(2)

AFHRL-TP-89-30

AIR FORCE

HUMAN RESOURCES LABORATORY

AD-A216 122

ESTIMATING THE CONTRIBUTION OF EXPERIMENTAL TESTS TO THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY

Raymond E. Christal

Universal Energy Systems
4401 Dayton-Xenia Road
Dayton, Ohio 45432-1894

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

December 1989
Interim Technical Paper for Period October 1988 - May 1989

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1989

3. REPORT TYPE AND DATES COVERED
Interim -- October 1988 to May 1989

4. TITLE AND SUBTITLE
Estimating the Contribution of Experimental Tests to the Armed Services Vocational Aptitude Battery

5. FUNDING NUMBERS
C - F41689-86-D-0052
PE - 61102F
PR - 2313
TA - T1
WU - 33

6. AUTHOR(S)
Christal, R.E.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Universal Energy Systems
4401 Dayton-Xenia Road
Dayton, Ohio 45432-1894

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFHRL-TP-89-30

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The only way to accurately determine how much an experimental test will add to the Armed Services Vocational Aptitude Battery (ASVAB) in predicting a subsequent criterion would be to administer the experimental test along with ASVAB in the operational setting. This simply is not feasible to do on a routine basis. Ordinarily, experimental tests are administered to individuals already in the service and an attempt is made to determine how much predictive efficiency these tests add to ASVAB subtests administered some months previously. This approach has been justly criticized because abilities could have changed between the time of ASVAB administration and experimental test administration. Because of the availability of ASVAB test-retest data on over 4,000 subjects, it was possible in the present research effort to address the ability changes in terms of a meta-analysis of differences between test and retest validities for grades in 69 technical school courses. It was hypothesized that the retest scores should have higher validities to the extent that abilities had changed. This follows from the fact that retest scores measure abilities at the time of course entry.

With the exception of slightly higher retest validities for ASVAB measures of technical knowledge, the test and retest validities centered at about the same level. It was concluded that reasonable

14. SUBJECT TERMS
abilities
ability measurement
Armed Services Vocational Aptitude Battery (ASVAB)
Learning Abilities Measurement Program (LAMP)
meta-analysis
tests
validation

15. NUMBER OF PAGES
24

16. PRICE CODE

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

Item 13 (Concluded):

trust can be placed in estimates of the contributions of experimental tests to ASVAB in the prediction of subsequent criteria, even though the ASVAB scores are collected some months prior to administration of the experimental tests.

Another approach to estimating the contribution of experimental tests to ASVAB would be to readminister ASVAB concurrently with the experimental tests. The present research indicates that this approach would, on the average, yield values comparable to those which would be obtained if the experimental tests had been administered at the operational testing stations, along with ASVAB.

Validities of experimental tests for concurrent criterion measures administered in the same testing session appear to be moderately inflated. This inflation is hypothesized to be due in part to a few non-cooperative subjects who operate at a reduced level of effort on all tests when they are told that scores will not affect their careers. It is recommended that, to the extent possible, such non-cooperative subjects be identified and removed from the validation sample. Estimated contributions may still be slightly inflated, but results from the present study suggest that such inflation is not likely to be large.

AFHRL Technical Paper 89-30 December 1989

ESTIMATING THE CONTRIBUTION OF EXPERIMENTAL TESTS TO THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY

Raymond E. Christal

Universal Energy Systems
4401 Dayton-Xenia Road
Dayton, Ohio 45432-1894

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

Reviewed by

Patrick C. Kyllonen
Learning Abilities Research Function

Submitted for publication by

Joseph L. Weeks
Chief, Cognitive Skills Assessment Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

This paper is concerned with how to estimate the contribution of experimental tests to the Armed Services Vocational Aptitude Battery (ASVAB) in accounting for criterion variance. Ordinarily, experimental tests are administered to individuals already in the service and an attempt is made to determine how much predictive efficiency these tests contribute to ASVAB subtests, which were administered some months previously. This approach has been justly criticized, in that ability levels could have changed between the time of ASVAB administration and experimental test administration. In the LAMP program, more serious problems are faced. Often an attempt is made to determine how much power the LAMP tests add to ASVAB when the criterion measure is administered concurrently with the experimental tests. Thus, important questions need to be answered. How much do the abilities of Air Force applicants change between the time of operational testing and entry into service? Would scores from ability tests administered at time of entry into technical schools have higher validities for course grades than would scores collected some months earlier in the operational testing program? How much are estimates of the validity added by LAMP tests to ASVAB inflated because of concurrency effects? It was possible to address these and other questions in the present research because of the availability of ASVAB test-retest data on over 4,000 cases. Investigation of ability changes between test and retest was made by meta-analyses of validities for technical school grades. It was hypothesized that retest scores should have higher validities, since they tap abilities at the time of school entry. Concurrency effects were analyzed by internal analysis of ASVAB test-retest data.

With the exception of slightly higher retest validities for ASVAB measures of technical knowledge, the test and retest validities centered at about the same level. The validities of ASVAB retest scores may have been slightly underestimated because of situational variance during the experimental retest situation; however, this situational variance will lead to underestimates of the validity of any experimental test for subsequent operational criteria. It was concluded that, until better information becomes available, the estimated contribution of experimental tests to ASVAB in the prediction of subsequent criteria be accepted at face value.

Another approach to estimating the contribution of experimental tests to ASVAB would be to readministered ASVAB concurrently with the experimental tests. The present research indicates that this approach would, on the average, yield values comparable to those which would be obtained if the experimental tests had been administered at the operational testing stations with ASVAB.

Validities of experimental tests for concurrent criterion measures (administered in the same testing session) appear to be moderately inflated. This inflation is hypothesized to be due in part to a few non-cooperative subjects who operate at a reduced level of effort on all tests when they are told that scores will have no effect on their careers. It is recommended that, to the extent possible, such non-cooperative subjects be identified and removed from validation samples. Estimated contributions may still be slighly inflated, but results from the present research suggest that such inflation is not likely to be large.

PREFACE

Support for this effort was provided by the Air Force Human Resources Laboratory (AFHRL), under contract F41689-86-D-0052. The data analyzed and reported in this paper were originally collected by Maj. John Welsh and Capt Randy Massey for use in the ASVAB quality control program. I am indebted to Mr. Calvin Fresne and Mrs. Janice Herford for assistance in locating data files and reorganizing them for analysis. Ms Carmen Peña provided assistance in proofing tables and locating references. Appreciation is expressed to Dr. William Alley, Dr. Patrick Kyllonen, Dr. William Tirre, Dr. Scott Chaiken, and Dr. Sidney Irvine for their suggestions and encouragement. I also wish to acknowledge the assistance of Dr. Malcom Ree and Mr. William Phalen for comments concerning analysis alternatives. Finally, I am deeply indebted to Dr. Lloyd Humphreys and Dr. Daniel Woltz, both of whom read early drafts of this paper and made many helpful suggestions leading to its improvement.

