

283

602979

Technical Research Note 144

**FEASIBILITY OF ABBREVIATED FORMS
FOR THE ENLISTMENT SCREENING TEST**

44-p
HC \$2.00
MF \$0.50

DDC
RECEIVED
JUL 31 1964
DDC-IRA B



U. S. ARMY PERSONNEL RESEARCH OFFICE

10/1

U. S. ARMY PERSONNEL RESEARCH OFFICE

An activity of the Chief, Research and Development

Charles S. Gersoni
Colonel, GS
Commanding

J. E. Uhlener
Director, Research Laboratories

Hubert E. Brogden
Chief Scientist

Initial distribution of this report has been made by the U. S. Army Personnel Research Office. Additional copies may be obtained as follows. Department of Defense agencies and their contractors should address requests for additional copies to: Director, Defense Documentation Center, ATTN: TIOG, Bldg. No. 5, Cameron Station, Alexandria, Virginia, 22314. Other agencies and persons may obtain information on cost of microfilm or photostatic copies from: Office of Technical Services, Department of Commerce, Washington, D. C. 20230, ATTN: Technical Reports Section. Copies may also be obtained on loan from local depository libraries. A list of these libraries may be obtained from: Documents Expediting Project, Library of Congress, Washington, D. C. 20540. The mailing address of USAPRO is: U. S. Army Personnel Research Office, Washington, D. C. 20315.

Technical Research Note 144

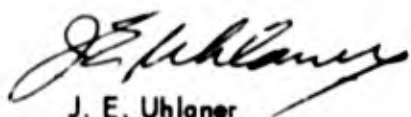
**FEASIBILITY OF ABBREVIATED FORMS FOR THE
ENLISTMENT SCREENING TEST**

Carrie Jean Waters and Emil F. Heermann

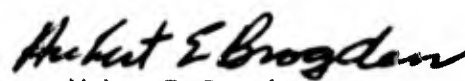
A. G. Bayroff, Task Leader

Submitted by Edmund F. Fuchs
Chief, Military Selection Research Laboratory

Approved by



J. E. Uhlener
Director, Research Laboratories



Hubert E. Brogden
Chief Scientist



Charles S. Gersoni
Colonel, GS
Commanding

Army Project Number
2J024701A722

Input Quality c-11

May 1964

USAPRO 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.

PREFACE

Congressional legislation has established the basis for procedures to screen input quality so that those who lack military trainability are rejected. The Armed Forces Qualification Test (AFQT) provides an overall screening measure. The Army Qualification Battery (AQB) provides differential measures of the individual's abilities as aids in determining qualification for service. Research to maintain these and auxiliary measures and to provide improved operational techniques for screening input is conducted by the INPUT QUALITY Task. Methods are also developed for providing estimates of the mental ability resources of the civilian pool available for military service for use in future research and manpower planning.

One objective of the INPUT QUALITY Task is to conduct methodological research basic to the preparation of future operational measures. The present publication reports on a portion of Subtask c "Methods for improving testing," FY 1963 and 1964 Work Programs, and deals with a promising technique for shortening the Enlistment Screening Test, which is used by recruiters to prescreen applicants for enlistment on their likelihood of passing the AFQT.

The entire research task is responsive to special requirements of the Deputy Chief of Staff for Personnel and the Assistant Secretary of Defense for Manpower, as well as to requirements to contribute to achievement of the objectives of DA R&D Project No. 2J024701A722, "Selection and Behavioral Evaluation."

FEASIBILITY OF ABBREVIATED FORMS FOR THE ENLISTMENT SCREENING TEST

BRIEF

Requirement:

A recent development in psychological testing points to the use of much shorter tests than those now in use for such limited objectives as estimating the likelihood that an individual will be able to qualify for military service on the Armed Forces Qualification Test (AFQT). The possibility of shortening the Enlistment Screening Test (EST)--used by recruiters to determine whether applicants for enlistment should be sent on to Armed Forces Examining Stations for further testing--was considered to merit investigation.

Procedure:

Forty 8-item tests were assembled from AFQT 7 and 8 on the basis of item data available from the standardization study. Two methods of selecting items and two methods of establishing cutting scores were applied experimentally. Effectiveness of the tests in predicting pass or fail AFQT was estimated.

Results:

The 8-item tests were highly comparable to the 48-item EST in the percentages of examinees classified the same as they were classified by AFQT. The simpler methods of selecting items and setting cutting scores proved as effective as the more elaborate and time-consuming methods.

Utilization of Findings:

The procedures evaluated will be field tested using 8-item forms.

