD	Kurtosis	Skewness	N
1. Maintenance & repair	7.17	1.62	.44	.83	30
2. Make environ. measurements	6.98	1.75	-.11	.43	127
3. Target detect., broadband	6.70	1.79	-.54	-.42	111
4. Sonar Op overall	6.57	1.71	.13	-.49	131
5. Set up equipment	6.52	1.88	-.48	-.47	122
6. Common sense	6.42	1.98	.19	-.73	137
7. Target detect., narrow band	6.42	2.03	-.35	-.48	43
8. Detect contacts	6.38	1.89	-.54	-.38	124
9. Set up for T. arrays	6.25	1.80	-.54	-.61	28
10. Flexibility/ motivation	6.20	2.11	-.53	-.52	138
11. PMS	6.02	1.61	.53	-.68	65
12. Factor score 3-trait	0.00	1.00	.23	-.66	131
13. Factor score 3-trt, 3-tsk	0.00	1.00	.14	-.64	115
14. Factor score 3 traits + PMS	0.00	1.00	1.26	-.99	64

Incremental Validity

Table 7 shows validity coefficients for the ASVAB and the ASVAB + GRIP batteries. The first row for each criterion shows the coefficients for the ASVAB tests and the second row the coefficients for the ASVAB plus GRIP tests. Coefficients for the subtest composites are the values adjusted for overfitting (Wherry correction applied to the raw coefficients). Unadjusted multiple regression coefficients are shown for SELC and G because overfitting did not occur for these variables--the predictor sets were fixed.

GRIP scores added to the predictiveness of ASVAB variables for the 11 listed in Table 7. GRIP tests did not improve predictiveness in the three not shown in Table 7 (i.e.,) maintenance and repair, PMS, or Factor score, 3 traits + PMS.

For subtest composites, GRIP variables added from .04 to .56 to the predictiveness of ASVAB variables with the median increment being .08. For SELC and G, GRIP increments range from .07 to .82 and .08 to .48, with medians of .12 and .13 respectively. Increments are very large for the criteria, "Set up for towed arrays" and "Target detection, narrow band." The very small <u>N</u>s of both criteria suggests that capitalization of chance may be magnifying the increments.

Results of tests to determine the statistical significance of incremental validity coefficients are shown in Table 8. The GRIP variable, which incremented ASVAB predictors, is shown in the right column.

Incremental validities for 10 of the 11 criteria are statistically significant. Six criteria (variables 2, 3, 4, 5, 7 and 8) have incremental validities that are statistically significant for all three predictor types; two (variables 1 and 9) have statistically significant incremental validities for the subtest and selector composites, but not for G; two (variables 6 and 11) have statistically significant incremental validities for only one predictor type, and one (variable 10) has no incremental validities that are statistically significant. All criteria for which incremental validities are statistically significant for only one type of selector, or for none, have small <u>N</u>s.

Based on a conservative interpretation of the data in Table 8, statistically significant incremental validity was found for six criteria (2, 3, 4, 5, 7, and 8). Three of these six measures are global measures of sonar operation.

DG_RIGHTS provided the incremental validity for 7 of the 10 criteria with statistically significant incremental validities and for six of the eight criteria that have statistically significant incremental validities on two or more predictor types and for five of the six criteria with significant incremental validities for three predictor types. PS_HINTS also has statistically significant validity for three predictor types. VG_FALSP has significant incremental validity for two predictor types. The effect found for SC_DIGS, although statistically significant for one predictor type, was very weak.

Table 7

Validity Coefficients[a] for Sets of ASVAB and ASVAB + GRIP Variables

Criterion	Data B	Composites	SELC	G	N
		Type of Predictor			
Common sense	AS[b]	.24**	.12	.12	130
	AS + GR	.30**	.21	.21	120
Flexibility/ motivation	AS	.24**	.18*	.15	130
	AS + GR	.30**	.27*	.25	120
Sonar Op overall	AS	.12	.11	.08	125
	AS + GR	.26**	.27**	.26*	115
Factor score 3-trait	AS	.24**	.16	.14	125
	AS + GR	.33**	.28**	.27*	115
Make environ. measurements	AS	.30**	.28*	.22*	121
	AS + GR	.37**	.33**	.32**	111
Set up equipment	AS	.19*	.15	.17	116
	AS + GR	.25**	.24	.25	107
Detect contacts	AS	.17	.04	.05	117
	AS + GR	.28**	.22	.21	108
Factor score 3-trt, 3-tsk	AS	.22*	.15	.14	109
	AS + GR	.30**	.28*	.27*	101
Set up for T. arrays	AS	.09	-.09	-.04	26
	AS + GR	.65**	.73*	.44	16
Target Detect., .broadband	AS	.14	.11	.10	105
	AS + GR	.18	.19	.19	96
Target detect, narrow band	AS	.21	-.01	-.04	39
	AS + GR	.41*	.47*	.38	33

[a]*R*s for ASVAB composites were adjusted for overfitting by means of the Wherry procedure. Unadjusted statistics were used for SELC and G.