TABLE OF CONTENTS

LIST OF TABLES

Estimating the Contribution of Experimental Tests to the Armed Services Vocational Aptitude Battery

I. INTRODUCTION

Advances in cognitive theory and the general availability of microcomputers have stimulated research during the last decade on the development of computer-based tests designed to measure individual differences in cognition. A question currently being addressed is whether these new theory-based tests will contribute anything to conventional paper-and-pencil ability tests in predicting subsequent learning and performance criteria. During the last few years, scientists working in the Air Force Learning Abilities Measurement Program (LAMP)[1] have conducted research involving validation of new experimental tests against concurrent or subsequent learning and performance measures . A normal practice is to obtain official scores from the previously administered Armed Services Vocational Aptitude Battery (ASVAB) and compute how much the experimental tests contribute to that battery in accounting for criterion variance. Results from this procedure relating to the incremental validity of the experimental tests may be inflated. The ASVAB validities may be attenuated because individuals have changed in the time between ASVAB and criterion test administrations. Because the new experimental tests are administered at or near the time of the criterion measures, they are not subject to this attenuation.

Humphreys and others (Humphreys, 1960; Humphreys & Davey, 1988; Humphreys, Davey, & Park, 1985; Humphreys, Parsons, & Park, 1979) have shown that when test scores or school grades are collected on the same individuals at several points in time, the intercorrelations of scores or grades yield a quasi-simplex matrix. That is, high correlations are obtained between successive scores or grades across a short time period; but the longer the time intervening between data collection, the lower the correlations. However, it should be noted that these studies dealt with children or with young adults in college. In both instances significant changes in knowledge across time should have been expected.

It may be that the knowledge changes in Air Force selectees between testing and service entry are far less dramatic than those demonstrated by Humphreys. Over 40% of the individuals tested in LAMP enter the Air Force within 6 months of operational testing, and roughly 87% come in within 12 months of operational testing. During this intervening time, many of these individuals are not in school at all, although some are tested in high school and take additional courses before graduation and entry into the Air Force.

The only way to accurately determine how much an experimental test will contribute to ASVAB in predicting a subsequent criterion would be to ad-

[1] LAMP is a basic research program which is jointly sponsored by the Air Force Office of Scientific Research (AFOSR) and the Air Force Human Resources Laboratory (AFHRL). LAMP seeks to understand how individuals process information in order to perceive, store, remember, solve problems, and acquire knowledge and skills. An ultimate goal is to develop new ability measures that can be used in the Air Force Personnel Selection and Classification Program.

minister it along with the ASVAB in the operational setting. This is simply not feasible to do on a routine basis at the present time.

A second approach would be to readminister the ASVAB along with experimental tests after subjects have entered the Air Force. Then one could find out how much power the experimental tests add to ASVAB in predicting concurrent or subsequent criteria. This would eliminate the time differential between ASVAB and experimental test administrations, but it might not accurately show what the validities and joint validities would have been had these measures been administered in the operational testing stations some months previously. Furthermore, insufficient testing time is available for this approach.

The third approach is the one currently being used in LAMP research; that is, to determine how much the scores on experimental tests, administered at or near the time of criterion data collection, contribute to operational ASVAB scores in accounting for variance in the criterion. This is undoubtedly the weakest approach, but the only one currently feasible. No information is yet available as to how fallible such estimates might be.

Thus, serious questions must be addressed regarding the LAMP validation process. How much do the abilities of Air Force applicants change between the time of operational testing and the time of entry into the service? Would scores from ability tests administered at time of entry have higher validities for technical school grades than would scores collected some months earlier in the operational testing program? How much are estimates of the validity contributed by LAMP tests inflated because of concurrency effects? The present study was designed to provide information bearing on such questions.

II. APPROACH

During the 1984-1987 time period, the Air Force readministered the ASVAB to random samples of airmen on the 6th day of their Basic Military Training. This is the same day that experimental LAMP data are collected. The present investigation consists of analyses of these ASVAB test-retest data in an effort to address the predictive-concurrent validity question.

III. METHOD

ASVAB test-retest scores were available on all subtests for 4,077 cases. The sample was divided into three subsamples: (a) those who entered the Air Force within 6 months after initial testing (N = 1,774), (b) those who entered between 6 and 12 months after testing (N = 1,785) and (c) those who entered 12 or more months after testing (N = 518).

Three types of analyses were conducted. First, test-retest correlations, means, and standard deviations were computed for all 10 ASVAB subtests. These data provide a rough indication of the stability of test scores across time. Second, test and retest data were correlated with technical school grades for a variety of courses. Due to concurrency effects, one would expect the retest scores to have higher validities. Finally, each of eight ASVAB subtests at Time 2 (retest) was selected as a "criterion," and in each instance, a second related test was selected as an "experimental" predictor. It was then possible to compare how much the experimental predictor added to "ASVAB" (the remaining eight subtests) in predicting the criterion under two conditions: (a) The ASVAB was administered months earlier under operational conditions, and (b) the ASVAB was readministered concurrently with the experimental and criterion variables. Each of these analyses is discussed below.

IV. RESULTS

Test-Retest Correlations, Means, and Standard Deviations

Table 1 presents the names of the various subtests in ASVAB, along with the abbreviations which are used throughout this paper. Table 2 reports test-retest correlations for the ASVAB subtests, separately for each of the three subsamples and the total sample. This table also reports median internal consistency reliability estimates for eight of the subtests, as reported by Ree, Mullins, Mathews, & Massey (1982). Table 3 reports test-retest means and standard deviations for the various subsamples and total sample.

Table 1

ASVAB TESTS AND ABBREVIATIONS

TEST	ABBR.
General Science	GS
Arithmetic Reasoning	AR
Word Knowledge	WK
Paragraph Comprehension	PC
Numerical Operations	NO
Coding Speed	CS
Auto-Shop Information	AS
Mathematics Knowledge	MK
Mechanical Comprehension	MC
Electrical Information	EI

Table 2

ASVAB TEST-RETEST CORRELATIONS

TEST	0-6MOS	6-12MOS	>12MOS	TOTAL	r11
N	1174	1785	518	4077	
GS	.744	.736	.737	.740	.84
AR	.768	.762	.773	.766	.90
WK	.774	.765	.798	.772	.92
PC	.493	.407	.350	.439	.80
NO	.690	.672	.620	.674	*
CS	.694	.651	.596	.663	*
AS	.800	.808	.776	.800	.88
MK	.839	.819	.817	.828	.87
MC	.716	.704	.714	.710	.86
EI	.692	.687	.662	.685	.83

Notes: r11 refers to internal consistency reliabilities. * = not computed.

