AD A076839

(2)

14

AGO-PRS-RESEARCH [illegible]

Number 52-63

11 October 1952

DDC
RECEIVED
NOV 14 1979
E

(6)

SUMMARY OF LITERATURE ON VALIDATION AND FACTOR STUDIES ON APTITUDE TESTS

PJ 23-52

DDC FILE COPY

Best Available Copy

This document has been approved for public release and sale; its distribution is unlimited.

THIS DOCUMENT IS BEST QUALITY PRACTICABLE
COPY FURNISHED TO DDC CONTAINED A
SIGNIFICANT NUMBER OF PAGES WHICH DO NOT
REPRODUCE LEGIBLY.

PERSONNEL RESEARCH SECTION

PR AND P BR, TAGO

ADB032

79 11 07 110

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DDC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

DISPOSITION FORM

For use of this form, see AR 340-15, the proponent agency is TAGCEN.

REFERENCE OR OFFICE SYMBOL	SUBJECT
PERI-TP	Clearance and Transmittal of Reports to Defense Technical Information Center (DTIC)

TO DDC-DAA-I
ATTN: H. A. Schreckengost

FROM Research Pubs Group,ARI DATE 6 Nov 79 CMT 1
(Price/48913)

1. The reports listed on Inclosure 1 are approved for public release with unlimited distribution (95 numbered ARI Research Memorandums, 51-1 thru 52-82).

2. These are among the unrecorded ARI reports which you identified to us 22 June 1979 as not in your retrieval system. The accompanying box contains at least one copy of each report listed in Inclosure 1, for your retention and reproduction.

3. We will be sending more previously unrecorded ARI reports to you shortly.

Helen S. Price

HELEN S. PRICE
Research Publications
Group, ARI

1 incl
list of reports

Source:(Army Research Institute for the Behavioral and Social Sciences)

✓

AFW

DA FORM 2496 REPLACES DD FORM 96, WHICH IS OBSOLETE.

52-60

SUMMARY OF LITERATURE OF VALIDATION AND FACTOR STUDIES OF APTITUDE TESTS

A

INTRODUCTION

The Army is continually faced with the problem of developing tests to predict performance of enlisted personnel in the wide variety of jobs existing in the Army. This is the focal problem in enlisted classification.

The ideal situation in predicting successful performance on the job exists when the components or factors comprising the elements of the job are known and pure factor tests are available. One would then merely select those tests which are most heavily weighted with the factors that predominate in the job to predict job success. Although the factorial composition of many aptitude tests is known, very little data is available on the factorial structure of jobs. As validation findings and factor analysis results become available for a wide range of jobs, personnel selection methods will become more systematic and effective.

In the absence of direct statistical data on the factorial structure of jobs, an attempt was made to draw inferences as to their factorial structure. These inferences are based upon a comparison of validity coefficients of aptitude tests of known factorial structure. It is then possible to determine the factor composition of the aptitude tests which are best able to predict success in different occupational groups.

Civilian findings are used along with military findings, since for counterpart occupations many of the basic factors are similar. It should be emphasized that this summary is intended only to present tentative generalizations based upon current information. It is hoped that this may serve as a guide for the construction and modification of test batteries used in selection procedures.

BACKGROUND

Guilford(1) states that although correlational analysis alone is sufficient to assess a test's contribution to a battery, factor analysis is necessary in order to define the nature of that contribution. Factor results constitute an indispensable aid to the test constructor who is interested in what his tests measure and why they are valid. By factorial methods one gains insight into the functions responsible for beta weights in regression equations. Guilford points out that the prediction of test validities when the factorial structure of both test and criterion are reasonably well known has been quite successful in the Air Force classification and selection program.

This present survey was undertaken to obtain knowledge of the factorial content and validities of aptitude tests as a source of guidance for revision of existing aptitude tests and for the construction of new tests. The central principle in setting up any battery would be to determine first the factors that account for the respective criterion variance to be predicted

and their relative weights, then to select the best tests in terms of strength and purity in these factors. Since there is little direct evidence of the factorial composition of various criteria in this survey, it is possible only to make generalizations based upon comparisons of the factorial composition of tests which correlate with the various criteria. Factorial analysis and correlation with criteria are available in the literature for many batteries of predictors, hence inferences can be drawn as to what in any particular test contributes to the correlation with the criteria.

SPECIFIC PURPOSE

This survey attempts to summarize the literature on factorial and validity studies for aptitude tests and to provide generalizations as to the effectiveness of factors in predicting success in various occupational groups. So much of the criteria data were obtained from the school situation that this survey emphasizes trainability for various occupations rather than on-the-job performance itself.