CONTENTS

| | Page |
|-------------------------------------------------------|------|
| BACKGROUND | 1 |
| OBJECTIVE | 2 |
| RESEARCH DESIGN | 2 |
| Samples | 3 |
| Item Selection Methods | 3 |
| Length of Experimental Tests | 3 |
| Establishing Cutting Scores on the Experimental Tests | 4 |
| RESULTS | 4 |
| Characteristics of Selected Items | 4 |
| Validity of the Experimental Tests | 7 |
| Validity of Short Tests and 48-Item Test Compared | 13 |
| Categorization of Examinees with the Short Tests | 18 |
| CONCLUSIONS | 19 |
| REFERENCES | 20 |

TABLES

Page

| | | |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 1. | Mean corrected p-values of experimental tests in back and cross samples | 5 |
| 2. | Median item validity coefficients for experimental tests in back and cross samples | 6 |
| 3. | Median back- and cross-validity coefficients for experimental tests developed to predict specific points of cut on AFQT | 7 |
| 4. | Back-validity coefficients for experimental tests developed by the a priori method to predict specific points of cut on AFQT | 9 |
| 5. | Cross-validity coefficients for experimental tests developed by the a priori method to predict specific points of cut on AFQT | 10 |
| 6. | Back-validity coefficients for experimental tests developed by the test selection method to predict specific points of cut on AFQT | 11 |
| 7. | Cross-validity coefficients for experimental tests developed by the test selection method to predict specific points of cut on AFQT | 12 |
| 8. | Comparison of cross-validity coefficients for 8-item experimental tests and 48-item EST | 13 |
| 9. | Experimental tests developed by a priori method--back samples: Percents passing and failing experimental tests (ET) and percents passing and failing AFQT | 14 |
| 10. | Experimental tests developed by a priori method--cross samples: Percents passing and failing experimental tests (ET) and percents passing and failing AFQT | 15 |
| 11. | Experimental tests developed by the test selection method--back samples: Percents passing and failing experimental tests (ET) and percents passing and failing AFQT | 16 |
| 12. | Experimental tests developed by the test selection method--cross samples: Percents passing and failing experimental tests (ET) and percents passing and failing AFQT | 17 |
| 13. | Average percentages which experimental tests and the EST classified the same as AFQT at four points of cut | 18 |

FEASIBILITY OF ABBREVIATED FORMS FOR THE ENLISTMENT SCREENING TEST

BACKGROUND

The Enlistment Screening Test (EST) is used by recruiters to determine whether male applicants for enlistment should be sent on to Armed Forces Examining Stations (AFES) for further testing. The test used at AFES to determine whether applicants for enlistment--and Selective Service registrants--have sufficient military trainability to be accepted for military service is the Armed Forces Qualification Test (AFQT). Pre-screening of applicants by recruiters enables the Army to reduce travel costs, as well as housing and other costs, for men who are not likely to attain a passing score on the AFQT.

The currently operational forms 3 and 4 of the EST, modeled on the AFQT, contain verbal, arithmetic reasoning, tool functions, and spatial relations subtests. Raw scores on the EST can be converted to percentile scores as are AFQT scores. However, the practice has been to provide the recruiter with a qualifying score--in raw score form--below which he should not send an applicant on for further processing. The EST tests now in use are 48-item tests and take 60 minutes for administration.

Two lines of development indicated that the possibility of drastically shortening the Enlistment Screening Test should be explored. First, experience with the Army Qualification Battery has shown that, where the objective is to divide a restricted segment of the population into pass-fail categories, short tests can be as effective as longer tests. Since the tests of the AQB are used chiefly with marginally acceptable personnel, test content of appropriate difficulty for men of higher ability levels and men of very low level could be omitted without appreciably impairing the effectiveness of the test. Tests shortened by exclusion of the inappropriate items proved to be effective measures of abilities for the supplementary screening of men scoring from the 10th to the 30th percentile on AFQT.

The promise of the research was indicated also by developments in test theory. Cronbach and Warrington (1952) offered research evidence that multiple-choice tests built to be used with a particular qualifying score may yield higher validity coefficients if the items are of uniform difficulty rather than covering a full range of item difficulty. The validity of the narrow-range test is not appreciably enhanced by increasing the length of the test beyond a certain number of items--and this number can be predetermined. P-values for the multiple-choice items making up such tests should be such that, for individuals whose true ability is at the qualifying score, probability of passing is .50 or greater after correction for chance success. Thus, the corrected p-values for the population should correspond to, or be somewhat larger than, the percentage of the population to be selected.