Data in Table 2 reveal some lack of stability in scores across time, especially for the Paragraph Comprehension, Coding Speed, Numerical Operations, and Electrical Information subtests. Part of this is attributable to a lack of internal consistency in the tests themselves. Paragraph Comprehension and Electrical Information have the two lowest reliability coefficients of those considered. Furthermore, Numerical Operations and Coding Speed are both speeded tests, and speeded tests are noted for their lack of stability across time. The test-retest correlations are attenuated somewhat by the use of several test forms. However, great pains have been taken to produce ASVAB forms which are equivalent in terms of item characteristics, and such attenuation is judged to be of minor consequence.

Data for the three time periods reported in Table 2 do not form a simplex. Although the correlations for the 6-12 months group are slightly lower that those for the 0-6 months group, several of the correlations for the over 12 months group are actually higher than those for the 0-6 months group. The most striking feature in Table 2 is the consistency of values across time. Of course, these data do not represent correlations on the same individuals across time. They are from independent subsamples. The small differences noted could be due entirely to sampling fluctuations.

Table 3 reports test-retest means and standard deviations for subtests, separately for each of the subsamples and the total sample. The test-retest means show a remarkable stability across time. Although some minor fluctuations can be observed from subtest to subtest, it appears that, on the whole, forgetting, learning, practice effects, regression effects and situational variance at Time 2 are fairly well balanced out. The standard deviations appear to be generally higher at Time 2, a matter which will be discussed later in this paper.

Although the data in Tables 2 and 3 suggest little difference between test and retest scores regardless of interval times, firm conclusions concerning

TABLE 3

ASVAB TEST-RETEST MEANS AND STANDARD DEVIATIONS

MEANS

	0-6 MONTHS (N=1774)		6-12 MONTHS (N=1785)		>12 MONTHS (N=518)		TOTAL SAMPLE (N=4077)	
TEST	TIME1	TIME2	TIME1	TIME2	TIME1	TIME2	TIME1	TIME2
GS	18.7	18.4	18.6	18.6	18.9	19.1	18.7	18.6
AR	22.9	22.1	22.8	21.9	23.3	22.6	22.9	22.1
WK	29.2	29.4	28.8	29.5	29.2	29.9	29.1	29.5
PC	12.5	12.4	12.4	12.3	12.4	12.5	12.4	12.4
NO	40.8	40.6	40.9	40.3	41.6	41.1	40.9	40.5
CS	53.3	54.3	52.3	53.4	53.0	53.0	52.8	53.7
AS	18.4	18.7	18.2	18.5	17.9	18.7	18.3	18.6
MK	16.6	16.2	16.9	16.3	17.2	16.8	16.8	16.3
MC	17.9	18.0	17.6	17.8	17.7	18.3	17.7	18.0
EI	14.1	13.9	13.9	13.9	13.8	14.0	14.0	13.9

STANDARD DEVIATIONS

	0-6 MONTHS (N=1774)		6-12 MONTHS (N=1785)		>12 MONTHS (N=518)		TOTAL SAMPLE (N=4077)	
TEST	TIME1	TIME2	TIME1	TIME2	TIME1	TIME2	TIME1	TIME2
GS	3.66	3.77	3.53	3.68	3.29	3.56	3.56	3.71
AR	4.64	5.35	4.59	5.17	4.52	4.93	4.60	5.22
WK	4.37	4.29	4.34	4.11	4.10	3.90	4.33	4.16
PC	1.93	1.99	1.92	1.92	1.78	1.92	1.91	1.95
NO	7.24	7.67	7.01	7.58	7.01	7.33	7.12	7.59
CS	11.94	12.30	11.41	12.23	11.60	11.48	11.67	12.18
AS	4.39	4.27	4.33	4.41	4.41	4.36	4.37	4.34
MK	5.04	5.23	4.74	5.13	4.68	5.02	4.87	5.16
MC	3.98	3.98	4.08	3.98	4.00	3.89	4.03	3.97
EI	3.15	3.37	3.12	3.36	3.19	3.44	3.14	3.38

ability changes across time cannot be drawn from comparative analyses of data from three different subsamples. Therefore, emphasis was shifted to the second approach, which involved validation of Time 1 (test) and Time 2 (retest) scores against technical school course grades.

Test-Retest Validation Against Technical School Grades

The ASVAB subtest test-retest correlations are low enough to allow for significant changes to have occurred in the abilities of applicants between the time of operational testing and the individuals' entry into the Air Force. To the extent that such changes did occur, one would expect the retest scores to have higher validities for technical school grades. This follows from the fact that the scores gathered just prior to entry into the technical school should be the best indicators of abilities at that time. Two approaches were taken to evaluate this hypothesis. One involved using a simple signs test of the differences in validity coefficients yielded by the test and retest data. The second involved more sophisticated meta-analyses.

Signs Tests. ASVAB scores for the 4,077 subjects included in the study were matched with operational files to obtain technical school grades. Table 4 reports test and retest validities of the 10 ASVAB subtests for grades in each of the 20 courses in which the matching process yielded 40 or more subjects. Thus, 200 pairs of validity coefficients were available to test the hypothesis. Because test and retest scores came from the same individuals, 200 t-values were computed using Hotelling's formula for correlated data, one for each test-retest validity pair. These t-values are reported in Table 5, ordered by ASVAB subtest. A negative t-value indicates that a particular retest validity coefficient is higher than its associated test validity coefficient.

A cursory inspection of the data in Table 5 reveals a weak tendency for the retest validities to be higher. There are 110 negative t-values and 90 positive ones. Nevertheless, applying a simple signs test produces a Chi-Square of only -2.225, which is non-significant at the .05 level. The overall hypothesis must be rejected. It appears that, in general, retest validities are *not* significantly higher than test validities. Although test-retest validities did not differ across all subtests, it was hypothesized that ability changes were more likely to occur among the ASVAB subtests measuring technical knowledge, than among those measuring general ability or perceptual speed. Certainly technical knowledge is subject to learning and forgetting across time. Therefore, the ASVAB subtests were sorted into three categories for further evaluation, as follows:

TECHNICAL KNOWLEDGE
Electrical Information
Mechanical Comprehension
General Science
Auto-Shop Information

GENERAL ABILITY
Arithmetic Reasoning
Word Knowledge
Mathematics Knowledge
Paragraph Comprehension

PERCEPTUAL SPEED
Numerical Operations
Coding Speed

Data in Table 5 reveal that the four technical knowledge subtests yielded 52 negative t-values but only 28 positive t-values. In this instance, the simple signs test produces a Chi-Square of -7.212, which is significant beyond the .01 level. Thus, the hypothesis of higher retest validities for ASVAB measures of technical knowledge is supported. The signs tests for differences in the general ability and perceptual speed ASVAB measures produced Chi-Squares of 0.812 and -0.366, respectively, neither of which is significant at the .05 level.

