Technical Research Note 199

AD



# USE OF ARMY SCHOOL SAMPLES IN ESTIMATING ACB TEST VALIDITY

Pauline T Olson

STATISTICAL RESEARCH AND ANALYSIS DIVISION

BESRL

U. S. Army
Behavioral Science Research Laboratory



August 1968

# BEHAVIORAL SCIENCE RESEARCH LABORATORY

## An activity of the Chief, Research and Development

J. E. UHLANER
Director

JAMES E. WIRRICK
COL, GS
Commanding



NOTICES

DDC AVAILABILITY: U. S. Military and Government agencies may obtain copies of this report directly from DDC. Other qualified users will request through the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, Springfield, Virginia 22151. Dissemination of this report to foreign nationals by DDC is not authorized.

DISTRIBUTION: Primary distribution of this report has been made by BESRL. Please address correspondence concerning distribution of reports to: U. S. Army Behavioral Science Research Laboratory, Attn: CRDBSRL, Washington, D. C. 20315.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Behavioral Science Research Laboratory.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Technical Research Note 199

AD

# USE OF ARMY SCHOOL SAMPLES IN ESTIMATING ACB TEST VALIDITY

Pauline T. Olson

STATISTICAL RESEARCH AND ANALYSIS DIVISION
Cecil D. Johnson, Chief

U. S. ARMY BEHAVIORAL SCIENCE RESEARCH LABORATORY

Office, Chief of Research and Development
Department of the Army

Washington, D. C. 20315

August 1968

Army Project Number
2T061101A91B

(ILIR)
DOL 33

BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

## FOREWORD

In-house Laboratory Independent Research (ILIR) provides for the application of BESRL's scientific talent in the exploration of new developments in experimental psychology, psychometrics, and statistical-mathematical models. Its objective is to extend laboratory capability in terms of knowledge and techniques from which applications can ultimately be made to a wide range of Army human factor problems. A portion of the FY 1968-69 effort of the Statistical Research and Analysis Division of BESRL has therefore been devoted to research under RDT&E Project 2T061101A91B. The Division, in its regular program, conducts supportive research related to human factors problems in which quantitative considerations enter substantially.

The present study, which was under the general guidance of Cecil D. Johnson, Chief of SRAD, and John Mellinger, Chief of the Statistical Research and Consultation Branch acting in an advisory capacity, deals with a reexamination of a statistical method commonly employed in estimating the effectiveness of a selection test when validity data are restricted to a segment of the population with which the test is to be used.

J. E. UHLANER, Director
U. S. Army Behavioral Science
Research Laboratory

## USE OF ARMY SCHOOL SAMPLES IN ESTIMATING ACB TEST VALIDITY

## BRIEF

Objective:

To reexamine the statistical method currently used for estimating the predictive efficiency of tests of the Army Classification Battery, specifically, the method by which validity coefficients are corrected for restriction of samples to a segment of a given population.

Procedure:

Descriptive statistics--means, standard deviations, intercorrelations--for subgroups of an entering population of 2480 enlisted men were the basis for computation of difference ratios for partial correlation coefficients between subgroups and total group, and also regression weights. Means and standard deviations of the difference ratios were compared.

Findings:

1. No appreciable effect of the curtailed sample on the estimated effectiveness of the cognitive tests in the total group was demonstrated.

2. Differences between statistics for curtailed samples and for the total population were found to be somewhat larger in the case of noncognitive measures than for cognitive tests.

Utilization of Findings:

Results raised some slight question as to the appropriateness of estimating the validity of noncognitive tests by statistical correction for restriction in the range of scores in curtailed samples. Some factors entering into the classification process as applied by interviewers at reception stations--but not quantified or recorded--may not correspond closely to characteristics measured by the noncognitive tests in the battery administered. Replication of the present study on more recent data is projected.

USE OF ARMY SCHOOL SAMPLES IN ESTIMATING ACB TEST VALIDITY

## CONTENTS

TABLES

# USE OF ARMY SCHOOL SAMPLES IN ESTIMATING ACB TEST VALIDITY

## BACKGROUND

Assignment of entering Army personnel to training in military occupational specialties (MOS) is generally based on a combination of the stated needs of the total Army system at the time assignment is made, the personal preferences and previous experience of the individuals available for assignment, and scores on the tests used by assignment specialists to predict success in the various Army occupations. The tests used are those of the Army Classification Battery (ACB) administered during reception station processing. Eleven paper-and-pencil tests, nine cognitive, one self-descriptive, and one on general information designed to measure interests, constitute the ACB (see Appendix A for a fuller description of the tests). The scores on the individual tests of the ACB are combined into eight single-digit scores by simple weighting of scores on specified tests. These composite test scores, called aptitude area scores, have demonstrated relationship to performance in Army occupations falling into broad groups. Cutting scores for the training courses are set so as to admit only individuals who have reasonable chance of success in the course.