OBJECTIVE

A methodological study was undertaken to determine whether effective short tests could be constructed to predict passing or failing AFQT at various score levels. The general approach was to estimate from statistics on available test items the potential usefulness of very short limited-range tests for predicting a given qualifying score--rather than to develop experimental tests for later validation. Results would be useful indications as to whether the EST could be drastically shortened and still predict the AFQT pass-fail criterion successfully.

The APRO study was designed as a basis for developing valid short tests as predictors of AFQT qualifying scores. While the Cronbach and Warrington article supported the feasibility of the research, the present study was not intended to be a test of their conclusions nor a direct application of their model. Conditions assumed in the Cronbach and Warrington article could not reasonably be met with the data available for analysis.

If short experimental tests developed by appropriate techniques should give promise of being as effective as current longer forms for their specific purpose, similar methods could be applied in developing subsequent EST forms. These would, of course, be subjected to the usual validation and standardization procedures. A successful outcome would yield an improved Army personnel management tool.

RESEARCH DESIGN

In each of four samples, ten 8-item tests were developed. Items for the experimental tests were selected from items comprising the standardization forms of AFQT 7 and 8. Item statistics and operational AFQT 5 and 6 scores were available for the standardization sample. Two different methods of selecting items for the hypothetical tests and two different procedures for setting the qualifying scores were tried out. Results obtained with the 40 experimental tests and with the EST were compared for effectiveness in predicting passing or failing AFQT.

For five of the ten tests developed in each sample, items were selected by an a priori method on the basis of p-values and correlation with AFQT qualifying score. For the other five tests in each sample, a test selection method was used to select the set of eight items most predictive of pass-fail AFQT at a given cutting score. In both cases, the five tests were constructed to predict passing or failing AFQT at the 10th, 21st, 31st, 50th, and 65th percentile scores, respectively. These percentiles are significant for qualifying under various enlistment programs. Each experimental test was dichotomized by two methods, normative equivalence and regression. The back- and cross-validity of the tests was evaluated by phi and tetrachoric coefficients.

Samples

Analysis was based on the 2000 cases collected in the standardization of AFQT 7 and 8 (Bayroff, Morton, Anderson, and Hilligoss, 1960), 1000 for each form. The group was divided into samples of 500 examinees, each stratified to be representative of the mobilization population in general mental ability. The two samples which served for the standardization of AFQT 7 (designated Samples A7 and B7) had also been administered AFQT 5 operationally at Armed Forces Examining Stations. The samples for AFQT 8 standardization (Samples A8 and B8) had also taken the operational AFQT 6. Thus, test items selected on the basis of results in one sample could be cross-validated in a second sample, and the process repeated in reverse, using the second sample for the item selection and the first for cross-validation.

Item Selection Methods

A priori method. Four verbal (VE) items and four arithmetic reasoning (AR) items were selected. These items met two requirements: difficulty (1 minus corrected p-value) at or near a given cutting score on AFQT 5-6, and high validity coefficient (ρ_{hi}) for the cutting score on AFQT 5-6. These item indices were computed in each of the four samples. No tool functions or spatial relations items were included in the experimental tests developed by the a priori method because these items usually have lower correlation with total AFQT score than do the arithmetic reasoning and verbal items.

Test selection method. Item intercorrelations and validity coefficients were used to select the most valid set of eight items--regardless of content--for the prediction of the specified cutting score on AFQT. Again, statistical procedures were applied separately in the four samples. Items were given unit weights in scoring the tests developed by this method.

Length of Experimental Tests

As a guide in estimating the number of items to be used in the experimental tests, a theoretical evaluation of the effect of length was based on typical data from AFQT 7-8. A form of the Spearman-Brown prophecy formula was used with assumed item-criterion correlations of .60, item intercorrelations of .70, and uniform p-values. The resulting estimated validity coefficients were as follows for tests consisting of from one to ten items:

| | | | | | | | | | | |
|---------------------------------|------|------|------|------|------|------|------|------|------|------|
| Number of Items | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Estimated Validity Coefficients | .600 | .651 | .671 | .682 | .688 | .693 | .696 | .699 | .701 | .702 |

Validity increased less than .01 with inclusion of the fifth item. After inclusion of the eighth item, validity increased less than .003 with the addition of each new item. On the basis of this analysis the number of items in a test was set at eight.

Establishing Cutting Scores on the Experimental Tests

Normative equivalence method. The cutting score on the experimental test was set at the same percentile score as the AFQT cutting score the test was developed to predict. That is, for an AFQT cutting score at the 31st percentile, the cutting score on the experimental test was set at the 31st percentile for the same base population. Because of the limited raw-score range on the experimental tests (0-8), it was not always possible to cut at exactly the same point as the AFQT cut. This method of dichotomizing tests approximately balances errors of classification (false positives and false negatives).