PROCEDURE

The overall procedure was to review validity studies, then to review factor analyses involving the types of tests found in the validity studies, and finally to attempt to integrate these reviews into an estimate of the differential validity of the factors in predicting success on the types of jobs included in the validity studies.

After the review of the literature, three reference sources were specifically utilized in reporting validity data. Three additional reference sources were specifically utilized in reporting factor data. The validity data were divided for purpose of analysis into civilian and military studies. In reporting the civilian studies, the tests were placed into groups according to content similarity. Jobs were classified into groups emphasizing psychological functions. As indices of validity, only investigations which could be expressed by means of a coefficient of correlation were reported. In the military studies, weighted mean validity coefficients for the tests of the Navy Classification Battery against grades of enlisted men in training courses were reported. The results of validating the tests of the Army Classification Battery against on-the-job criteria were reported along with orthogonal factor loadings of the tests. Approximations of the factorial structure of test groups according to "factor families" were reported to provide generalizations about the fundamental factors that characterize test areas.

REFERENCE SOURCES

1. The literature of the last ten years was reviewed for studies reporting relationships between aptitude test scores and measures of school or on-the-job success. A survey of factorial studies was also made. Only studies based upon the performance of adults in the United States were reviewed.

After reviewing the literature, the validity data of only three references were specifically utilized. There were:

a. Ghiselli and Brown's survey of the "Validity of Aptitude Tests for Predicting Trainability of Workers," Personnel Psychology, 1951, Volume 4, Number 3, pp. 243-260.

b. U. S. Navy, Test and Research Division, Validity Data for the United States Navy Basic Test Battery, March 1945.

c. Department of the Army, TAGO, Personnel Research Section, Personnel Research and Procedures Branch, Statistical Manuals I and II, December 1944, July 1945.

The results of Ghiselli and Brown's survey are given in Tables 1 through 4.. Since their work includes a review of the literature from 1919, the findings can be considered as adequately sampling the range of civilian occupations. The validity coefficients for two forms of the Navy Basic Test Battery against grades for 22 types of Navy school courses were summarized and tabulated. The average validity coefficients obtained were based upon total sample numbers ranging from approximately 100 to 3500 for each course. The correlations obtained from the Army's statistical manuals were similar to those found in the Navy survey, hence the Navy data were considered for purposes of generalization as representing the trend of validity coefficients in the military. In order to include in this study correlational data based upon actual job performance, reference is made to specific Army studies.

2. Three main sources were utilized for securing factorial data. They were:

a. French's Survey of the Factorial Composition of Aptitude and Achievement Tests, Prepublication Draft, 1949.

b. The Air Force's Aviation Psychology Program Research Reports, Printed Classification Tests, Report No. 5.

c. Department of the Army, TAGO, Personnel Research Section, Personnel Research and Procedures Branch, A Comparison of Army and Airman Classification Batteries. (In preparation)

French's comprehensive survey served to cover the civilian factorial studies, while the Air Force and Army studies explored the factorial composition of the military classification batteries.

SURVEY OF CIVILIAN VALIDITY STUDIES

1. Classification of Tests. In the Ghiselli and Brown study the tests were classified principally upon superficial similarity in content and to some extent on common usage and trait name. Ghiselli and Brown state that although the different tests were grouped together on an intuitive basis, they found little difficulty in placing a test in a given category. The types of tests used were as follows:

a. Intelligence

b. Immediate Memory

c. Substitution

d. Arithmetic

e. Number Comparison

f. Name Comparison

g. Cancellation

h. Tracing

i. Location

j. Pursuit

k. Spatial Relations

l. Speed of Perception

m. Mechanical Principles

n. Dotting

o. Tapping

p. Finger Dexterity

q. Hand Dexterity

r. Arm Dexterity

2. Classification of Jobs. Chiselli and Brown classified occupations into job families emphasizing psychological functions. They recognized that any rationale for job families is subject to criticism. The occupational organization, however, is understandable and meaningful for the purposes of this study. The categories used were the following:

a. Supervisory Occupations

General supervisors

Foremen

b. Clerical Occupations

General Clerks

Recording Clerks (Typists, stenographers, recording machine operators)

Computing Clerks (Bookkeeper, calculating machine operators)

c. Sales Occupations

Sales Clerks (routine sales)

Salesmen (higher sales jobs)

d. Protective Service Occupations

e. Personal Service Occupations

f. Vehicle Operators

g. Trades and Crafts

Mechanical Repairman

Electrical Workers

Structural Workers

Processing Workers

Complex Machine Workers

Machining Workers

h. Manipulative and observational occupations

Machine Tenders

Bench Workers and Assemblers

Inspectors

Packers and Wrappers

Gross Manual Workers

3. Classification of Criteria. In the portion of the analysis dealing with civilian occupations, only investigations utilizing grades in occupational training courses or pass-fail results of courses were considered.