One of the basic responsibilities of the Behavioral Science Research Laboratory is to maintain and improve the ACB tests and the test composites as used in an optimal assignment model. The final measure of the quality of a classification test is its effect on the magnitude of the objective function, in this case, predicted performance of the individuals in the jobs to which they are assigned. The validity of the tests, determined as the strength of the relationship of a test with a criterion variable, is used as a substitute measure of classification effectiveness in most BESRL research. Since it is not always possible to obtain a satisfactory measure of job performance, grades in Army school training courses have been frequently used as the criterion.

Army school success and job performance can be measured directly only for subsamples of individuals already assigned to training or job. Relationship of predictors to school grades or evaluations of job performance can therefore be obtained only on the separate occupations. By the nature of the classification process, comparative performance ratings for a single individual across several jobs or training schools cannot be obtained. Even within an occupational subgroup, validity is known only for those who have successfully met the school requirements.

Extension of results obtained from a selected subgroup to a hypothetical whole on the basis of known relationships among variables has been an accepted statistical procedure for many years. As early as 1903, K. Pearson (1) noted the effects of truncating a sample by selection on one variable, and presented a method for the extension of sample results to the complete distribution. Since that time, Aitken (2), Lawley (3), Birnbaum (4), Guilford (5), Gulliksen (6), and others have presented methods generalized to several selector variables.

Within the Behavioral Science Research Laboratory, extension of results on a restricted MOS sample to the uncurtailed recruit population by use of statistical relationships among variables on which complete information is available has been an accepted practice. The Army Classification Battery is administered to all entering recruits, and means, variances and covariances of the tests for the full Army population are therefore known. About criterion measures, however, or about any other measurements made within MOS subgroup, only limited information is available. Selection of a subgroup from a widely variant population usually results in restriction of the range of scores on variables subsequently measured, especially if the group has been formed by rejection of would-be members because of unacceptable scores on closely related variables. Restriction of the range of scores results in restriction of the variances and covariances of variables for the subgroup. For this reason, extension of subgroup results to the uncurtailed group is sometimes called correction for restriction in range. Strictly speaking, correction for restriction in range should be made with information available on the full set of variables entering into selection.

In Army classification procedures, information in addition to that provided by scores on the ACB is used by the assignment interviewer in making his recommendation for the assignment of recruits. However, the exact nature of this information is not known. Nor is it known how the information is used by the interviewer nor how interviewer recommendations are used in the assignment process. Correction procedures have therefore used the ACB tests as if they were the only variables of direct selection.

The basic statistical model for making adjustments in variances and covariances among variables measured in a subgroup so as to represent the same measurements in the uncurtailed distribution (including the subgroup) makes certain assumptions about the constitution of the subgroup, the variables, and the measurements. One assumption is that the subgroup is assembled from the whole distribution by specifying that individuals assigned to the subgroup meet certain standards in size of scores or linear combination of scores for the selector variables. The exact nature of these standards need not be known by the person making the adjustments. Other assumptions are:

1. The regression of the variables whose variances and covariances are to be corrected is linear on the variables of direct selection.

2. The variances and covariances of the affected variables are constant for fixed values of the selector variables.

3. The errors of measurement are the same in the restricted group and in the whole distribution.

If these assumptions are fully met, the partial regression weights and partial covariances obtained after elimination of the effects of the selector variables should be the same for subgroup and total group. In the present study, this equality is examined using experimental data.

Behavioral Science Research Laboratory studies in which subgroup results were extended to universe values have generally assumed that the required conditions prevailed; or, if not, that the technique was sufficiently robust to tolerate moderate violation of the assumptions. However, detailed evaluation of the degree of conformity to these conditions using an extensive amount of empirical data has not been made.

The present study is based on data collected for validation of tests developed as potential components of the Army Classification Battery. The group of enlisted men represented most of the new input to the Army for a period of 18 months. Assignment of individuals to MOS schools was accomplished in the usual way, through the use of priorities resulting from Army needs, preferences of the individuals, and occupational aptitudes as indicated by scores on the ACB. Additional tests were administered to all members of the group, but these scores were not used in the classification process. Hence, a unique opportunity existed. Complete information was available on variables used in subdividing the population into occupational subgroups as well as on additional variables for which the range of scores in the subgroups was restricted because of their relationship to the variables of direct selection. The correction process could be carried out and results compared with actual results observed for the total group. Suitability of the correction process could also be examined by testing the conditions required by the underlying assumptions. An hypothesis that variables are differentially affected by correction based on an incomplete set of selectors was also of interest. Some of the additional tests given were noncognitive in nature, being made up of items related to occupational preferences, past experience, and personal descriptions. Scores on these tests might be related to supplementary factors used by the interviewers in assigning men to occupational groups. If correction is based on the ACB alone, restriction in the range of scores due to relationship with the supplementary factors would be left uncorrected, and a poor estimate of the population covariances for the noncognitive tests would result.