Linear regression method. The score which was the best predictor of a given cutting score on AFQT was determined for each experimental test. The experimental test score derived from the regression equation was rounded up to the next integral value for use as the cutting score. For example, if the regression method resulted in a cutting score of 3.7, the experimental test was cut so that scores of 3 or below were classified as rejects and 4 or greater were classified as accepted.

RESULTS

Characteristics of Selected Items

Differences between mean corrected p-values and target AFQT percentile scores were smaller in the tests constructed by a priori methods than in those developed by test selection procedures. This result was expected, since the test selection procedure placed no restriction on p-values. Table 1 shows mean corrected p-values, as computed in back and cross samples, for the experimental tests developed by the two methods to predict the five cutting scores on AFQT.

In general, the test selection procedure chose items with lower validity coefficients than were obtained by the a priori method (Table 2). Since the test selection method depends upon item intercorrelation as well as on the relative value of the validity coefficients, the tests developed by this procedure may be just as valid as those developed by the a priori method.

BLANK PAGE

Table 1

MEAN CORRECTED P-VALUES^a OF EXPERIMENTAL TESTS IN BACK AND CROSS SAMPLES

| Sample | 10th %ile | | 21st %ile | | 31st %ile | | 50th %ile | | 65th %ile | |
|------------|-------------|-------------------|-------------|-------------------|-------------|-------------------|-------------|-------------------|-------------|-------------------|
| | A Priori | Test Selection | A Priori | Test Selection | A Priori | Test Selection | A Priori | Test Selection | A Priori | Test Selection |
| A7 - Back | 86 | 76 | 79 | 72 | 70 | 64 | 52 | 50 | 39 | 39 |
| A7 - Cross | 86 | 77 | 80 | 72 | 71 | 65 | 54 | 50 | 41 | 41 |
| B7 - Back | 82 | 77 | 78 | 71 | 71 | 66 | 51 | 48 | 36 | 47 |
| B7 - Cross | 81 | 78 | 78 | 71 | 70 | 66 | 50 | 46 | 34 | 45 |
| A8 - Back | 83 | 78 | 79 | 69 | 71 | 67 | 51 | 50 | 36 | 35 |
| A8 - Cross | 82 | 78 | 78 | 69 | 69 | 67 | 49 | 51 | 36 | 35 |
| B8 - Back | 81 | 78 | 77 | 68 | 69 | 56 | 48 | 49 | 36 | 34 |
| B8 - Cross | 81 | 76 | 77 | 69 | 69 | 58 | 48 | 50 | 35 | 34 |

^aDecimal points omitted.

Table 2

MEDIAN ITEM VALIDITY COEFFICIENTS^a FOR EXPERIMENTAL TESTS IN BACK AND CROSS SAMPLES

| Sample | 10th %ile | | 21st %ile | | 31st %ile | | 50th %ile | | 65th %ile | |
|------------|-------------|-------------------|-------------|-------------------|-------------|-------------------|-------------|-------------------|-------------|-------------------|
| | A Priori | Test Selection | A Priori | Test Selection | A Priori | Test Selection | A Priori | Test Selection | A Priori | Test Selection |
| A7 - Back | 36 | 32 | 52 | 42 | 52 | 45 | 39 | 36 | 32 | 34 |
| A7 - Cross | 37 | 30 | 48 | 45 | 52 | 46 | 42 | 36 | 30 | 32 |
| B7 - Back | 42 | 46 | 49 | 46 | 49 | 49 | 42 | 41 | 36 | 38 |
| B7 - Cross | 40 | 36 | 50 | 45 | 48 | 48 | 39 | 44 | 37 | 36 |
| A8 - Back | 40 | 38 | 49 | 50 | 47 | 52 | 41 | 42 | 30 | 38 |
| A8 - Cross | 38 | 34 | 46 | 44 | 54 | 48 | 42 | 38 | 35 | 39 |
| B8 - Back | 41 | 25 | 43 | 54 | 53 | 48 | 42 | 38 | 36 | 34 |
| B8 - Cross | 41 | 30 | 48 | 46 | 48 | 48 | 42 | 38 | 38 | 35 |

^aDecimal points omitted.

BLANK PAGE

Validity of the Experimental Tests

The validity coefficients, both phi and tetrachoric, in back and cross samples were substantial (Table 3). Coefficients were largest for experimental tests developed for the 31st percentile on AFQT, and dropped off for tests developed for cutting scores above and below--a result attributable to the fact that the AFQT has proportionately more items around the 31st percentile level.