4. Indices of Validity. Only those investigations were included in the study in which test validity could be expressed by means of a coefficient of correlation. All of these coefficients were assumed to be equivalent and were lumped together in computing the average validity coefficients. For each type of test as it applies to each occupational group, the weighted mean validity coefficients were computed through Fisher's Z transformation. Thus, generalizations with respect to the validities of different types of tests for the different occupational groups are given in the form of these average coefficients.

SURVEY OF MILITARY VALIDITY STUDIES

1. Classification of tests and occupations. The two sets of parallel tests comprising the two forms of the Navy Classification Battery were correlated with grades of enlisted men in training courses for specific Navy jobs. The weighted mean validity coefficients for each type of test for each course was computed. Thus, generalizations with respect to validities of different types of tests for the different schools are given in the form of these average coefficients. The types of Navy tests used were as follows:

a. General Classification (sentence completion, opposites, analogies)

b. Reading (items on reading passages)

c. Arithmetic (problems)

d. Mechanical Aptitude (block counting, mechanical comprehension, surface development)

e. Mechanical Knowledge Mechanical (Mechanical tools and information)

f. Mechanical Knowledge Electrical (Electrical tools and information)

g. Clerical Aptitude (alphabetizing, name checking, number checking, spelling)

h. Spelling

i. Radio Code

2. A series of Army studies are reported in which the Army Classification Battery was validated against on-the-job criteria. In the main, carefully constructed rating scales completed by supervisors and co-workers were utilized in securing data of job success. The raters often received instructions directly from personnel technicians to insure an adequate, uniform quality of rating. The Army Classification Battery is comprised of the following tests:

a. Reading and Vocabulary
b. Arithmetic Reasoning
c. Pattern Analysis
d. Mechanical Aptitude
e. Clerical Speed
f. Radio Code
g. Shop Mechanics
h. Automotive Information
i. Electrical Information
j. Radio Information

SURVEY OF FACTORIAL STUDIES

1. Factor loadings of .15 and greater for the tests of the Army Classification Battery are reported in Table '. They were obtained from an orthogonal solution of an intercorrelation matrix of Army, Air Force and Navy test scores of 500 servicemen.

2. Table ' presents approximations of the factorial structure of test groups according to "factor families." Factors were placed into "factor families" according to the degree of similarity of factors. For example, a verbal comprehension factor and a word fluency factor would be considered as belonging to a verbal factor family. Tests were placed in groups or areas on the basis of similarity of factor structure of individual tests. The tests considered as belonging to a verbal abilities test group would be one that repeatedly comes out clear and strong with high loading on every type of verbal factor. Such tests as vocabulary,

sentence completion, opposites, and word groupings would appear in the verbal abilities test group. The approximations are based upon a survey of all available factor studies, emphasizing the military findings. For each test area, the fundamental traits or factors that characterize the area are estimated with emphasis upon whatever factors appear in that area and few if any others. It should be stressed that the terms "similarity of tests" and "similarity of factors," mean only that the verbal appellation applied to the tests and factors on the basis of subject judgments are similar.

PROCEDURAL SHORTCOMINGS

It has been pointed out by Ghiselli and Brown that an analysis of this type has a number of objections. Only insofar as the reader is willing to agree that the deficiencies are relatively minor will the generalizations reached be meaningful. The support for this approach is that errors might be cancelled out in some instances due to the large number of cases and the independent studies made by the investigators cited. In addition, trends in validity and factor structure observed running through the various occupational groups and types of tests provide a safeguard against over-generalization.

Specific shortcomings of this procedure are:

1. The basis on which job-courses were grouped together in occupational families may be seriously questioned. In reviewing the various occupational categories it may be seen that some are quite broad, such as general clerk, while others are very homogeneous, such as protective service workers.

2. The bases for the classification of tests into groups vary considerably from one grouping to another. It also seems likely there will be differences in the reliability of different tests placed in the same group, and perhaps the reliability of a given test will vary significantly from one occupational group to another.

3. Serious questions may be raised with respect to criteria. In some instances the measure of success was grades on tests, while in others the measure was instructors' ratings of students' products. For the on-the-job criteria, supervisors' and co-workers' ratings were used. The stability of ratings varies according to occupations and number of raters, as well as with the ability of the rater.