## METHOD

### Sample

The study used information collected from a total of 2480 men at the beginning of their first enlistment. The men were subsequently assigned to 21 different MOS school courses but because some of the subgroups were very small, closely related MOS were combined for the analysis. Subgroup size and MOS (by the MOS classification structure in use when the tests were administered) are shown in Table 1.

Table 1

IDENTIFICATION OF SUBGROUPS BY MOS AND SIZE

| Subgroup | MOS | Title | N |
|---|---|---|---|
| 1 | 231.1<br>231.2 | Artillery Fire Control Repairman<br>Light Fire Control Equipment Repairman | 111 |
| 2 | 230.0<br>250.0 | Ordnance Electronics Entry<br>Electronic Repair Helper | 305 |
| 3 | 293.1 | Radio Relay and Carrier Operator | 244 |
| 4 | 294.1<br>296.0<br>296.1 | Field Carrier Equipment Repairman<br>Field Radio Repairman<br>Field Radio Repairman | 188 |
| 5 | 321.1 | Field Communications Lineman | 265 |
| 6 | 357.1 | Guided Missile Installation Electronic Equipment Repair | 103 |
| 7 | 440.0 | Metalworking Helper | 157 |
| 8 | 511.1<br>530.0 | Carpenter<br>Chemical Operations Helper | 275 |
| 9 | 612.1 | Construction Machine Operator | 177 |
| 10 | 670.0<br>680.0 | Aircraft Maintenance Crewman<br>Aircraft Components Repair Apprentice | 264 |
| 11 | 722.1<br>730.0<br>053.1 | Cryptography<br>Finance Clerk<br>Radio Teletype Operator | 281 |
| 12 | 271.1<br>281.1 | Fixed Station Receiver Repairman<br>Microwave Radio Repairman | 110 |
| 13 | | Pooled Samples 1-12 | 2480 |

Predictor Tests

Variables of direct selection were tests of the nine-test Army Classification Battery. Two additional tests, the Classification Inventory and the General Information Test, have been added to the ACB since the data for the present study were collected. Of the other tests administered to the twelve subsamples, fourteen were cognitive and nine noncognitive.

| Variable No. | Army Classification Battery Tests |
|---|---|
| 1 | Verbal |
| 2 | Arithmetic |
| 3 | Pattern Analysis |
| 4 | Mechanical Aptitude |
| 5 | Army Clerical Speed |
| 6 | Army Radio Code Aptitude |
| 7 | Shop Mechanics |
| 8 | Automotive Information |
| 9 | Electronics |
| | **Cognitive Tests** |
| 10 | Object Completion |
| 11 | Word Squares |
| 12 | Pattern Analysis |
| 13 | Practical Situations |
| 14 | Reaction to Signals |
| 15 | Mechanical Principles |
| 16 | Spatial Orientation |
| 17 | Letter Combinations |
| 18 | Hidden Figures |
| 19 | Attention to Detail |
| 20 | Patterns |
| 21 | Army Perceptual Speed |
| 22 | Associative Memory |
| 23 | Subtraction - Division |
| | **Noncognitive Tests** |
| 24 | General Information Test |
| 25 | Classification Inventory |
| 26 | Clerks (a priori scale) |
| 27 | Electronics (a priori scale) |
| 28 | Mechanics (a priori scale) |
| 29 | General Adjustment (empirical scale) |
| 30 | Clerks (empirical scale) |
| 31 | Mechanics (empirical scale) |
| 32 | Mechanics suppressor (empirical scale) |

### Statistical Procedure

To test for possible differences in partial correlation coefficients and regression weights from the individual subgroups to the pooled sample of 11 other subgroups, partial correlation coefficients and partial regression weights were computed for each combination of variables. Differences in weights or coefficients were then compared to the combined error of the difference.

### Formulas

Formulas used in calculation of the difference ratios for partial correlation and for regression weights are shown below:

| | |
|---|---|
| $N_i$ | Number of persons in subgroup i; i = 1 to 12 |
| $N_p$ | Number of persons in the pooled group, $N_p = \sum_{i=1}^{12} N_i$ |
| ${}_pC_{xx}$ | 9 x 9 matrix of covariances for the ACB tests in the pooled group |
| ${}_pC_{xy}$ | 9 x 23 matrix of covariances between the ACB tests and the experimental tests in the pooled group |
| ${}_pC_{yy}$ | 23 x 23 matrix of covariances among the experimental tests in the pooled group |
| ${}_iC_{xx}$, ${}_iC_{xy}$ and ${}_iC_{yy}$ | Same as above for subgroup i; i may equal 1 to 12 |
| ${}_{pi}C_{xx}$, ${}_{pi}C_{xy}$ and ${}_{pi}C_{yy}$ | Same as above for the pooled <u>eleven</u> subgroups, omitting the <u>i</u>$^{th}$ subgroup. ${}_{p2}C_{xx}$ is the covariance matrix for the pooled samples 1, 3-12, ACB tests |
| ${}_iR_{yy}$, ${}_{pi}R_{yy}$ | Partial correlation coefficients for the i$^{th}$ subgroup and for the pooled 11 other subgroups |
| ${}_iW_{xy}$, ${}_{pi}W_{xy}$ | Partial regression weights for the i$^{th}$ subgroup and for the pooled 11 other subgroups |
| ${}_p\bar{C}$ | ${}_pC_{yy} - {}_pC_{xy}' \, {}_pC_{xx}^{-1} \, {}_pC_{xy}$, matrix of partial covariances in the pooled group (all 12 subgroups) |