Table 3

MEDIAN BACK- AND CROSS-VALIDITY COEFFICIENTS^a FOR EXPERIMENTAL TESTS DEVELOPED TO PREDICT SPECIFIC POINTS OF CUT ON AFQT

| Cutting Score | A Priori | | | | Test Selection | | | |
|---------------|----------|-----------|--------|-----------|----------------|-----------|--------|-----------|
| | Back | | Cross | | Back | | Cross | |
| | ϕ | r_{tet} | ϕ | r_{tet} | ϕ | r_{tet} | ϕ | r_{tet} |
| 10th %ile | 53 | 84 | 54 | 82 | 59 | 88 | 53 | 84 |
| 21st %ile | 68 | 91 | 68 | 91 | 72 | 92 | 66 | 90 |
| 31st %ile | 71 | 93 | 70 | 92 | 77 | 94 | 71 | 92 |
| 50th %ile | 62 | 83 | 64 | 84 | 64 | 88 | 59 | 81 |
| 65th %ile | 50 | 72 | 54 | 76 | 58 | 82 | 54 | 77 |

^a Decimal points omitted.

Test construction methods. For tests developed by the a priori method, validity coefficients showed little loss in the cross samples, and in several cases a gain. This result may reflect the use of the most stable item types of the AFQT (verbal and arithmetic reasoning) and stable item statistics (primary dependence on p-values) as the basis for the a priori approach.

In the back samples, tests developed by the test selection procedure had higher validity coefficients than did tests developed by the a priori method, regardless of which method of setting cutting scores was used (Tables 4 and 6). In the cross samples (Tables 5 and 7), the a priori tests held up better than did those developed by test selection, so that the cross validity of the a priori tests was as high as that of the tests resulting from the test selection procedure. For a given underlying relationship between two variables, the maximum value of phi varies as a function of the cuts on the two variables. For this reason, tetrachoric coefficients, which provide estimates of the underlying relationship between two variables, were also obtained (Chesire, Saffir, and Thurstone tables).

In general, the test selection approach did not increase test validity over that attained in cross validation when the a priori method was used.

Cutting score method. The phi's in both back and cross samples for the normative equivalence method tended to be slightly higher than those for the regression method. The value of phi varies as a function of the cut on the two variables as well as of the relationship between the variables; as the cut points become increasingly different, the value of phi decreases. Since the normative equivalence method cuts the experimental test as near the AFQT cut as possible, the somewhat higher phi's for this method could be anticipated. As would be expected, no consistent differences appeared between validity coefficients obtained by the two methods when the comparison was based upon tetrachoric coefficients. It is understood, of course, that the magnitude of the correlation coefficients is not considered to be a test of the adequacy of the methods for establishing cut scores.

The normative equivalence and regression methods were expected to yield differing proportions of false positives and false negatives, with the regression method providing a higher percentage of false positives, when the cut scores are below the mean. In theory, total errors should be greater for the normative equivalence method. However, the findings indicated no difference in most instances and, where differences were found, the regression method usually appeared to be favored (Tables 9 to 12). Choice between the two methods depends therefore upon the relative value management places on the two types of error, since it is only here that real differences appear. With a high percentage of false positives, travel costs are increased; with a high percentage of false negatives, usable personnel are lost to the armed services.

BLANK PAGE

Table 4

BACK-VALIDITY COEFFICIENTS^a FOR EXPERIMENTAL TESTS DEVELOPED BY
THE A PRIORI METHOD TO PREDICT SPECIFIC POINTS OF CUT ON AFQT

| Cutting Method | Sample | 10th %ile | | 21st %ile | | 31st %ile | | 50th %ile | | 65th %ile | |
|-----------------------|--------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|
| | | Ø | r _{tet} | Ø | r _{tet} | Ø | r _{tet} | Ø | r _{tet} | Ø | r _{tet} |
| Normative equivalence | A7 | 54 | 84 | 69 | 92 | 75 | 94 | 61 | 83 | 50 | 72 |
| Regression | A7 | 48 | 75 | 69 | 92 | 70 | 93 | 61 | 83 | 50 | 72 |
| Normative equivalence | B7 | 66 | 90 | 68 | 88 | 73 | 94 | 61 | 82 | 50 | 73 |
| Regression | B7 | 66 | 90 | 67 | 92 | 71 | 95 | 61 | 82 | 50 | 73 |
| Normative equivalence | A8 | 56 | 87 | 68 | 90 | 68 | 90 | 64 | 83 | 48 | 69 |
| Regression | A8 | 52 | 82 | 65 | 92 | 60 | 86 | 64 | 83 | 48 | 69 |
| Normative equivalence | B8 | 52 | 84 | 70 | 90 | 74 | 93 | 63 | 85 | 54 | 76 |
| Regression | B8 | 52 | 84 | 64 | 89 | 71 | 93 | 64 | 87 | 54 | 76 |

^aDecimal points omitted.