4. The most serious shortcomings arise out of the validity coefficients themselves. Different types of coefficients are considered to be equivalent. It is also assumed that for any given type of test applied to a given job, differences in magnitude of the validity coefficients reported are simply the result of sampling errors. However, it is known that some of the coefficients are not based upon samples drawn from the same population. There is also the problem of differential restriction in range of ability represented in different samples.

5. Estimating the loadings of test areas on various "factor families" is an attempt to make gross numerical approximations based upon generalizations. These "average" loadings were not estimated by any mathematical procedure. However, all the apparently relevant studies were considered, and such factors as size of sample, number of variables in solution and types of solution were taken into account before an estimate of median loadings was made.

RESULTS AND CONCLUSIONS

CIVILIAN FINDINGS

1. Clerical Occupations. The findings for the clerical occupations are given in Table 1. It will be seen from this table that by and large the best predictions of success in training for all clerical occupations taken together are given by arithmetic tests, closely followed by number comparison tests.

For general clerks all the tests investigated have appreciable validity. Intelligence, arithmetic, and number and name comparison tests have substantial and about equal validity. Substitution, spatial relations, and mechanical principles tests have somewhat less validity and are about equally good.

There is a wealth of information concerning tests for recording clerks. Cancellation and arithmetic tests are the most substantial, followed by intelligence, number comparison, speed of perception, and possibly hand dexterity tests. The predictive efficiency of name comparison, dotting, finger dexterity and arm dexterity tests is practically zero. Correlations are low or moderately low for the remaining tests of immediate memory, substitution, tracing, location, pursuit, spatial relations, mechanical principles, and tapping.

In the case of computing clerks, substitution, arithmetic, and number comparison tests have moderately high validities, and although only limited measures were available, the coefficients obtained for immediate memory and location tests are high enough to appear to have promise. The coefficients for intelligence, name comparison, cancellation, tracing, tapping, and dotting tests are low or negligible.

From a factorial approach (see Table 8 and factor studies), those tests with high loadings on a reasoning factor seem in general to be most valid for clerical jobs. Those tests which have high loadings on a perceptual speed factor also account for much valid variance in a number of the reported coefficients. Thus, for predicting success in clerical occupations reasoning and perceptual speed factors should be included.

Table 1.* Mean validity coefficients of various tests for clerical occupations.

TEST	GENERAL CLERKS	RECORDING CLERKS	COMPUTING CLERKS
Intelligence	.4[illegible]	.2[illegible]	.14
Immediate Memory		.24	.4[illegible]
Substitution	.21	.24	.34
Arithmetic	.43	.41	.3[illegible]
Number Comparison	.42	.29	.3[illegible]
Name Comparison	.40	.01	.1[illegible]
Cancellation		.53	.11
Tracing		.17	.08
Location		.24	.40
Pursuit		.21	
Spatial Relations	.27	.24	
Speed of Perception		.31	
Mechanical Principles	.24	.21	
Tapping		.23	.16
Dotting		.13	.16
Finger Dexterity		.09	
Hand Dexterity		.30	
Arm Dexterity		.09	

*Tables 1 to 4 are taken from the Ghiselli and Brown Survey (pp. 253-256). The size of the populations are given in their report.

2. Protective Service, Personal Service, and Vehicle Operation Occupations. The findings for the protective service, personal service, and vehicle operation occupations are given in Table 2. The only safe generalizations that can be made are for the spatial relations tests which yield a moderate validity. Intelligence tests appear to have high validity except for vehicle operators. The arithmetic tests appear to be the best predictors for personal service workers, and strangely enough, to have moderate negative validity for the protective service workers. The mechanical principles tests have an understandably high relationship for vehicle operators. These tests are generally highly loaded with mechanical experience factors which prove helpful in operating and maintaining vehicles.

Table 2. Mean validity coefficients of various tests for protective services, personal services, and vehicle operators

	PROTECTIVE SERVICES	PERSONAL SERVICES	VEHICLE OPERATORS
Intelligence	.34	.31	.19
Arithmetic	-.25	.39	
Spatial Relations	.32	.42	.45
Mechanical Principles	.28		.48

3. Trades and Crafts. It may be seen from Table 3 that for mechanical repairmen, electrical workers, structural workers, and machining workers, the intelligence, arithmetic, spatial relations, and mechanical principles tests generally gave good predictions. For processing workers, none of the tests investigated is particularly good. For complex machine operators, intelligence and spatial relations tests are fairly good. Name comparisons, number comparisons, pursuit, hand dexterity, and arm dexterity tests have reasonable promise.