$_i\bar{C}$ — ${}_iC_{yy} - {}_iC'_{xy}\,{}_iC^{-1}_{xy}\,{}_iC_{xy}$, matrix of partial covariances in the $i^{th}$ group

$_{pi}\bar{C}$ — ${}_{pi}C_{yy} - {}_{pi}C'_{xy}\,{}_{pi}C^{-1}_{xy}\,{}_{pi}C_{xy}$, matrix of partial covariances in the pooled 11 subgroups, excluding the $i^{th}$ subgroup

$_iD_W$ — ${}_iW_{xy} - {}_{pi}W_{xy}$

$_iD_R$ — ${}_iR_{yy} - {}_{pi}R_{yy}$

The following computations make use of the elements of matrices defined above:

$$ {}_i\Delta_{w(jk)} = \sqrt{S^2_{B(jk)}\left(\frac{1}{N_i-10} + \frac{1}{N_p-N_i-10}\right)} \text{ where } S^2_{B(jk)} = {}_p\bar{C}_{jj} \cdot {}_pC^{-1}_{xx(kk)} $$

$$ {}_i\Delta_{r(jk)} = \sqrt{S^2_{r(jk)}\left(\frac{1}{N_i-9} + \frac{1}{N_p-N_i-9}\right)} \text{ where } S^2_{r(jk)} = 1-R^2_{jk} \text{ for the pooled sample} $$

$$ \text{Ratio of Comparison} = \frac{Djk}{\Delta jk} $$

Partial variances were compared by the F-test, ${}_{pi}C_{ii}/{}_iC_{ii}$.

### Statistical Processing

Although it was desirable to know whether statistically discernible differences existed between the restricted subgroup and the total subgroup from which the subgroup was selected, an appropriate statistical test for making this evaluation directly was not available. Consequently, it was decided to test for differences between the two parts, the subgroup and the corresponding pooled group (composed of the 11 other subgroups) on the premise that if the two independent parts were not different, the subgroup would certainly not be different from the whole (composed of all 12 subgroups).

Intercorrelations, means, and standard deviations for the full set of 32 variables were available from the statistical processing of a previous study. Covariances computed from these available measures were used to obtain partial correlation coefficients and partial covariances among the 23 variables not used in the assignment process and partial regression weights for these 23 variables on the nine tests of the selection battery.

The differences between the partial correlation coefficients obtained for the subgroup and those obtained for the pooled remaining 11 subgroups were computed for each pair of variables in the 23-variable matrix. These differences were divided by the standard error based on the entire 12 subgroups weighted by the square root of the sum of reciprocals of the degrees of freedom for the subgroup and the corresponding pooled 11 subgroups. Since multiple tests were made, there was no effort to single out individual differences which might be statistically significant. However, a comparison was made between the obtained distributions of difference ratios and the normal distribution. A similar comparison was made for difference ratios for regression weights.

## RESULTS

Means and standard deviations for the difference ratio distributions for partial correlation coefficients and for regression weights are shown in Table 2. In all except one subgroup, the standard deviation of the distribution of difference ratios was equal to or greater than one (+1), indicating somewhat flattened distributions with a greater number of large ratios than is found in the normal distribution. A like result was obtained for partial correlation coefficients and for regression weights.

The frequency and proportion of differences greater than two standard errors were tabulated (Table 3). The percent of area under the normal curve with abscissa 2.0 or greater is 4.55. Averaged across the 12 subgroups, 6 percent of the regression weight ratios were equal to or greater than 2.0.