Table 5

CROSS-VALIDITY COEFFICIENTS^a FOR EXPERIMENTAL TESTS DEVELOPED BY
THE A PRIORI METHOD TO PREDICT SPECIFIC POINTS OF CUT ON AFQT

| Cutting Method | Sample | 10th %ile | | 21st %ile | | 31st %ile | | 50th %ile | | 65th %ile | |
|-----------------------|--------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|
| | | Ø | r _{tet} | Ø | r _{tet} | Ø | r _{tet} | Ø | r _{tet} | Ø | r _{tet} |
| Normative equivalence | A7 | 50 | 83 | 68 | 90 | 76 | 93 | 65 | 84 | 51 | 73 |
| Regression | A7 | 36 | 62 | 66 | 91 | 70 | 91 | 65 | 84 | 51 | 73 |
| Normative equivalence | B7 | 54 | 80 | 70 | 91 | 74 | 92 | 59 | 80 | 52 | 76 |
| Regression | B7 | 54 | 80 | 68 | 92 | 70 | 92 | 59 | 80 | 52 | 76 |
| Normative equivalence | A8 | 58 | 85 | 68 | 91 | 74 | 92 | 64 | 84 | 56 | 76 |
| Regression | A8 | 52 | 80 | 64 | 87 | 68 | 89 | 64 | 84 | 56 | 76 |
| Normative equivalence | B8 | 55 | 86 | 71 | 93 | 68 | 89 | 66 | 86 | 59 | 80 |
| Regression | B8 | 55 | 86 | 66 | 90 | 67 | 89 | 63 | 86 | 59 | 80 |

^aDecimal points omitted.

Table 6

BACK-VALIDITY COEFFICIENTS^a FOR EXPERIMENTAL TESTS DEVELOPED BY THE
TEST SELECTION METHOD TO PREDICT SPECIFIC POINTS OF CUT ON AFQT

| Cutting Method | Sample | 10th %ile | | 21st %ile | | 31st %ile | | 50th %ile | | 65th %ile | |
|-----------------------|--------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|
| | | \emptyset | r_{tet} | \emptyset | r_{tet} | \emptyset | r_{tet} | \emptyset | r_{tet} | \emptyset | r_{tet} |
| Normative equivalence | A7 | 57 | 88 | 68 | 90 | 75 | 94 | 57 | 80 | 50 | 74 |
| Regression | A7 | 38 | 75 | 68 | 90 | 77 | 95 | 57 | 80 | 50 | 74 |
| Normative equivalence | B7 | 73 | 92 | 78 | 94 | 77 | 93 | 68 | 89 | 66 | 86 |
| Regression | B7 | 73 | 92 | 78 | 94 | 79 | 98 | 68 | 88 | 66 | 86 |
| Normative equivalence | A8 | 61 | 88 | 74 | 94 | 79 | 96 | 71 | 91 | 64 | 86 |
| Regression | A8 | 61 | 88 | 74 | 94 | 79 | 98 | 71 | 91 | 60 | 84 |
| Normative equivalence | B8 | 54 | 80 | 71 | 90 | 72 | 92 | 59 | 82 | 52 | 73 |
| Regression | B8 | 31 | 62 | 66 | 90 | 72 | 94 | 61 | 87 | 55 | 79 |

^aDecimal points omitted.

Table 7

CROSS-VALIDITY COEFFICIENTS^a FOR EXPERIMENTAL TESTS DEVELOPED BY THE
TEST SELECTION METHOD TO PREDICT SPECIFIC POINTS OF CUT ON AFQT

| Cutting Method | Sample | 10th %ile | | 21st %ile | | 31st %ile | | 50th %ile | | 65th %ile | |
|-----------------------|--------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|
| | | \emptyset | r_{tet} | \emptyset | r_{tet} | \emptyset | r_{tet} | \emptyset | r_{tet} | \emptyset | r_{tet} |
| Normative equivalence | A7 | 58 | 88 | 66 | 91 | 71 | 91 | 58 | 80 | 47 | 66 |
| Regression | A7 | 57 | 88 | 66 | 91 | 72 | 92 | 58 | 80 | 47 | 66 |
| Normative equivalence | B7 | 52 | 84 | 66 | 90 | 76 | 92 | 61 | 84 | 54 | 77 |
| Regression | B7 | 52 | 84 | 66 | 90 | 75 | 94 | 64 | 85 | 54 | 77 |
| Normative equivalence | A8 | 54 | 84 | 68 | 90 | 71 | 91 | 59 | 82 | 59 | 82 |
| Regression | A8 | 54 | 84 | 68 | 90 | 71 | 93 | 59 | 82 | 61 | 83 |
| Normative equivalence | B8 | 46 | 74 | 68 | 89 | 65 | 86 | 60 | 80 | 55 | 77 |
| Regression | B8 | 52 | 83 | 65 | 87 | 67 | 89 | 56 | 80 | 52 | 74 |

^aDecimal points omitted.