From a factorial viewpoint, those tests which have high loadings on a reasoning factor appear to be the best predictors for all occupations in this group except for the machining workers which show up second best. For mechanical repairman, other factors contributing to the valid variance are space, perceptual speed, and mechanical experience. For electrical workers, in addition to reasoning, space and mechanical experience factors appear to be important. For structural workers, reasoning, space, and mechanical experience factors are the most significant. For processing workers, intelligence, perceptual speed, space, and psycho-motor factors are the more significant factors. For complex machine workers, reasoning, space, perceptual speed, and psycho-motor factors are the more important ones. For machining workers, perceptual speed, reasoning, space, mechanical experience, and psycho-motor factors are the most significant.

Table 3. Mean validity coefficients of various tests for trades and crafts.

	Mechanical Repair	Electrical Workers	Structural Workers	Processing Workers	Complex Machine Operators	Machining Workers
Intelligence	.34	.41	.23	.25	.34	.28
Immediate Memory	.14	.07	.15			-.05
Substitution	.34	.34	.19			.27
Arithmetic	.36	.43	.34	.14	-.08	.32
Number Comparison	.22	.16	-.04	.24	.42	.02
Name Comparison		.09	-.01	.14	.57	-.02
Cancellation					.20	.51
Tracing	.21	.29	.24	.17	.22	.21
Location	.24	.24	.23	.24	.28	.24
Pursuit	.17	.12	.18	-.13	.41	.19
Spatial Relations	.35	.33	.27	.29	.36	.31
Speed of Perception	.34		.19			.18
Mechanical Principles	.35	.39	.31			.33
Tapping	-.01	.11	.20	-.01	.24	.05
Dotting	.20	-.14	.13	.02	.26	.14
Finger Dexterity	.15	.15	.24	.21	.11	.24
Hand Dexterity	.17	.38			.40	
Arm Dexterity	.08	-.10			.34	-.03

4. Manipulative and Observational Occupations. It may be seen from Table 4 that no conclusions may be made for machine tenders. For bench workers, arithmetic and dexterity tests yield good predictions. Factorially, the reasoning and psycho-motor factors give the best results for bench workers.

For inspectors, no test of any great value is indicated. The best predictors are spatial relations tests and arithmetic tests.

Information concerning packers and wrappers is restricted to arithmetic, intelligence, and spatial relations tests. The first of these tests is rather good while the other two are of considerably less value.

For gross manual workers the only data available concerned intelligence tests which appear to have little value. Perhaps this is so, since individuals engaged in such occupations need a certain minimal intellectual ability, and having that minimum, it may not enter into consideration as an important source of variance.

Table 4. Mean validity coefficients of various tests for manipulative and observational occupations.

	Machine Tenders	Bench Workers	Inspectors	Packers and Wrappers	Gross Manual Workers
Intelligence	-.31	.02	.19	.22	-.03
Arithmetic		.39	.27	.43	
Tracing		.16	.09		
Location		.29	.19		
Pursuit		.28	.09		
Spatial Relations		.24	.27	.24	
Speed of Perception		.26	.22		
Tapping		.16	.10		
Dotting		.22	.08		
Finger Dexterity	.21	.44	.00		
Hand Dexterity		.30			
Arm Dexterity		.54			

5. Conclusions for Civilian Occupations. When one compares the magnitudes of the validity coefficient of the various tests for the different occupational groups, few consistent differences appear. In general those tests that are most effective for one group are the most effective for the other groups, although the tests which are next highest in effectiveness do vary to a degree. This is not to imply that some of the types of tests considered have any differential validity, but rather that occupational differences in test validities are likely to be much smaller than is generally believed. In general the tests with high loadings on a general reasoning or intellectual factor account for the largest contribution to the valid variance.

NAVY FINDINGS

The findings for the Navy tests correlated with grades of enlisted men in training schools are given in Table 5. It may be seen that for each course with the exception of Cooks and Bakers, Machinist's Mates, Radio and Signal, at least one test in the batteries has a validity coefficient of .50 or above after correction for restriction in range of the predictor since the test was employed with a cutting score as a basis for selection. The Arithmetic Reasoning test has the highest average validity coefficient in the first battery for all courses. The other tests, in descending order of magnitude of average validity coefficients are: General Classification, Reading, Mechanical Aptitude, Mechanical Knowledge Electrical, and Mechanical Knowledge Mechanical. The results of the second battery are, in general, the same.

From a factorial viewpoint, the tests with the highest loadings on a reasoning factor make the largest contribution to the valid variance. For many of the technical courses a mechanical experience factor appears to account for a considerable portion of the correlation with the criteria. Little differentiation in the pattern of validities is obtained from the various types of mechanical tests. This is a further indication of a general mechanical experience factor present in these tests.