Meta-Analyses. As a second, and much more powerful approach, a series of meta-analyses were conducted for the three categories of tests mentioned above (technical knowledge, general ability, and perceptual speed), and for all subtests combined. Results of these analyses are presented

Table 4

TEST-RETEST CORRELATIONS WITH SCHOOL GRADES

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
81130 (APPRN. SECURITY SPECIALIST)											
TIME1	268	0.507	0.286	0.459	0.386	0.110	0.176	0.273	0.364	0.182	0.284
TIME2	268	0.482	0.392	0.454	0.336	0.170	0.174	0.281	0.346	0.130	0.353
43131 (APPRN. TACTICAL AIRCRAFT MAINTENANCE SPECIALIST)											
TIME1	223	0.435	0.416	0.378	0.364	0.156	0.112	0.316	0.403	0.412	0.382
TIME2	223	0.394	0.416	0.315	0.316	0.106	0.129	0.389	0.419	0.358	0.479
81150 (SECURITY SPECIALIST)											
TIME1	182	0.448	0.271	0.345	0.266	0.223	0.088	0.255	0.334	0.326	0.149
TIME2	182	0.468	0.336	0.337	0.271	0.274	0.169	0.268	0.324	0.275	0.257
70230B (APPRN. ADMINISTRATION SPECIALIST-STAFF SUPPORT)											
TIME1	118	0.354	0.427	0.336	0.275	-0.024	0.150	0.257	0.367	0.392	0.332
TIME2	118	0.369	0.433	0.215	0.439	0.091	0.142	0.258	0.481	0.267	0.251
64530 (APPRN. INVENTORY MANAGEMENT SPECIALIST)											
TIME1	114	0.264	0.451	0.329	0.205	0.037	0.016	0.195	0.273	0.059	0.078
TIME2	114	0.250	0.345	0.274	0.181	0.084	-0.066	0.202	0.323	0.186	0.171
81132 (APPRN. LAW ENFORCEMENT SPECIALIST)											
TIME1	110	0.447	0.328	0.444	0.474	0.126	0.204	0.325	0.383	0.372	0.310
TIME2	110	0.542	0.406	0.401	0.362	0.209	0.167	0.418	0.346	0.340	0.455
57130 (APPRN. FIRE PROTECTION SPECIALIST)											
TIME1	89	0.293	0.302	0.253	0.294	0.085	0.147	0.221	0.343	0.323	0.279
TIME2	89	0.417	0.3867	0.243	0.167	0.134	0.063	0.226	0.233	0.451	0.333
90230 (APPRN. MEDICAL SERVICES SPECIALIST)											
TIME1	86	0.482	0.350	0.379	0.310	0.020	-0.076	0.399	0.323	0.206	0.293
TIME2	86	0.473	0.357	0.462	0.286	-0.077	0.018	0.508	0.228	0.330	0.465
63130 (APPRN. FUEL SPECIALIST)											
TIME1	84	0.150	0.432	0.075	0.214	0.320	0.230	0.221	0.446	0.152	0.096
TIME2	84	0.071	0.401	0.093	0.240	0.242	0.269	0.083	0.414	0.173	0.101
73230 (APPRN. PERSONNEL SPECIALIST)											
TIME1	74	0.354	0.404	0.513	0.334	0.078	-0.054	0.142	0.409	0.142	0.093
TIME2	74	0.241	0.475	0.441	0.379	0.112	0.151	0.123	0.376	0.231	0.267

Table 4 (Cont.)

TEST-RETEST CORRELATIONS WITH SCHOOL GRADES

42632 (APPRN. JET ENGINE MECHANIC)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	73	0.281	0.200	0.201	0.327	-0.021	0.318	0.329	0.253	0.416	0.307
TIME2	73	0.348	0.191	0.220	0.217	-0.027	0.240	0.269	0.256	0.301	0.222

43132 (APPRN. STRATEGIC AIRCRAFT MAINTENANCE SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	69	0.449	0.457	0.402	0.450	0.251	0.293	0.264	0.418	0.306	0.460
TIME2	69	0.569	0.417	0.396	0.485	0.348	0.124	0.291	0.453	0.371	0.392

42335 (APPRN. AEROSPACE GROUND EQUIPMENT MECHANIC)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	71	0.090	0.446	-0.035	0.061	0.375	0.026	0.263	0.268	0.232	0.207
TIME2	71	0.182	0.421	0.055	0.155	0.439	0.206	0.255	0.443	0.253	0.330

46130 (APPRN. MUNITIONS SYSTEMS SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	55	0.392	0.422	0.351	0.343	0.458	0.175	0.447	0.366	0.271	0.386
TIME2	55	0.623	0.398	0.166	0.108	0.257	0.094	0.479	0.421	0.427	0.586

64531 (APPRN. MATERIEL STORAGE AND DISTRIBUTION SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	52	0.261	0.254	0.383	0.349	0.211	0.372	0.087	0.161	0.149	0.034
TIME2	52	0.290	0.146	0.371	0.473	0.113	0.196	-0.046	0.204	0.110	0.208

12230 (APPRN. AIRCREW LIFE SUPPORT SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	52	0.122	0.288	0.158	0.152	0.004	0.211	0.422	0.255	0.237	0.214
TIME2	52	0.278	0.175	0.251	0.264	0.090	0.235	0.259	0.262	0.281	0.270

43133 (APPRN. AIRLIFT AIRCRAFT MAINTENANCE SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	45	0.402	0.347	0.315	0.296	0.183	0.212	0.197	0.366	0.349	0.377
TIKME2	45	0.318	0.321	0.288	0.421	0.148	0.236	0.307	0.460	0.480	0.320

29130 (APPRN. TELECOMMUNICATIONS OPERATIONS SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	43	0.382	0.314	0.402	0.144	0.230	0.205	0.384	0.320	0.418	0.419
TIME2	43	0.519	0.174	0.586	0.403	0.288	0.189	0.342	0.330	0.432	0.370

42330 (APPRN. AIRCRAFT ELECTRICAL SYSTEMS SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	43	0.075	0.104	-0.070	0.328	-0.262	-0.023	0.094	0.233	0.353	0.101
TIME2	43	0.251	0.212	-0.028	0.162	-0.112	0.037	0.186	0.170	0.141	0.137

60531 (APPRN. AIR CARGO SPECIALIST)

TEST	N	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
TIME1	40	0.182	0.347	0.202	0.210	-0.043	0.096	-0.033	0.269	0.060	0.114
TIME2	40	0.295	0.334	0.273	0.159	-0.040	0.105	0.135	0.254	0.136	0.114