BLANK PAGE

Validity of Short Tests and 48-Item Test Compared

Cross validity coefficients (phi's and tetrachoric's) of the 48-item EST, Forms 3 and 4, for four cutting scores on AFQT 5-C were computed from data available from an earlier study (Bayroff, Thomas, and Kehr, 1959). Data were not available for the 50th percentile cutting score. To permit comparison, the median of the cross-validity coefficients of the appropriate 8-item tests (reported in Tables 5 and 7) was computed for each of the four AFQT cutting scores. Cross-validity coefficients for the EST and median cross-validity coefficients for the experimental tests are shown in Table 8. While the validity coefficients obtained for the EST ranged from .91 to .97 higher than the median coefficients obtained for the 8-item tests, the much shorter experimental tests compared favorably with the EST in predicting pass-fail on AFQT.

Table 8
COMPARISON OF CROSS-VALIDITY COEFFICIENTS^a FOR
8-ITEM EXPERIMENTAL TESTS AND 48-ITEM EST

| Cutting Score | ϕ | | r_{tet} | |
|------------------|---------|----------------------------|-----------|-------------------------------|
| | 48-Item | 8-Item (Median ϕ) | 48-Item | 8-Item (Median r_{tet}) |
| 10th %ile | 60 | 54 | 87 | 84 |
| 21st %ile | 72 | 67 | 92 | 90 |
| 31st %ile | 74 | 71 | 93 | 92 |
| 65th %ile | 59 | 54 | 82 | 76 |

^aDecimal points omitted.

Table 9

EXPERIMENTAL TESTS DEVELOPED BY A PRIORI METHOD--BACK SAMPLES:
 PERCENTS PASSING AND FAILING EXPERIMENTAL TESTS (ET) AND
 PERCENTS PASSING AND FAILING AFQT^a

| Cutting Method | Sample | 10th %ile | | | | 21st %ile | | | | 31st %ile | | | | 50th %ile | | | | 65th %ile | | | |
|-----------------------|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|----|----|
| | | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | | |
| Normative equivalence | A7 | 88 | 05 | 04 | 03 | 76 | 15 | 05 | 04 | 66 | 24 | 06 | 04 | 41 | 39 | 11 | 09 | 25 | 52 | 12 | 11 |
| Regression | A7 | 89 | 03 | 06 | 02 | 78 | 12 | 08 | 02 | 68 | 20 | 10 | 02 | 41 | 39 | 11 | 09 | 25 | 52 | 12 | 11 |
| Normative equivalence | B7 | 88 | 06 | 03 | 03 | 76 | 14 | 06 | 04 | 65 | 24 | 06 | 05 | 40 | 40 | 10 | 10 | 23 | 55 | 09 | 13 |
| Regression | B7 | 88 | 06 | 03 | 03 | 78 | 12 | 08 | 02 | 69 | 20 | 10 | 01 | 40 | 40 | 10 | 10 | 23 | 55 | 09 | 13 |
| Normative equivalence | A8 | 87 | 06 | 03 | 04 | 76 | 14 | 06 | 04 | 65 | 22 | 08 | 05 | 42 | 40 | 10 | 08 | 22 | 54 | 10 | 14 |
| Regression | A8 | 89 | 04 | 05 | 02 | 78 | 12 | 08 | 02 | 67 | 17 | 13 | 03 | 42 | 40 | 10 | 08 | 22 | 54 | 10 | 14 |
| Normative equivalence | B8 | 88 | 05 | 04 | 03 | 73 | 16 | 04 | 07 | 65 | 24 | 06 | 05 | 40 | 42 | 08 | 10 | 25 | 54 | 10 | 11 |
| Regression | B8 | 88 | 05 | 04 | 03 | 77 | 12 | 08 | 03 | 68 | 20 | 10 | 02 | 46 | 36 | 14 | 04 | 25 | 54 | 10 | 11 |

^aTotal percent passing ET may be obtained by adding the first and third columns for each ET. Percent classified the same (pass both plus fail both) may be obtained by adding the first and second columns for each ET.