The Clerical and Radio Code tests generally show less validity than do the reasoning tests. The clerical tests appear to have a moderate order of validity for clerical courses.

ARMY FINDINGS

In Table 6 are shown the validity coefficients obtained for the Army Classification Battery tests against ratings of job proficiency of enlisted men on various Army jobs. These correlations are uncorrected for restriction in range and therefore considerably smaller than they would be for a heterogeneous validation group. It may be seen from Table 6 that the Arithmetic Reasoning test exhibits general validity for all the jobs represented in this study. This is further evidence of the validity of reasoning factors for a wide range of occupations. The Pattern Analysis test which also has a moderately high reasoning factor content does not exhibit as high an order of validity as does the Arithmetic Reasoning test. Perhaps the greater validity of the Arithmetic Reasoning test for this group of jobs may be attributable to a number factor.

A very significant result is the high order of correlations found for the mechanical and technical information tests with the various occupational groups reported. In general, the Electrical and Radio Information tests, as a group, exhibit similar validity patterns. The Mechanical Aptitude, Shop Mechanics and Automotive Information tests, as a group, also exhibit similar validity patterns. It seems plausible that a mechanical experience factor appears to be common to all the information tests in this battery.

Table 5.* Weighted Navy validity coefficients (corrected for restriction in range)**

	BASIC TEST BATTERY 1						BASIC TEST BATTERY 2								
COURSE	GCT	READ.	ARITH.	MAT	MEM	MKE	GCT	READ.	ARITH.	MAT	MEM	MK3	CLER. APT.1	SPELL 1	RADIO CODE 2
Avn. Machinist's Mate							.44	.31	.31	.48	.64	.56			
Aven. Ordnancemen	.48	.47	.52	.32	.32	.40	.53	.39	.37	.28	.36	.48			
Avn. Radioman	.49	.43	.34	.27	.29	.35	.54	.47	.45	.37	.22	.30			
Basic Engineering	.51	.48	.60	.53	.47	.42	.50	.42	.56	.52	.62	.50			
Cooks and Bakers	.14	.17	.30	.12	.19	.3'									
Diesel	.42	.35	.38	.25	.42	.46	.71	--	.53	--	.76	.55			
Electrical	.52	.49	.57	.41	.35	.45	.50	.38	.51	.50	.36	.57			
Fire Control	.52	.50	.55	.41	.17	.19	.57	.47	.49	.34	.35	.54			
Fire Control (O)							.51	.43	.40	.41	.21	.30			
Gunnery	.39	.42	.36	.33	.38	.38	.31	.29	.30	.46	.53	.52			
Hospital Corps	.47	.40	.37	.29	.19	.30	.72	.59	.56	.46	.25	.5.			
Hospital Corps (WR)							.62	.52	.46	.20	--	--			
Machinist's Mate	.37	.29	.42	.46	.40	.44									
Quartermaster	.42	.45	.52	.30	.11	.12	.48	.46	.41	.30	.08	.18	.42	.17	
Radar	.60	.62	.61	.46	.34	.37	.65	.52	.48	.50	.30	.50			
Radio	.35	.32	.42	.18	.37	.1(	.25	.23	.2?	.12	.13	.18	.25	.20	.1'
Signal	.44	.41	.43	.2?	.07	.10	.44	.32	.37	.32	.13	.32	.46	.2?	.2-
Storekeeper	.49	.40	.53	.26	.21	.2(	.56	.47	.54	.3?	.14	.32	.43	.25	.01
Storekeeper (WR)	.42	.46	.50	.36	--	--	.52	.36	.44	.17	--	--	.48	.27	
Torpedo	.33	.37	.20	.28	.39	.40	.24	.21	.2k	.34	.33	.39			
Yeoman	.49	.48	.60	.25	.31	.24	.46	.36	.33	.25	.10	.13	.54	.31	.2'
Yeoman (WR)	.63	.59	.62	.38	--	--	.65	.48	.47	.2(	--	--	.46	.31	

* SAMPLE NUMBERS range from approximately 100 to 3'00 for each course.

** The criterion and predictor ranges are restricted by virtue of the fact that the predictor was employed with a cutting score as the basis for selection.