[b]AS = ASVAB, GR = GRIP

*F ratio $p < .05$.
**F ratio $p < .01$.

Table 8

F Ratio Tests of Incremental Validity of GRIP Variables

	Type of Predictor			
Criterion	Subtest Composite	Selector Composite	G	GRIP Variable
Common sense	5.20*	4.04*	3.65	DG_RIGHTS
Flexibility/ motivation	6.48*	5.55*	4.85*	PS_HINTS
Sonar Op overall	8.05*	8.31**	7.55**	DG_RIGHTS
Factor score 3-traits	6.90**	6.48*	6.40*	DG_RIGHTS
Make environ. measurements	7.69**	9.42**	8.30**	DG_RIGHTS
Set up equipment	1.51	4.10*	3.89	DG_RIGHTS
Detect contacts	7.27**	5.04*	4.64*	DG_RIGHTS
Factor score 3-trt, 3-tsk	7.37**	5.68*	5.66*	DG_RIGHTS
Set up for T. arrays	12.25**	5.41*	3.13	VG_FALSP
Target detect., broadband	1.26	2.23	2.45	PR_RIGHTS
Target detect, narrow band	2.65	4.04	4.88*	SC_DIFS

*$p < .05$.
**$p < .01$.

The criteria for which DG_RIGHTS was significantly predictive suggests that its significant effect for detect contacts resulted from its association with ability to exercise judgment in organizing work and in manipulating equipment rather than from a perceptual ability that it might be measuring. The significant association of PS_HINTS with flexibility/motivation seems consistent with an ability to reason efficiently from diverse hints to the identification of objects.

Corrected Validity Coefficients

Tests performed on the raw validity coefficients are appropriate for determining the statistical significance of incremental validity of GRIP variables. However, accurate estimation of the predictiveness of variables for criteria requires correction for artifacts which distort the true relationships. Restriction in range and attenuation, because of unreliability in predictors and criteria, result in raw validity coefficients which underestimate the true correlations. On the other hand, the multiple regression procedure overestimates the true relationship because capitalization on chance occurs when the most predictive subset is selected from a set of predictors.

Tables 9, 10, and 11 provide estimates of the predictive validity of ASVAB, SELC, G, ASVAB + GRIP, SELC + GRIP, and G + GRIP variables, which are more accurate than those shown in Table 6. Table 9 (as well as 10 and 11) shows the successive modifications to the raw validity coefficients which result from applying corrections for restriction in range, attenuation and finally Stein and Wherry adjustments. The fourth and fifth columns of Table 9 show the Stein and Wherry estimates of validity of the ASVAB predictors in the population. Coefficients for the Stein or Wherry adjustments, which were calculated as "0" or negative, are shown as "0"s in the table.

The conservative Stein correction provides validity estimates which are 0 for six of 14 criteria. For two of these six criteria (variables 3 and 7) the ASVAB + GRIP battery has a validity coefficient > 0. The Wherry estimates range from .09 to .47 more than the Stein estimates. The differences are highest for variables with the smallest Ns.

For the ASVAB + GRIP battery, only the Stein and Wherry estimates are shown (columns 6 and 7). The coefficients corrected for restriction of range and for attenuation such as are shown for ASVAB (columns 2 and 3) are not included. Table 8 shows the degrees of freedom for the Stein and Wherry adjustments and for the incremental validity tests.

Predictiveness of ASVAB and ASVAB + GRIP Subsets

The Wherry estimates of validities for the ASVAB subsets range from .15 to .55 with a median of .41. Stein estimates range from 0 to .44 and have a median of .22.

Wherry estimates of validities for the ASVAB + GRIP subsets range from .31 to .64 with a median of .48. Stein estimates for the ASVAB + GRIP validities range from 0 to .57 and the median is .35. GRIP variables provided a median increase of 28 percent based on the Wherry estimates and 16 percent based on the Stein estimates.