Table 2

MEANS AND STANDARD DEVIATIONS OF DIFFERENCE RATIOS

| | Coefficients | | | |
|---|---|---|---|---|
| | Partial Correlation | | Regression | |
| Sample | Mean | SD | Mean | SD |
| 1 | .205 | 1.04 | .084 | 1.11 |
| 2 | -.012 | 1.07 | -.037 | 1.12 |
| 3 | .003 | 1.09 | .047 | 1.08 |
| 4 | .107 | 1.14 | .045 | 1.04 |
| 5 | .780 | 1.08 | .180 | 1.08 |
| 6 | -.266 | 1.09 | -.540 | .97 |
| 7 | .051 | 1.09 | -.230 | 1.01 |
| 8 | .075 | 1.22 | -.019 | 1.17 |
| 9 | .000 | 1.00 | -.036 | 1.11 |
| 10 | .060 | .99 | -.144 | 1.01 |
| 11 | -.168 | 1.21 | -.032 | 1.15 |
| 12 | -.117 | 1.06 | -.083 | 1.02 |

Table 3

NUMBER AND PERCENTAGE OF DIFFERENCE RATIOS GREATER THAN 2.0

| | Coefficients | | | |
|---|---|---|---|---|
| | Partial Correlation | | Regression | |
| Sample | Number | Percentage | Number | Percentage |
| 1 | 17 | 7.7 | 10 | 4.8 |
| 2 | 16 | 6.3 | 17 | 8.2 |
| 3 | 22 | 8.7 | 10 | 4.8 |
| 4 | 19 | 7.5 | 13 | 6.3 |
| 5 | 15 | 5.9 | 9 | 4.3 |
| 6 | 16 | 6.3 | 8 | 3.9 |
| 7 | 16 | 6.3 | 8 | 3.9 |
| 8 | 23 | 9.1 | 20 | 9.7 |
| 9 | 12 | 4.7 | 14 | 6.8 |
| 10 | 11 | 4.3 | 13 | 6.3 |
| 11 | 21 | 8.3 | 15 | 7.2 |
| 12 | 18 | 7.1 | 11 | 5.3 |
| | Total 206 | Mean 6.8 | Total 148 | Mean 6.0 |

In the case of partial covariance, 6.8 percent were equal to or greater than 2.0. For regression weights, three subgroups showed fewer large ratios than the normal curve; for partial correlation coefficients only one subgroup. A further breakdown of the distributions of ratios equal to or greater than 2.0 is shown by variable in Tables 4 and 5. Of the total number of tests made on partial correlation coefficient differences for cognitive variables, 6.1 percent of the ratios were greater than 2.0 (Table 4). (There were 22 ratios for each variable across the 12 subsamples.) For the noncognitive scales, 7.8 percent were greater than 2.0. There is overlap in these percentages, since the tabulation was made for each variable with the 22 remaining variables, assuming $\frac{D_{jk}}{\Delta_{jk}} \geqq 2.0$. One tally was made for variable j and one for variable k. The 6.1 percent shown for variable 1 includes large ratios with other cognitive tests as well as those with noncognitive scales. Among the ratios for regression coefficients (on ACB tests) 6.0 percent were greater than 2.0 for the cognitive variables and 5.9 percent were greater than 2.0 for the noncognitive scales (Table 5).

In order to examine differences in the intercorrelations separately for cognitive tests and noncognitive tests, the summary shown in Table 6 was made. Here, the large differences are divided into three categories: 1) cognitive tests versus cognitive tests; 2) noncognitive scales versus noncognitive scales; and 3) cognitive measures versus noncognitive measures. Category 1 showed 6.6 percent, category 2 showed 12 percent, and category 3 showed 5.4 percent.