Table 5

DIFFERENCES IN TEST VALIDITIES FOR COURSES, ORDERED BY TESTS AND T-VALUES

COURSE	TST	T	COURSE	TST	T	COURSE	TST	T	COURSE	TST	T
81132	AR	-1.308	81130	CS	0.038	90230	MC	-1.732	73230	MO	-0.313
81150	AR	-1.200	70230	CS	0.090	64530	MC	-1.704	81130	MO	-0.190
57130	AR	-1.082	46130	CS	0.094	57130	MC	-1.673	60531	MO	0.046
73230	AR	-1.046	29130	CS	0.112	46130	MC	-1.246	43133	MO	0.236
42330	AR	-0.938	81132	CS	0.502	73230	MC	-0.973	12230	MO	0.639
70230	AR	-0.105	42632	CS	0.850	43133	MC	-0.881	64531	MO	0.807
90230	AR	-0.082	64530	CS	0.896	43132	MC	-0.712	43131	MO	0.905
43131	AR	0.000	57130	CS	1.002	12230	MC	-0.497	63130	MO	0.977
60531	AR	0.097	43132	CS	1.591	60531	MC	-0.475	90230	MO	1.147
42632	AR	0.119	64531	CS	1.823	42335	MC	-0.234	46130	MO	2.557
46130	AR	0.223				63130	MC	-0.204			
43133	AR	0.232	64531	EI	-2.054	29130	MC	-0.151	29130	PC	-1.702
42335	AR	0.320	46130	EI	-2.008	64531	MC	0.351	70230	PC	-1.701
63130	AR	0.485	90230	EI	-1.972	81132	MC	0.378	81130	PC	-1.190
43132	AR	0.485	43131	EI	-1.946	81150	MC	0.921	64531	PC	-1.091
81130	AR	1.051	73230	EI	-1.849	43131	MC	0.973	42335	PC	-0.801
64531	AR	1.110	81150	EI	-1.824	81130	MC	1.086	43133	PC	-0.768
12230	AR	1.184	81132	EI	-1.675	42632	MC	1.361	12230	PC	-0.752
29130	AR	1.371	81130	EI	-1.497	42330	MC	1.489	73230	PC	-0.396
64530	AR	1.874	42335	EI	-1.305	70230	MC	1.722	43132	PC	-0.287
			64530	EI	-1.218				63130	PC	-0.231
81132	AS	-1.887	63130	EI	-1.097	42335	MK	-2.421	81150	PC	-0.068
90230	AS	-1.823	57130	EI	-0.628	70230	MK	-1.928	90230	PC	0.194
60531	AS	-1.787	12230	EI	-0.531	43133	MK	-0.898	64530	PC	0.217
70230	AS	-1.644	42330	EI	-0.271	64530	MK	-0.865	60531	PC	0.321
43131	AS	-1.579	60531	EI	0.000	46130	MK	-0.621	43131	PC	0.740
43133	AS	-0.942	43133	EI	0.384	64531	MK	-0.558	42632	PC	1.009
43132	AS	-0.376	29130	EI	0.417	43131	MK	-0.545	42330	PC	1.143
46130	AS	-0.356	43132	EI	0.795	43132	MK	-0.500	57130	PC	1.158
81150	AS	-0.260	42632	EI	0.810	29130	MK	-0.135	81132	PC	1.304
64530	AS	-0.121	70230	EI	1.131	12230	MK	-0.082	46130	PC	1.609
81130	AS	-0.094				42632	MK	-0.037			
57130	AS	-0.075	81130	GS	-2.717	60531	MK	0.120	29130	WK	-2.436
42330	AS	0.094	46130	GS	-2.619	81150	MK	0.271	12230	WK	-1.128
42335	AS	0.095	12230	GS	-1.699	63130	MK	0.447	90230	WK	-1.125
73230	AS	0.279	81132	GS	-1.630	81130	MK	0.511	42335	WK	-0.869
29130	AS	0.415	57130	GS	-1.628	73230	MK	0.541	60531	WK	-0.835
42632	AS	0.769	43132	GS	-1.500	42330	MK	0.652	42632	WK	-0.238
63130	AS	1.491	29130	GS	-1.390	81132	MK	0.762	42330	WK	-0.213
64531	AS	1.582	42330	GS	-1.245	90230	MK	1.486	43132	WK	0.082
12230	AS	2.091	60531	GS	-1.183	57130	MK	2.059	63130	WK	0.094
			42335	GS	-0.994				64531	WK	0.112
42335	CS	-1.927	42632	GS	-0.831	81132	MO	-1.305	57130	WK	0.125
73230	CS	-1.728	81150	GS	-0.416	70230	MO	-1.235	81150	WK	0.161
81150	CS	-1.231	70230	GS	-0.347	42330	MO	-1.147	43133	WK	0.251
90230	CS	-1.028	64531	GS	-0.294	43132	MO	-0.970	81130	WK	0.547
42330	CS	-0.470	90230	GS	0.130	81150	MO	-0.894	81132	WK	0.863
63130	CS	-0.442	64530	GS	0.230	42335	MO	-0.638	73230	WK	0.955
43131	CS	-0.291	43133	GS	0.663	64530	MO	-0.605	64530	WK	0.993
12230	CS	-0.220	43131	GS	0.918	57130	MO	-0.598	43131	WK	1.519
43133	CS	-0.172	63130	GS	0.977	42632	MO	-0.550	46130	WK	1.608
60531	CS	-0.069	73230	GS	1.200	29130	MO	-0.546	70230	WK	1.884

Table 6

META-ANALYSES RESULTS

20 SCHOOLS

TEST GROUP	T1	T2	R12	DF	T
Tech. Knowledge	0.286	0.322	0.737	7321	-4.499
General Ability	0.336	0.335	0.727	7321	0.124
Perceptual Speed	0.141	0.148	0.668	3659	-0.526
Total	0.280	0.294	0.720	18307	-2.662

69 SCHOOLS

TEST GROUP	T1	T2	R12	DF	T
Tech. Knowledge	0.295	0.321	0.737	11337	-4.047
General Ability	0.315	0.315	0.727	11337	0.000
Perceptual Speed	0.124	0.136	0.668	5667	-1.120
Total	0.270	0.283	0.720	28347	-3.065

Note: Values under T1 and T2 are average validity coefficients computed using Fisher's z-transformation and weighted by N's. The R12 column is the average time 1-time 2 test intercorrelations, again computed using Fisher's z's and weighted by N's. The last column reports Hotelling's t's based on correlated data for the differences in test and retest average validities within each category.

in Table 6, not only for data in the 20 schools reported in Table 5, but for all schools in which the matching process yielded 15 or more subjects. The general approach was to treat data from each school as being from a separate study. The data were combined using a method recommended by Hedges and Olkin (1985, pp. 230-232). First, the test and retest validities within each category were converted to Fisher's z's and weighted by N-3. Next, the weighted z's were averaged, and the averages were converted back into correlations. Finally, Hotelling's t's were computed for the differences in test and retest average correlations within each subtest category.