Table 7. Correlation of the Army Classification Battery Test with job proficiency ratings of enlisted men on various Army jobs

TEST	FIELD INSPECTION	CENTRAL OFFICE INSTAL.	FIRE CONTROL MAINTENANCE	CARTOGRAPHY	RADAR REPAIR	FIXED STATION RADIO REP.	PREVENTIVE MEDICINE
	N = 200	N = 198	N = 84	N = 71	N = 58	N = 31	N = 19
Reading and Vocabulary	.16	.12	.07	.50	.07	.37	.29
Arithmetic Reasoning	.19	.13	.18	.46	.17	.27	.31
Pattern Analysis	.06	.06	.10	.26	.20	.13	-.12
Mechanical Aptitude	.10	.10	.10	.48	.28	.23	.41
Army Clerical Speed	.00	.18	-.10	.41	.05	-.10	.24
Army Radio Code	.19	.11	-.04	.28	.15	.15	-.18
Shop Mechanics	.10	.12	.26	.54	.23	.19	.44
Automotive Information	.08	.11	.23	.60	.34	.31	.68
Electrical Information	.07	.28	.20	.5	.1	.48	.1
Radio Information	.04	.23	.01	.47	.35	.46	.13

It is interesting to note that in the validity studies reported, those tests which tap factors other than reasoning are more highly correlated with ratings of on-the-job proficiency than with ratings of success in training. Conversely, the tests with a high reasoning content appear to predict trainability better than do the non-reasoning tests. In general, all of the tests show greater differential validity against on-the-job criteria than they show against training grades. Perhaps one explanation for this trend is that in the training situation an individual is evaluated, in part, on a generalized impression of intellectual level or learning ability. There is also the tendency in the school situation to emphasize theoretical aspects. This would tend to maximize the validity of tests tapping an intelligence or reasoning factor. On the other hand, the on-the-job evaluations may be expected to represent more of the actual job performance. The latter situation would tend to maximize the validity of tests which tap the more specific elements which account for the variance in performance on-the-job.

FACTOR RESULTS

1. The Army Classification Battery, the Airman Classification Battery, and the Navy Mechanical Test, along with the Army and Air Force interest inventories, were administered to a group of 500 soldiers and airmen. An intercorrelation matrix of 44 variables was computed and an orthogonal solution obtained. Table 7 presents factor loadings of .15 and above on seven factors for the tests in the Army Classification Battery. It may be seen from Table 7 that the Arithmetic Reasoning test has significant loadings on general reasoning, numerical facility, verbal comprehension, visualization, and mechanical experience factors. These factors account for much of the variance in the different criteria investigated. Hence, the pattern of general validity obtained with the Arithmetic Reasoning test in light of its factorial structure is understandable. It should also be noted that both the Pattern Analysis and Reading and Vocabulary tests have moderate loadings on general reasoning and verbal comprehension. These tests also exhibit their greatest predictive value in the school situation. An interesting finding is that the Mechanical Aptitude, Shop Mechanics, and Automotive Information tests have their highest loadings on the mechanical experience factor and have very similar loadings on the other factors. This indicates that the three tests may be replaced by one of the tests without a great loss of predictive efficiency. A similar conclusion may be reached about the Electrical Information and Radio Information tests.

2. In order to make generalizations about the factorial structure of test groups or areas, the loadings of these tests according to factor families were approximated. These approximations are intended only as indications of the trend of the factorial structure of test groups. The loadings are based upon median loadings obtained from the available literature on factor studies with particular emphasis on military findings. In Table 8 are given the factor loadings of seven test groups on various factor families.

Most reasoning tests, but not all, prove to have variance in a factor identified only as general reasoning, which is best measured by arithmetic reasoning. Other reasoning factors have been identified, but are difficult to purify.

Table 7. Factor loading of .15 and above for the tests of the Army Classification Battery from the orthogonal solution of 44-variable matrix

VARIABLE	FACTORS						
	PERCEP SPEED	GENERAL REASONING	VERBAL COMPREHENSION	MECHANICAL EXPERIENCE	NUMERICAL FACILITY	INFORMATION (TECHNICAL)	VISUALIZATION
Reading and Vocabulary	15	27	55	23	37	24	17
Arithmetic Reasoning		47	38	29	48	15	30
Pattern Analysis	36	47	23	26	23		45
Mechanical Aptitude	18		22	49	25	20	43
Clerical Speed	34		31		55		18
Radio Code	37	21		21	29	24	
Shop Mechanics	19	21	32	59	23	18	37
Automotive Information		17		76	15	15	23
Electrical Information		27	15	41	17	59	17
Radio Information				18		62	

Table 8. Approximations of factor loadings of test groups on various factor families