## Table 4

### CORRELATION COEFFICIENT DIFFERENCE RATIOS LARGER THAN 2.0
(by variable and subgroup)

| Variables | Subsamples | | | | | | | | | | | | All Samples | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Number | Percent |
| **Cognitive** | | | | | | | | | | | | | | |
| 1 | 2 | 3 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 1 | 2 | 1 | 16 | 6.1 |
| 2 | 1 | 2 | 1 | 2 | 2 | 1 | 0 | 1 | 1 | 1 | 6 | 2 | 20 | 7.6 |
| 3 | 1 | 0 | 6 | 2 | 0 | 4 | 0 | 2 | 2 | 0 | 0 | 1 | 18 | 6.8 |
| 4 | 2 | 5 | 2 | 0 | 0 | 3 | 0 | 0 | 1 | 1 | 2 | 0 | 16 | 6.1 |
| 5 | 3 | 1 | 1 | 2 | 0 | 3 | 3 | 3 | 1 | 0 | 1 | 2 | 20 | 7.6 |
| 6 | 0 | 2 | 3 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 2 | 7 | 18 | 6.8 |
| 7 | 2 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 2.3 |
| 8 | 2 | 0 | 1 | 2 | 2 | 2 | 2 | 1 | 0 | 2 | 0 | 1 | 15 | 5.7 |
| 9 | 0 | 0 | 1 | 2 | 1 | 1 | 0 | 1 | 1 | 1 | 2 | 4 | 14 | 5.3 |
| 10 | 1 | 2 | 2 | 3 | 3 | 1 | 2 | 4 | 1 | 0 | 1 | 0 | 20 | 7.6 |
| 11 | 1 | 3 | 3 | 4 | 2 | 1 | 0 | 2 | 0 | 2 | 1 | 0 | 19 | 7.2 |
| 12 | 2 | 0 | 3 | 1 | 0 | 0 | 1 | 4 | 1 | 1 | 1 | 1 | 15 | 5.7 |
| 13 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 0 | 1 | 1 | 0 | 13 | 4.9 |
| 14 | 0 | 0 | 1 | 2 | 3 | 1 | 1 | 1 | 1 | 2 | 5 | 0 | 17 | 6.4 |
| | | | | | | | | | | | | | Tot. 227 | Avg. 6.1 |
| **Noncognitive** | | | | | | | | | | | | | | |
| 15 | 3 | 2 | 2 | 1 | 1 | 1 | 0 | 3 | 5 | 0 | 1 | 3 | 22 | 8.3 |
| 16 | 0 | 1 | 3 | 3 | 1 | 4 | 2 | 0 | 0 | 0 | 1 | 1 | 16 | 6.1 |
| 17 | 3 | 0 | 1 | 2 | 4 | 2 | 1 | 5 | 2 | 1 | 2 | 0 | 23 | 8.7 |
| 18 | 1 | 3 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 5 | 1 | 25 | 9.5 |
| 19 | 2 | 2 | 1 | 0 | 3 | 1 | 2 | 5 | 0 | 1 | 3 | 1 | 21 | 8.0 |
| 20 | 2 | 2 | 0 | 4 | 0 | 0 | 1 | 2 | 1 | 1 | 0 | 3 | 16 | 6.1 |
| 21 | 2 | 1 | 3 | 0 | 0 | 0 | 3 | 2 | 2 | 2 | 1 | 3 | 19 | 7.2 |
| 22 | 1 | 0 | 1 | 1 | 2 | 1 | 5 | 4 | 1 | 2 | 3 | 1 | 22 | 8.3 |
| 23 | 3 | 0 | 4 | 2 | 1 | 0 | 3 | 2 | 0 | 1 | 2 | 4 | 21 | 8.0 |
| | | | | | | | | | | | | | Tot. 185 | Avg. 7.8 |

Grand Total 412

Avg. for All Tests 6.8

## Table 5

REGRESSION COEFFICIENT DIFFERENCE RATIOS LARGER THAN 2.0
(by variable and subgroup)

| Variables | Subsamples 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | All Samples Number | All Samples Percent |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cognitive | | | | | | | | | | | | | | |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 1 | 0 | 7 | 6.5 |
| 2 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 6 | 5.6 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 1.9 |
| 4 | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 7 | 6.5 |
| 5 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 7 | 6.5 |
| 6 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 2 | 10 | 9.3 |
| 7 | 0 | 2 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 10 | 9.3 |
| 8 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | 7 | 6.5 |
| 9 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | 2.8 |
| 10 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 8 | 7.4 |
| 11 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 4 | 3.7 |
| 12 | 0 | 0 | 1 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 5 | 4.6 |
| 13 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 3.7 |
| 14 | 2 | 1 | 0 | 0 | 2 | 2 | 0 | 1 | 0 | 1 | 2 | 0 | 11 | 10.2 |
| | | | | | | | | | | | | | Tot. 91 | Avg. 6.0 |
| Noncognitive | | | | | | | | | | | | | | |
| 15 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 2 | 2 | 1 | 10 | 9.3 |
| 16 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | .9 |
| 17 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 9 | 8.3 |
| 18 | 0 | 3 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 2 | 1 | 10 | 9.3 |
| 19 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 1 | 7 | 6.5 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 3 | 2.8 |
| 21 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 6 | 5.6 |
| 22 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 6 | 5.6 |
| 23 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 5 | 4.6 |
| | | | | | | | | | | | | | Tot. 57 | Avg. 5.9 |

Grand Total 148

Avg. for All Tests 6.0

Table 6

NUMBER AND PERCENTAGE OF CORRELATION DIFFERENCE RATIOS GREATER THAN 2.0
(by cognitive and noncognitive variables)

| | 14 Cognitive Tests | | 9 Noncognitive Tests | | 14 Cognitive vs. 9 Noncognitive | |
|---|---|---|---|---|---|---|
| Sample | Number | Percent | Number | Percent | Number | Percent |
| 1 | 4 | 4.4 | 3 | 8.3 | 10 | 7.9 |
| 2 | 6 | 6.6 | 1 | 2.8 | 9 | 7.1 |
| 3 | 8 | 8.8 | 3 | 8.3 | 11 | 8.7 |
| 4 | 8 | 8.8 | 4 | 11.1 | 7 | 5.6 |
| 5 | 6 | 6.6 | 5 | 13.9 | 4 | 3.2 |
| 6 | 8 | 8.8 | 2 | 5.6 | 6 | 4.8 |
| 7 | 4 | 4.4 | 7 | 19.4 | 5 | 4.0 |
| 8 | 9 | 9.9 | 11 | 30.6 | 3 | 2.4 |
| 9 | 4 | 4.4 | 5 | 13.9 | 3 | 2.4 |
| 10 | 5 | 5.5 | 4 | 11.1 | 2 | 1.6 |
| 11 | 7 | 7.7 | 5 | 13.9 | 9 | 7.1 |
| 12 | 3 | 3.3 | 2 | 5.6 | 13 | 10.3 |
| Samples 1-12 | 72 | 6.6 | 52 | 12.0 | 82 | 5.4 |