The data in Table 6 reveal that there are no significant differences between Time 1 (test) and Time 2 (retest) validities for the ASVAB subtests associated with either general learning ability or perceptual speed. Subtest validities for the technical knowledge area are significantly higher at Time 2, both in the 20-school and 69-school samples, but the magnitudes of the differences are very small. The average retest validities across all schools in both the 20-school and 69-school samples are significantly higher at Time 2, but this was in the presence of many degrees of freedom, and due solely to differences in the technical knowledge subtests.[2]

In summary, analyses of the zero-order validities present a clear indication that ability changes for Air Force applicants between time of testing and entry into the service are restricted to measures of technical knowledge, and are relatively minor.

[2] Some observers may object to the meta-analysis approach because subjects attending the various schools cannot be assumed to have been drawn from the same population. This is certainly true, since the entry ability level requirements vary from course to course. In order to respond to this possible criticism, all validity values in the 20-school sample were corrected for restriction in range due to selection, and the average validities were recomputed. The revised Time 1-Time2 average validities were: technical knowledge, .405 vs. .425.; general ability, .501 vs. .485; perceptual speed, .197 vs .215, and total, .408 vs. .411. These differences appear to be less than those computed for the uncorrected validities. Hotelling t's were not computed because of problems in dealing with double curtailments associated with the intercorrelation terms.

Regression Analyses. Another way of evaluating the impact of ability changes is to determine the differences in these multiple correlations against technical school grades which are obtained using Time 1 and Time 2 data. This was accomplished for each of the 20 technical schools courses in which the matching process yielded 40 or more subjects. Again, the mean multiple R's were based upon weighted z's. The results yielded a multiple R of .546 for Time 1 (test) and .539 for Time 2 (retest). Thus, even considering the joint action of all ASVAB subtests, there is no indication of appreciable loss of predictive power across time.

Internal Regression Analyses Using ASVAB Test-Retest Data

It appears from the preceding analyses that, if ability changes occur for Air Force selectees between time of operational testing and time of entry into the service, the changes are associated with technical knowledge and are relatively minor. Estimates of the value of new tests in adding power to the operational ASVAB in predicting technical school grades would appear to be fairly trustworthy. However, on many occasions LAMP experimental tests are validated against intermediate learning criteria which are administered in the same experimental testing session. It is entirely possible that the correlations between the experimental tests and criterion measures are inflated because of test covariance associated with motivational decrements present in an experimental testing situation. It has been conjectured that some individuals operate with decreased motivation when they are told that the test results will be used only for purposes of research, and this negatively impacts their performance on both experimental and criterion measures administered during an experimental testing session. This type of behavior would be expected to increase the correlations between the experimental and criterion measures. In the case of LAMP validation studies, where the contribution of experimental tests to ASVAB is being evaluated against a concurrent criterion measure, it is important to know how much of the computed contribution might be associated with this type of situational variance.

Data indicating existence of the problem are shown in Table 7. The last three columns of this table report the average correlation (computed using Fisher's z-transformations) of each ASVAB

TABLE 7

TIME1,TIME2 MEANS AND AVERAGE INTERCORRELATIONS FOR TIME1,TIME2 AND TIME1-TIME2
TOTAL SAMPLE (N=4077)

	MEAN SCORES		STD.DEVIATIONS		AVG.INTERCORRELATIONS		
TEST	TIME1	TIME2	TIME1	TIME2	T1-T1	T1-T2	T2-T2
GS	18.66	18.59	3.56	3.71	0.375	0.377	0.393
AR	22.89	22.08	4.60	5.22	0.364	0.377	0.401
WK	29.05	29.50	4.33	4.16	0.322	0.324	0.357
PC	12.44	12.37	1.91	1.95	0.285	0.278	0.314
NO	40.91	40.52	7.12	7.59	0.131	0.144	0.193
CS	52.83	53.70	11.67	12.18	0.133	0.130	0.159
AS	18.25	18.62	4.37	4.34	0.241	0.232	0.267
MK	16.81	16.31	4.87	5.16	0.368	0.365	0.392
MC	17.74	17.96	4.03	3.97	0.349	0.343	0.350
EI	14.00	13.94	3.14	3.38	0.320	0.300	0.359
AVG.	24.36	24.36	4.96	5.17	0.291	0.290	0.320

subtest with the other nine subtests, separately for Time 1, for Time 2, and between Time 1 and Time 2. Notice that the Time 1-Time 2 correlations are almost equivalent to the Time 1-Time 1 correlations, providing additional evidence on the lack of ability changes across time. However, notice that the Time 2-Time 2 correlations are systematically higher than either the Time 1-Time 1 or the Time 1-Time 2 correlations. This may be attributable to common situational variance associated with the experimental testing situation at Time 2. That is, some individuals were either tired, under stress, or unmotivated during the Time 2 experimental retesting session, and they systematically made low scores on all subtests. One would expect this factor to result in lower means, but comparing the means from Time 1 to those from Time 2 is meaningless because of the interacting factors of learning, forgetting, practice, and regression effects mentioned previously.

The standard deviations appear to be generally higher at Time 2, and this too is attributed to situational variance present in the experimental testing session. Poorly motivated subjects could actually score so low that they would not have been accepted into the Air Force had they performed at the same level during the operational administration.

The fact that situational variance in an experimental testing session increases test intercorrelations has implications for the LAMP program. It indicates that the correlations observed between experimental tests and criterion tests administered in the same experimental testing session are likely to be inflated. Furthermore, the computed contribution of the experimental tests to ASVAB scores collected some months previously would also be inflated. The question is how seriously such estimates are inflated.

The ASVAB test-retest data were used to roughly evaluate the amount of inflation likely in estimating the contribution of experimental tests to ASVAB. The approach was to use each of eight ASVAB subtest scores at Time 2 as a criterion measure. In each instance, a related subtest was selected as being an experimental predictor variable, while the remaining eight ASVAB subtest scores at Time 1 and again at Time 2 were treated as though they were the complete ASVAB. Three comparisons were evaluated for each of the eight criterion measures as defined by the experimental design shown in Table 8.

Table 9 reports the R-squares for each of the six comparison equations, separately for eight criteria. The NO and CS subtests were not used as criteria in this analysis because of their unreliability, and because they tend to correlate only with each other.