Table 9

Corrected Validity Coefficients[a] for ASVAB and ASVAB + GRIP Variables

	Corrected					ASVAB+GRIP		% Incr. R		
Criterion	Raw	RRng	Atten	Stn	Wher	Stn	Wher	Stn	Wher	N
1. Common sense	27	37	42	27	41	37	48	33	18	120
2. Flexibility/ motivation	27	44	49	35	47	41	53	16	11	120
3. Sonar Op overall	15	27	31	0	30	31	45	--	51	115
4. Factor score 3-trait	27	40	45	31	43	48	55	54	28	115
5. Make environ. measurements	32	49	54	44	53	57	64	29	20	111
6. Set up equipment	14	32	37	22	36	35	46	59	28	107
7. Detect contacts	22	27	32	0	29	19	40	--	36	108
8. Factor score 3-trt, 3-tsk	26	38	43	28	41	45	55	64	33	101
9. Set up for T. arrays	18	20	25	0	15	0	48	--	226	16
10. Target detect., broadband	17	29	32	0	31	0	36	--	16	96
11. Target detect., narrow band	26	22	28	0	23	0	41	--	77	33
12. Maintenance & repair	39	43	50	0	47	--	--	--	--	29
13. PMS	29	46	52	22	51	--	--	--	--	61
14. Factor score, 3 traits + PMS	39	51	57	34	55	--	--	--	--	60

[a]Decimal points omitted from the Rs.

Table 10

Corrected Validity Coefficients[a] for SELC and SELC + GRIP Variables

Criterion	Raw	Corrected RRng	Corrected Atten	SELC + GRIP Stn	SELC + GRIP Wher	% Incr. R Stn	% Incr. R Wher	N
1. Common sense	12	29	31	32	35	3	13	120
2. Flexibility/ motivation	18	40	44	46	48	5	9	120
3. Sonar Op overall	11	26	28	32	36	21	32	115
4. Factor score 3-traits	16	36	39	41	43	10	15	115
5. Make environ. measurements	20	43	47	53	55	13	17	111
6. Set up equipment	15	34	37	39	43	8	16	107
7. Detect contacts	08	19	21	22	28	5	33	108
8. Factor score 3-trt, 3-tsk	15	33	37	38	41	8	16	101
9. Set up for T. arrays	-09	-20	-22	0	37	--	--	16
10. Target detect., broadband	11	26	28	25	30	-11	7	96
11. Target detect., narrow band	-01	-03	-04	0	30	--	--	33
12. Maintenance & repair	17	38	42	--	--	--	--	29
13. PMS	18	41	45	--	--	--	--	61
14. Factor score, 3 traits + PMS	19	42	47	--	--	--	--	60

[a]Decimal points omitted from the Rs.

Table 11

Corrected Validity Coefficients[a] for G and G + GRIP Variables

Criterion	Corrected			G + GRIP		% Incr. R		
	Raw	RRng	Atten	Stn	Wher	Stn	Wher	N
1. Common sense	12	28	31	30	34	-3	10	120
2. Flexibility/ motivation	15	39	43	43	46	0	7	120
3. Sonar Op overall	08	24	27	32	36	19	33	115
4. Factor score 3-traits	14	35	38	41	43	8	13	115
5. Make environ. measurements	22	44	48	52	53	8	10	111
6. Set up equipment	17	35	38	38	41	0	8	107
7. Detect contacts	05	17	19	20	26	5	37	108
8. Factor score 3-trt, 3-tsk	14	33	36	38	41	6	14	101
9. Set up for T. arrays	-04	-18	-20	0	37	--	--	16
10. Target detect., broadband	10	25	28	24	29	-14	4	96
11. Target detect., narrow band	-04	-04	-05	0	29	--	--	33
12. Maintenance & repair	12	38	40	--	--	--	--	29
13. PMS	15	39	43	--	--	--	--	61
14. Factor score, 3 traits + PMS	14	40	44	--	--	--	--	60

[a]Decimal points omitted from the Rs.

Criteria 1, 2, 4, 5, 8, and 14, which have Stein estimates > .26 and Wherry estimates > .40, are the most predictable criteria. Three of these criteria, (variables 4, 8, and 14) are global estimates of performance. Of the three, variable 14, the most general because it measures both sonar operation and maintenance abilities, was the most predictable, However, this variable only describes non-technicians. An equivalent variable for technicians would include variables 1, 2, 3, and 12, but the sample size is too small for the calculation to be meaningful. Variables 4 and 8, the 3- and 6-element global estimates of sonar operation performance, are about equally predictable.