Since the difference ratios for the partial correlation coefficients did not reveal the possible difference in variance for a variable in a subgroup as compared to the same variable in the pooled 11 other groups, these comparisons were made separately. The variance ratios were computed and the values compared to the F-distribution (Table 7). Again, no attempt was made to single out individually significant differences. Since nine noncognitive tests in twelve samples were being analyzed, there were 108 separate variance ratios for noncognitive variables. Of the 108, 29 (or 27 percent) were in the upper 5 percent of the central F-distribution. Similarly for the 14 cognitive tests, there were 168 separate variance ratios of which 38 (23 percent) were in the upper 5 percent of the central F-distribution.

Table 7

FREQUENCY AND PERCENTAGE OF F-TESTS FALLING IN THE UPPER 5 PERCENT OF THE CENTRAL F-DISTRIBUTION

| Sample | Frequency | Percentage |
|---|---|---|
| 1 | 11 | 47.8 |
| 2 | 7 | 30.4 |
| 3 | 3 | 13.0 |
| 4 | 6 | 26.1 |
| 5 | 2 | 8.7 |
| 6 | 5 | 21.7 |
| 7 | 7 | 30.4 |
| 8 | 6 | 26.1 |
| 9 | 2 | 8.7 |
| 10 | 5 | 21.7 |
| 11 | 5 | 21.7 |
| 12 | 8 | 34.8 |
| All Groups | 67 | 24.3 |

## IMPLICATIONS OF THE FINDINGS

Interpretation of the statistical measures presented here is not clearcut. There is some indication that the subgroups examined did not conform to the restriction-in-range model, but the evidence is not consistent enough to be decisive. Some range restriction remained after the subgroup results were equated on the ACB tests. The partial correlation coefficients and variances for the noncognitive scales seem less amenable to the extension process than those for the cognitive tests. This difference may perhaps be explained by the close relationship between the preference-experience factors used by the classification specialists and the noncognitive scales.

Since correction for restriction in range is a common procedure in studies involving MOS school samples of enlisted men for the validation of noncognitive scales as well as for the more conventional cognitive tests, the question of appropriateness of model should be examined further.

The current version of the Army Classification Battery includes two personal information and preference scales. These measures may relate to classification factors used in the assignment process but not quantified and recorded. Large amounts of test data similar to data used in the present analysis were collected on men entering the Army in 1965. The scores are now in machine-usable form and could be examined in a replication of the present study at relatively low cost. Decision on whether or not subgroup results on noncognitive scales are appropriate for extension by methods commonly used with the linearly related cognitive tests should be deferred until after such replication. The projected study should also yield further clarification of results of present extension methods for cognitive tests known to be linearly related to the criterion.

## LITERATURE CITED


1. Pearson, K. On the influence of natural selection on the variability and correlation of organs. Phil. Transcript Royal Society, Series A, 1903. 200, 1-66.

2. Aitkin, A. C. Note on selection from a multivariate normal population. Proceedings Edinburgh Mathematical Society, 1934. 4, 106-110.

3. Lawley, D. N. A note on Karl Pearson's selection formula. Proceedings Royal Society Edinburgh, Sect. A, 1943-44. 52, I, 28-30.

4. Birnbaum, Z. W., E. Paulson and F. C. Andrews. On the effect of selection performed on some coordinates of a multi-dimensional population. Psychometrika, 1950. 15, 191-204.

5. Guilford, J. P. Fundamental Statistics in Psychology and Education. 4th Ed. McGraw-Hill, New York, 1965. 341-344.

6. Gulliksen, H. Theory of Mental Tests. John Wiley and Sons, New York, 1962. 128-172.

## APPENDIX

TESTS OF THE ARMY CLASSIFICATION BATTERY

The 11 tests in the Army Classification Battery are described below. With the exception of the Radio Code Aptitude Test, all the tests in the battery are paper-and-pencil tests. Items are four-choice alternatives in VE, AR, SM, AI, ELI, and GIT. In other tests, number of choices vary. The last two tests, CI and GIT, had not been introduced when data for the present study were collected.

1. Verbal Test, VE (50 items). Each item requires the examinee to select the correct synonym for the underlined word in a short sentence.

2. Arithmetic Reasoning, AR (40 items). Each item is a reasoning problem involving application of arithmetic processes.

3. Pattern Analysis, PA (50 items). For a set of items, a two-dimensional pattern with numbered lines is presented along with the corresponding three-dimensional figure made by folding the pattern along the indicated lines. The edges of the figure are lettered. The examinee is required to identify the lettered edge of the figure corresponding to a numbered line in the pattern. The numbers in the pattern are the item numbers and the letters of the figure are used to form five-alternative responses for each item.