Table 10 summarizes differences between Comparisons 1, 2, and 3 from Table 9. Comparison 1 is the estimated contribution of an experimental test to ASVAB, when ASVAB is administered at Time 1, and data for the experimental test and criterion data are collected at Time 2 in the same experimental testing session. Comparison 2 reports the actual contribution of the experimental test to ASVAB. In this latter instance, the experimental test and ASVAB data are collected at Time 1, and the criterion data are collected at Time 2. The difference between Comparison 1 and Comparison 2 estimates reflects the degree of inflation in the estimated contribution of an experimental test to ASVAB, when both the experimental test and criterion data are collected concurrently in the same experimental testing session. It appears that such inflation is not very serious. As shown in Table 10, the average estimated variance contribution across all eight criteria in the total sample is 12.8%, while the actual contribution is 10.4%. Application of an 18.75% discount would bring the average estimated contribution into alignment.

The data on Comparison 3 presented in Table 10 suggest that if ASVAB were readministered along with the experimental tests and criterion measures, the estimated contribution of the experimental tests to ASVAB would generally not be inflated. If sufficient testing time were available, this might be the preferred approach.

Table 8

EXPERIMENTAL DESIGN

Y= CRITERION (ONE OF EIGHT ASVAB TESTS).

X= PREDICTOR TEST BEING EVALUATED (A RELATED ASVAB TEST).

ASVAB= REMAINING 8 TESTS IN ASVAB

TIME1= TIME OF OPERATIONAL ADMINISTRATION OF ASVAB

TIME2= 6TH DAY OF BASIC TRAINING. SAME DAY AS LAMP TESTS ARE ADMINISTERED.

COMPARISON 1

R-SQUARED1 Y(TIME2) AS CRITERION
X(TIME2) AND ASVAB(TIME1) AS PREDICTORS

R-SQUARED2 Y(TIME2) AS CRITERION
ASVAB(TIME1) AS PREDICTORS

(R-SQUARED1) - (R-SQUARED2)= ESTIMATED UNIQUE CONTRIBUTION OF EXPERIMENTAL TEST TO ASVAB IN ACCOUNTING FOR CRITERION VARIANCE. THIS IS THE WAY THE AIR FORCE AND OTHER SERVICES ESTIMATE THE VALUE OF NEW TESTS. THIS ESTIMATE IS KNOWN TO BE BIASED IN AN UPWARD DIRECTION, BUT HOW MUCH IS UNKNOWN.

COMPARISON 2

R-SQUARED3 Y(TIME2) AS CRITERION
X(TIME1) AND ASVAB(TIME1) AS PREDICTORS

R-SQUARED4 Y(TIME2) AS CRITERION
ASVAB(TIME1) AS PREDICTORS

(R-SQUARED3) - (R-SQUARED4)= ACTUAL UNIQUE VARIANCE CONTRIBUTED BY EXPERIMENTAL TEST TO ASVAB IN ACCOUNTING FOR CRITERION VARIANCE, WHEN THE EXPERIMENTAL TEST IS ADMINISTERED ON THE SAME OCCASION AS THE OPERATIONAL ASVAB.

COMPARISON 3

R-SQUARED5 Y(TIME2) AS CRITERION
X(TIME2) AND ASVAB(TIME2) AS PREDICTORS

R-SQUARED6 Y(TIME2) AS CRITERION
ASVAB(TIME2) AS PREDICTORS

(R-SQUARED5) - (R-SQUARED6)= UNIQUE VARIANCE CONTRIBUTED BY THE EXPERIMENTAL TEST TO ASVAB IN ACCOUNTING FOR THE CRITERION VARIANCE, WHEN ASVAB IS READMINISTERED CONCURRENTLY WITH THE CRITERION AND EXPERIMENTAL TESTS.

Table 9

SQUARED MULTIPLE CORRELATIONS FOR COMPARISON EQUATIONS

TESTS	TOTAL SAMPLE (N=4077) EQU1 R-SQ	EQU2&4 R-SQ	EQU3 R-SQ	EQU5 R-SQ	EQU6 R-SQ	0-6MOS SAMPLE (N=1774) EQU1 R-SQ	EQU2&4 R-SQ	EQU3 R-SQ	EQU5 R-SQ	EQU6 R-SQ
C - P										
GS-WK	0.511	0.384	0.503	0.536	0.442	0.511	0.398	0.510	0.531	0.448
AR-MK	0.527	0.342	0.487	0.550	0.390	0.549	0.362	0.498	0.573	0.416
WK-GS	0.464	0.326	0.431	0.486	0.383	0.477	0.357	0.456	0.504	0.417
PC-WK	0.318	0.199	0.275	0.342	0.260	0.355	0.237	0.310	0.376	0.283
AS-MC	0.468	0.382	0.445	0.486	0.418	0.463	0.373	0.436	0.482	0.417
MK-AR	0.556	0.383	0.548	0.550	0.398	0.573	0.394	0.557	0.569	0.415
MC-AS	0.453	0.364	0.426	0.471	0.401	0.448	0.357	0.426	0.471	0.404
EI-AS	0.528	0.422	0.517	0.531	0.431	0.518	0.412	0.516	0.526	0.430

TESTS	6-12MOS SAMPLE (N=1785) EQU1 R-SQ	EQU2&4 R-SQ	EQU3 R-SQ	EQU5 R-SQ	EQU6 R-SQ	OVER 12MOS SAMPLE (N=518) EQU1 R-SQ	EQU2&4 R-SQ	EQU3 R-SQ	EQU5 R-SQ	EQU6 R-SQ
C - P										
GS-WK	0.513	0.376	0.506	0.535	0.430	0.510	0.381	0.496	0.564	0.467
AR-MK	0.516	0.340	0.481	0.540	0.379	0.479	0.276	0.480	0.503	0.329
WK-GS	0.451	0.301	0.410	0.474	0.356	0.468	0.326	0.436	0.465	0.357
PC-WK	0.279	0.158	0.245	0.301	0.224	0.324	0.221	0.277	0.384	0.329
AS-MC	0.478	0.394	0.457	0.492	0.421	0.463	0.386	0.438	0.495	0.434
MK-AR	0.551	0.385	0.537	0.542	0.398	0.525	0.322	0.548	0.506	0.331
MC-AS	0.452	0.364	0.423	0.474	0.401	0.472	0.392	0.442	0.474	0.416
EI-AS	0.528	0.426	0.509	0.526	0.417	0.573	0.470	0.564	0.578	0.486

NOTES: Equations are numbered as they are described in the experimental design.