The ASVAB variables are most predictive for the traits of common sense and flexibility/ motivation and for the task rating, make environmental measurements. They were poor predictors of ability to set up for towed arrays, and for passive target detection abilities (variables 10 and 11). In the area of sonar operation, the largest increments of GRIP variables for ASVAB variables were for criteria the ASVAB variables already predict well.

Predictiveness of SELC and SELC + GRIP Subsets

Table 10 provides estimates of corrected validity coefficients for SELC and SELC + GRIP. Stein and Wherry estimates are not calculated for Tables 10 or 11 because correction for capitalization on chance was not required for SELC. In contrast, Stein and Wherry corrections are given for the SELC + GRIP battery in order to provide a conservative estimate of the validity increases from GRIP variables. Since both SELC and GRIP variables are fixed for this step, calculation of Stein and Wherry adjustments may actually over correct for selection. The SELC + GRIP estimates would be several points larger if they were calculated the same way as the SELC estimates.

The conservative Stein and Wherry adjustments are used in this report as estimates of predictiveness of the SELC + GRIP and G + GRIP (Table 11) predictor sets. In contrast, the estimates of predictiveness for the SELC and G predictors are the attenuated values and the differences and percentage differences shown in Tables 10 and 11 in columns 4 through 8 relate to the differences between an attenuated value and Stein and Wherry corrections. Thus, the coefficients in columns 4 through 8 underestimate the true increments that GRIP tests add to the ASVAB predictor sets.

The raw coefficients (first column) support the research hypothesis that the selection process for STGs does not measure adequately the perceptual abilities required for target detection. The raw coefficient was negative for one and low positive for the other two target detection criteria (variables 11, 7 and 10).

SELC has negative predictiveness for one other criterion (variable 9). It has positive coefficients for 12 criteria. The tasks for which it has negative correlations were specific to only a small percentage of STG billets (19 and 28%, respectively, of the billets in the sample).

Both Wherry and Stein estimates of the validity coefficients for SELC range from -.22 to .47, and have medians of .37. Eight coefficients are > .30 and five (variables 2, 5, 12, 13, and 14) > .40.

Wherry estimates for the SFLC + GRIP battery range from .28 to .55 and have a median of .37. Stein estimates for these values range from 0 to .53 with median being .32.

GRIP variables provided median increases in validity coefficients of 15 percent based on the Wherry estimates and 5 percent based on the Stein estimates. However, these are underestimates of the true increase in predictiveness from the GRIP variables. To provide a better estimate, Stein and Wherry estimates were calculated for SELC and were subtracted from those shown for SELC + GRIP in Table 10. The median increase for both Wherry and Stein estimates was 18 percent.

Predictiveness of G and G + GRIP Subsets

Corrected validity coefficients for G and G + GRIP variables (Table 11) are similar to but generally smaller than comparable coefficients for SELC. All but one of the Wherry and Stein estimates of the coefficients of G for criteria are within .02 (usually smaller) of the comparable estimates in Table 10 for SELC. Median estimate of the predictiveness of G is .37.

Median increase from GRIP variables is 8 percent based on the Wherry estimates and 0 percent based on Stein estimates. However, if the Wherry and Stein estimates for G were calculated the same way as those for the G + GRIP set, the median increase is 13 percent for the Wherry and 11 percent for the Stein estimates.

Summary of Findings

1. ASVAB variables have substantial predictiveness for STG job performance. Median Wherry and Stein estimates for ASVAB subtests were .41 and .22, respectively. For both SELC and G they were .37. Thus, most estimates show the predictiveness of ASVAB variables for STG job performance to be in the high moderate range (Cohen, 1969).

2. Statistically significant measures of incremental validity were found for four GRIP scores, DG_RIGHTS, PS_HINTS, VG_FALSP and SC_DIFS. DG_RIGHTS, which provided statistically significant incremental validity for 7 of the 10 criteria for which incremental validity was found, is the most promising of the GRIP variables. The number and variety of the criteria for which it was statistically significant indicate that DG_RIGHTS is measuring a mental ability that is used in a wide variety of traits and tasks.

3. The predictiveness of DG_RIGHTS occurred despite the variables's low reliability. Further research should be conducted to improve the reliability of DG_RIGHTS or a similar measure of DSpan. The GRIP measure for DG_RIGHTS was very short. Thus, simply increasing the length of the test may improve its reliability to where it is acceptable.

4. The other measures with significant incremental validity are less promising for Navy personnel classification. PS_HINTS appears to measure a fairly narrow ability and the test requires a computer terminal with a complete alpha keyboard. The test cannot be administered on the computers presently used for Navy classification testing because those computers do not have complete alpha keyboards.