4. Mechanical Aptitude, MA (45 items). Each item includes a figure illustrating some physical principle and a question with two, three, or four alternative responses.

5. Army Clerical Speed, ACS. In Part I, Number Reversal (60 items), each item consists of 2 numbers. The examinee indicates whether or not the second number is exactly the reverse of the first. In Part II, Coding (50 items), there is a key in which each word is followed by a number associated with it. Each item presents a word followed by all fifteen numbers in the key. The examinee is to pick the number corresponding to the word in the key.

6. Army Radio Code, ARC. This is an auditory test recorded on tape which includes instructions to the examinees. The first part of the test is composed of 270 learning exercises designed to teach the examinee the code signals for the three letters I, N, and T. These items are presented at approximately 4 to 7 words per minute. Immediately after the learning exercises, a test of 150 items is given to measure how accurately the three code signals can be recognized at varying speeds. The first 75 items are presented at approximately 15 words per minute, and the second 75 at approximately 21 words per minute. Responses are recorded on machine scorable answer sheets presenting the three alternatives for each item.

7. Shop Mechanics, SM (40 items). Each item presents a drawing illustrating some mechanical principle or tool usage and a question.

8. Automotive Information, AI (40 items). Each item is a question about the identification or operation of automobile parts. Many of the items are based on pictures or diagrams.

9. Electronics Information, ELI (40 items). This test contains an equal number of both verbal items and picture items. The picture items require the examinee to associate pictured objects in terms of how they function electronically. The verbal items require demonstration of knowledge of electronics principles.

10.[1] Classification Inventory, CI (125 items). The test consists of self-description items in which the examinee indicates which choice most closely reflects his personal background, attitudes, self-evaluation, experiences, etc. Items are heterogeneous in content, empirically selected to predict combat effectiveness in the Korean War and rated ability to adapt to rigorous combat training and unit maneuvers in more recent combat-simulated situations.

11.[1] General Information Test, GIT (50 items). Questions cover objective items of information about various avocational pursuits to determine the degree of similarity to the knowledge patterns of effective combat men, sampled in the same situations as for the Classification Inventory.

August 1964

---

[1] These two tests were incorporated into the ACB subsequent to collection of the data analyzed in the present study.

Unclassified
Security Classification

## DOCUMENT CONTROL DATA - R & D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

| | |
|---|---|
| 1. ORIGINATING ACTIVITY (Corporate author) U. S. Army Behavioral Science Research Laboratory, OCRD, Washington, D. C. 20315 | 2a. REPORT SECURITY CLASSIFICATION Unclassified |
| | 2b. GROUP |

3. REPORT TITLE

USE OF ARMY SCHOOL SAMPLES IN ESTIMATING ACB TEST VALIDITY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Pauline T. Olson

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| August 1968 | 28 | 6 |

| | |
|---|---|
| 8a. CONTRACT OR GRANT NO. | 9a. ORIGINATOR'S REPORT NUMBER(S) Technical Research Note 199 |
| b. PROJECT NO. Army R&D PJ No. 2T061101A91B | |
| c. (ILIR) DOL | 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) |
| d. 33 | |

10. DISTRIBUTION STATEMENT

Qualified requestors may obtain copies of this report directly from DDC. Available, for sale to the public, from the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, Springfield, Virginia 22151.

| | |
|---|---|
| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY Office, Assist. Secretary of Army (R&D) Washington, D. C. |

13. ABSTRACT

Results are reported of the reexamination of the statistical method currently employed in estimating the predictive effectiveness of the tests of the Army Classification Battery (ACB) when validity data are restricted to a segment of the entering population with which the tests are used. Difference ratios were computed for partial correlation coefficients and for regression weights. Computations were based on data obtained on 2480 enlisted men at the beginning of their first enlistment. For analysis, the total sample was subdivided into 12 MOS-identified subgroups. Comparison of means and SD's of the difference ratios revealed no appreciable differences in the estimated effectiveness of the cognitive tests in the restricted sample and the total group. Somewhat larger differences between statistics for the restricted samples and the total population were found in the case of noncognitive measures. While results of the present study are inconclusive, it is anticipated that a projected replication based on more recent data will yield results for clear-cut decision on appropriateness of subgroup results on noncognitive scales for estimating validity by methods commonly used with linearly related cognitive tests.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| *Statistical methodology | | | | | | |
| *Estimating test validity | | | | | | |
| *Army Classification Battery | | | | | | |
| *Sample restriction | | | | | | |
| Psychometrics | | | | | | |
| Difference ratios | | | | | | |
| Noncognitive measures | | | | | | |
| Cognitive tests | | | | | | |
| *Statistical models | | | | | | |
| *Correlation techniques | | | | | | |
| Military psychology | | | | | | |