TEST GROUP	FACTOR FAMILY I REASONING	II VERBAL	III NUMBER	IV SPACE	V MECH. EXP.	VI MEMORY	VII PSYCHOMOTOR	VIII PERC. SPEED
Arith. Reasoning	60	40	50	30	30	00	00	00
Verbal Reasoning	45	30	00	00	00	00	00	00
Verbal Abilities	30	55	35	00	20	00	00	00
Numerical Operations	00	00	80	00	00	00	00	00
Spatial Tests	45	20	20	50	00	00	00	35
Mechanical Tests								
Mechanical Inf.	00	20	00		70	00	00	30
Mechanical Prin.	00	25	00	25-50	55	00	00	00
Elec. Inf.	20	25	00		65	00	00	00
Memory	00	00	00	00	00	75	00	00
Psychomotor	00	00	00	00	00	00	65	00
Perceptual Speed	00	00	00	00	00	00	00	65

The verbal tests fall into two groups: general vocabulary and reading comprehension. The vocabulary tests usually prove to be purest and strongest as measures of the verbal factor. The reading comprehension tests are usually of a complex factorial composition. Since all the factorial components of the reading comprehension tests are better measured by purer and stronger tests for them, it may be advisable that a general vocabulary test be utilized when it is desired to measure the verbal factor.

The unique feature of the numerical operations tests is their variance in the numerical factor. Of all mathematical tests, numerical operations has shown its strength wherever a measure of sheer numerical facility is wanted.

The data on spatial tests indicate that the principal unique functions measured can be explained by several factors and an additional factor which appears to be closely related. These space factors seem to involve relating different stimuli to different responses, either stimuli or responses being arranged in spatial order. The related factor, visualization, is strongest in tests that present a stimulus either pictorially or verbally, and in which manipulation or transformation to another visual arrangement is involved. It has been especially difficult to construct items which measure visualization and that do not involve a degree of reasoning. If a visualization problem is too difficult, it is probably solved by reasoning. By reducing the difficulty, reasoning variance seems to be reduced. If the problems are made too easy, however, they can be solved without visualization, possibly by a simple space ability.

Many varieties of mechanical tests have been constructed which were thought to cover familiarity with machinery, a natural "knack" for understanding new mechanisms, or an "inborn" aptitude for success in mechanical tasks. The understanding of the validity that can be called unique to mechanical tests lies in the mechanical experience factor. It would appear that "mechanical aptitude" tests generally tap an acquired trait. Opportunities for mechanical experience are not equal, and hence it may be reasonable to expect two factors, one attributable to the power to learn and the other to opportunity. It may be that two such factors are inextricably mingled in any analysis.

An exploration of the area of memory tests revealed that there exist several memory factors. They are: paired-associates memory, visual-memory, and object-name memory. It appears that little differential validity is obtained with the various memory factors.

The tests in the psychomotor area cover such tests as dexterity, tapping, and reaction time. In general, they include tests which require manipulative, visual, and auditory responses.

Perceptual speed tests require the comparison of a perceived configuration with a remembered one. Tests of a clerical type generally have a substantial variance in the perceptual-speed factor.

4. General Conclusions. There appear to be a limited number of "factors" appearing in both test and criteria which can be adapted to vocational prediction in terms of test batteries. The usual multiple regression procedure which allots optimum weights to the several tests of a battery can be successful only to the extent that the battery is appropriate to the criterion. It is generally agreed that tests are useful in a battery if they contribute uniquely to the criterion variance. It is believed that factorial analysis and factor hypothesis prior to test-battery construction are a necessary step in the process of developing uniquely contributing tests. The problem is made more difficult by the fact that jobs for which success is to be predicted are, in more factorially-complex than the tests which are used in the prediction. This implies factorial analysis of job structure as well as of tests. Reasoning test data are particularly difficult to interpret. Validity coefficients are high for reasoning tests against school grade criteria in many different training courses. They are surpassed by validities of non-reasoning tests on job-performance criteria. An explanation might be made in terms of greater need for "reasoning" in school courses than in the corresponding job performance. It could be argued that reasoning tests have the validities which are demonstrated on job performance criteria only because of a general intellectual halo effect incorporated into the performance rating. The delineation of the participation of reasoning factors in job structures requires the inclusion of performance criteria for many types of jobs in factor studies. Such studies should include various types of reasoning tests as reference variables.

PERSONNEL

Program Coordinator: Julius Uhlaner
Project Director: Edmund F. Fuchs
Research Associate. Joseph Zeidner

PREPARATION OF REPORT: 12 May 1952

REFERENCES

1. GUILFORD, J. P. Aviation Psychology Program Research Reports, Printed Classification Tests, Report No. 5, U. S. Government Printing Office, Washington 25, D. C.