Table 10

ESTIMATED CONTRIBUTION OF EXPERIMENTAL TESTS TO ASVAB IN THE PREDICTION OF SELECTED CRITERION MEASURES

		TOTAL SAMPLE (N=4,077)			0-6MOS SAMPLE (N=1,774)		
		COMP. 1	COMP. 2	COMP. 3	COMP. 1	COMP. 2	COMP.3
CRI	PRE	EQ1-EQ2	EQ3-EQ2	EQ5-EQ4	EQ1-EQ2	EQ3-EQ2	EQ5-EQ4
GS	WK	12.7	11.9	9.4	11.3	11.2	8.3
AR	MK	18.5	14.5	16.0	18.7	13.6	15.7
WK	GS	13.8	10.5	10.3	12.0	9.9	8.7
PC	WK	11.9	7.6	8.2	11.8	7.3	9.3
AS	MC	8.6	6.3	6.8	9.0	6.3	6.5
MK	AR	17.3	16.5	15.2	17.9	16.5	15.4
MC	AS	8.9	6.2	7.0	9.1	6.9	6.7
EI	AS	10.6	9.5	10.0	10.6	10.4	9.6
	MEAN	12.8	10.4	10.4	12.6	10.3	10.0

		6-12MOS SAMPLE (N=1,785)			OVER 12MOS SAMPLE (N=518)		
		COMP. 1	COMP. 2	COMP. 3	COMP. 1	COMP. 2	COMP. 3
CRI	PRE	EQ1-EQ2	EQ3-EQ2	EQ5-EQ4	EQ1-EQ2	EQ3-EQ2	EQ5-EQ4
GS	WK	13.7	13.0	10.5	12.9	11.5	9.7
AR	MK	17.6	14.1	16.1	20.3	20.4	17.4
WK	GS	15.0	10.9	11.8	14.2	11.0	10.8
PC	WK	12.1	8.7	7.7	10.3	5.0	5.5
AS	MC	8.4	6.3	7.1	7.7	5.2	6.1
MK	AR	16.6	15.2	14.4	20.3	22.6	17.5
MC	AS	8.8	5.9	7.3	8.0	5.0	5.8
EI	AS	10.2	8.3	10.9	10.3	9.4	9.2
	MEAN	12.8	10.3	10.7	13.0	11.3	10.3

NOTES: Reported values are differences in R-squares. Equations are numbered as they are described in the experimental design.

COMP. 1= CRITERION & EXPERIMENTAL TEST AT TIME 2, ASVAB AT TIME 1
Criterion & experimental test at Time 2, ASVAB at Time 1.

COMP. 2= CRITERION AT TIME 2, EXPERIMENTAL TEST AND ASVAB AT TIME 1
Criterion at Time 2, experimental test and ASVAB at Time 1.

COMP. 3= CRITERION, EXPERIMENTAL TEST, & ASVAB AT TIME2
Criterion, experimental test & ASVAB at Time 2.

V. DISCUSSION

The basic problem addressed in this paper is how to estimate the contribution of an experimental test to ASVAB in predicting a criterion measure. The only sure method would be to administer the experimental test along with the ASVAB at official testing stations under operational conditions. Then one could estimate the contribution of the experimental test by evaluating the difference in R-squares yielded by a full model (containing the ASVAB subtests and the experimental variable as predictors) and a restricted model (using only ASVAB subtests as predictors). Because a large amount of additional testing time in the operational environment would be required by this approach, it is usually not feasible.

In the typical situation, experimental tests are administered to individuals already in the service and an attempt is made to see how much predictive efficiency is added by these experimental tests to ASVAB subtests administered some months previously. This approach has been criticized, in that ability levels could have changed between the time of ASVAB testing and experimental test administration. The present study is an attempt to evaluate the changes in the abilities of Air Force personnel between the time of their selection and their entry into service.

With the exception of slightly higher retest validities for ASVAB measures of technical knowledge, the test and retest validities centered at about the same level. Although the validities of Time 2 ASVAB scores may have been slightly underestimated due to variance associated with a reduced level of motivation for some examinees during the experimental testing session, this same factor will lead to underestimates of the validity of any experimental test for subsequent operational criteria. It is concluded that reasonable trust can be placed in estimates of the contributions of experimental tests to ASVAB in the prediction of school grades, even though the ASVAB scores are collected some months prior to administration of the experimental tests.

Results of this study also address a potentially more serious problem. Experimental test scores and intermediate learning criterion scores are often collected in the same experimental testing session. In this case, the correlation computed between experimental and criterion test scores can be inflated by variance associated with the experimental testing situation, and the estimated contributions of the experimental tests to ASVAB are likely to be inflated. Results from analyses conducted using the ASVAB test-retest data indicate that such inflation does indeed exist, but it is not large. The average estimated contribution of an experimental test to ASVAB across eight criterion measures was 12.8%, while the actual computed contribution when the experimental tests were administered along with ASVAB at Time 1 was 10.4%.

Another approach to estimating the contribution of an experimental test to ASVAB would be to readminister ASVAB concurrently with the experimental test. Data in this study indicate that this approach would, on the average, yield values comparable to those which would be obtained if the experimental test had been administered at the testing stations, along with ASVAB.

VI. CONCLUSIONS

With regard to validation studies against concurrent criteria, some inflation in the estimated contribution of experimental tests to ASVAB should be expected. This appears to be due to a subset of individuals who, because of factors occurring in an experimental testing situation, do poorly on both experimental tests and concurrently administered criterion measures. Even if no adjustments are made, the degree of inflation in the estimated contributions is not large. Results from the present investigation suggest that an 18.8% reduction in the estimated contribution would bring such estimates closer to actual contributions. In many studies, such discounts may not be necessary, if an attempt is made to identify and remove unmotivated subjects from analyses.[3]

Until better information becomes available, it is recommended that the estimated contribution of experimental tests to ASVAB in the prediction of subsequent criteria be accepted at face value. In the case of concurrent criteria, it is recommended that an attempt be made to eliminate obviously unmotivated subjects from analyses (such as those who score at or near chance levels on all variables). Estimated contributions may still be slightly inflated, but present results suggest that such inflation is not likely to be large.

References

Hedges, L.V., & Olkin, I. (1985). *statistical methods for meta-analysis*. Orlando: Academic Press, Inc.

Humphreys, L.G. (1960). Investigations of the simplex. *Psychometrika*, 25, 475-483.

Humphreys, L.G., & Davey, T.C. (1988). Continuity in intellectual growth from 12 months to 9 years. Intelligence, 12, 183-197.

Humphreys, L.G., Davey, T.C, & Park, R.K. (1985). Longitudinal correlation analysis of standing height and intelligence. *Child Development*, 56, 1465-1478.

Humphreys, L.G., Parsons, C.K., & Park, R.K. (1979). Application of a simplex process model to six years of cognitive development in four demographic groups. *Applied Psychological Measurement*, 3, 51-64.

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

[3] There is a special problem in evaluating tests of information processing speed. No comparable tests were evaluated in the present study, and there is reason to believe that highly speeded tests are the ones most likely to suffer from situational variance during experimental testing sessions. It is especially important to remove unmotivated subjects from validation studies involving tests of processing speed.