VG_FALSP is based on a small sample size; there is no obvious rationale to explain its statistically significant relationship with set up for towed arrays; and it is collected from a test that requires 26 minutes to administer. This variable is not promising for future research.

Although it is logical that SC_DIFS, a measure of perceptual ability, be significantly predictive for narrow band target detection, the weak effect found indicates that SC_DIFS is unlikely to be useful as a predictor and it is not promising for future research unless its reliability can be substantially improved.

5. Median incremental validity of the GRIP variables for ASVAB subtests was 28 percent by Wherry and 16 percent by Stein estimates. Estimates of the incremental validity of GRIP variables indicate that they provide median increases of about 18 percent for SELC and about 12 percent for G.

6. The ASVAB variables are most predictive for the traits of common sense and flexibility/ motivation and for the task rating, make environmental measurements. They were poor predictors of ability to set up for towed arrays, and for passive target detection abilities. GRIP variables did not increment the validities for variables measuring maintenance and repair skills as opposed to sonar operation skills.

7. The findings of the research do not support one of its major hypotheses, that properly chosen perceptual tests could improve the job performance characteristics of STGs. Only two perceptual scores were significantly predictive (VG_FALSP and SC_DIFS) and neither was promising for future research. Part of the lack of predictiveness may result from the low reliability coefficients of GRIP scores.

CONCLUSIONS

1. The current ASVAB selector composite provides effective selection for STG personnel. It effectively predicts all aspects of STG rating performance except "narrow band target detection" and "ability to set up towed arrays."

2. Although DSpan substantially improves predictiveness for STG job performance, its low reliability makes it inappropriate for operational use in its current form.

RECOMMENDATIONS

Research should be continued with DSpan to (a) improve the reliability of the test and (b) evaluate its predictiveness for job performance of personnel in other ratings.

REFERENCES

Buckner, D. E., & McGrath, J. J. (1961). *A comparison of performance on single and dual sensory mode vigilance tasks* (TR 206-8). Goleta, CA: Human Factors Research, Inc.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, J., & Cohen P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.

Cory, C. H. (1974). *An evaluation of computerized tests as predictors of job performance in three Navy ratings: I. Development of the instruments* (NPRDC-TR-75-2). San Diego: Navy Personnel Research and Development Center.

Cory, C. H. (1976). *An evaluation of computerized tests as predictors of job performance: II. Differential validity for global and job element criteria* (NPRDC-TR-76-28). San Diego: Navy Personnel Research and Development Center.

Dobbins, D. A., Tiedemann, J. G., & Skordahl, D. M. (1961). Field study of vigilance under highway driving conditions (A. P. R. P. Technical Research Note 118).

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

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

Kennedy, E. (1988). Estimation of the squared cross-validity coefficient in the context of best subset regression. *Applied Psychological Measurement, 12*, 231-237.

Mackie, R. R., Ridihalgh, R. R., & Shultz, T. E. (1981). *New criteria for the selection and evaluation of Sonar Technicians* (NPRDC-TR-81-13). San Diego: Navy Personnel Research and Development Center.

Mackie, R. R., McCauley, M. E., & Ridihalgh, R. R. (1975). *Research on new predictors for the selection and evaluation of Sonar Technicians: Phase 1* (NPRDC-TR-75-24). San Diego: Navy Personnel Research and Development Center.

Mackworth, N. H. (1957). Vigilance. *The Advancement of Science, 53*, 389-393.

Schmidt, F. L. Hunter, J. E., & Dunn, W. L. (In review). *Potential utility increases from adding new tests to the Armed Services Vocational Aptitude Battery (ASVAB).* San Diego: Navy Personnel Research and Development Center.

Stein, C. (1960). Multiple regression. In I. Olkin, et al. (Eds.), *Contributions to probability and statistics*. Palo Alto, CA: Stanford University Press.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*, 1-32.

Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.

Wechsler, D. (1981). *Manual for Wechsler adult intelligence scale--revised*. New York: Psychological Corporation.

Wherry, R. J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Annuals of Mathematical Statistics*, 2, 440-457.

DISTRIBUTION LIST

Distribution:
Chief of Naval Research (ONT-222)
Chief of Naval Research (ONT-20)
Office of the Assistant Secretary of Defense (FM&P) (MMPP/AP))
Defense Technical Information Center (DTIC) (2)

Copy to:
Chief of Naval Operations (OP-136L)
Chief of Naval Operations (OP-136F)
Director, Military Personnel Policy Division (OP-13)