Table 10

EXPERIMENTAL TESTS DEVELOPED BY A PRIORI METHOD--CROSS SAMPLES:
 PERCENTS PASSING AND FAILING EXPERIMENTAL TESTS (ET)
 AND PERCENTS PASSING AND FAILING AFQT^a

| Cutting Method | Sample | 10th %ile | | | | 21st %ile | | | | 31st %ile | | | | 50th %ile | | | | 65th %ile | | | |
|-----------------------|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------|----|----|----|
| | | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | | | | |
| Normative equivalence | A7 | 88 | 05 | 04 | 03 | 76 | 14 | 06 | 04 | 66 | 24 | 06 | 04 | 42 | 40 | 10 | 08 | 26 | 51 | 12 | 11 |
| Regression | A7 | 89 | 02 | 07 | 02 | 78 | 12 | 08 | 02 | 68 | 19 | 11 | 02 | 42 | 40 | 10 | 08 | 26 | 51 | 12 | 11 |
| Normative equivalence | B7 | 87 | 05 | 04 | 04 | 74 | 16 | 04 | 06 | 65 | 24 | 06 | 05 | 39 | 40 | 10 | 11 | 23 | 56 | 08 | 13 |
| Regression | B7 | 87 | 05 | 04 | 04 | 78 | 13 | 07 | 02 | 68 | 20 | 10 | 02 | 39 | 40 | 10 | 11 | 23 | 56 | 08 | 13 |
| Normative equivalence | A8 | 86 | 06 | 03 | 05 | 75 | 15 | 05 | 05 | 65 | 24 | 06 | 05 | 40 | 42 | 08 | 10 | 24 | 55 | 09 | 12 |
| Regression | A8 | 89 | 04 | 04 | 03 | 77 | 12 | 08 | 03 | 67 | 20 | 10 | 03 | 40 | 42 | 08 | 10 | 24 | 55 | 09 | 12 |
| Normative equivalence | B8 | 88 | 05 | 05 | 02 | 74 | 17 | 03 | 06 | 64 | 23 | 07 | 06 | 41 | 42 | 08 | 09 | 25 | 56 | 08 | 11 |
| Regression | B8 | 88 | 05 | 05 | 02 | 77 | 13 | 07 | 03 | 67 | 20 | 10 | 03 | 46 | 35 | 15 | 04 | 25 | 56 | 08 | 11 |

^aTotal percent passing ET may be obtained by adding the first and third columns for each ET. Percent classified the same (pass both plus fail both) may be obtained by adding the first and second columns for each ET.

- 15 -

Table 11

EXPERIMENTAL TESTS DEVELOPED BY THE TEST SELECTION METHOD--BACK SAMPLES:
 PERCENTS PASSING AND FAILING EXPERIMENTAL TESTS (ET)
 AND PERCENTS PASSING AND FAILING AFQT^a

| Cutting Method | Sample | 10th %ile | | | | 21st %ile | | | | 31st %ile | | | | 50th %ile | | | | 65th %ile | | | |
|-----------------------|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------|----|----|----|
| | | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | | | | |
| Normative equivalence | A7 | 85 | 07 | 02 | 06 | 77 | 13 | 07 | 03 | 64 | 26 | 04 | 06 | 38 | 41 | 09 | 12 | 24 | 53 | 11 | 12 |
| Regression | A7 | 89 | 03 | 06 | 02 | 77 | 13 | 07 | 03 | 68 | 22 | 08 | 02 | 38 | 41 | 09 | 12 | 24 | 53 | 11 | 12 |
| Normative equivalence | B7 | 89 | 06 | 03 | 02 | 78 | 15 | 05 | 02 | 64 | 26 | 04 | 06 | 39 | 45 | 05 | 11 | 30 | 53 | 11 | 06 |
| Regression | B7 | 89 | 06 | 03 | 02 | 78 | 15 | 05 | 02 | 69 | 22 | 08 | 01 | 47 | 36 | 14 | 03 | 30 | 53 | 11 | 06 |
| Normative equivalence | A8 | 89 | 05 | 04 | 02 | 76 | 16 | 04 | 04 | 64 | 27 | 03 | 06 | 41 | 45 | 05 | 09 | 30 | 53 | 11 | 06 |
| Regression | A8 | 89 | 05 | 04 | 02 | 76 | 16 | 04 | 04 | 69 | 22 | 08 | 01 | 41 | 45 | 05 | 09 | 23 | 59 | 05 | 13 |
| Normative equivalence | B8 | 87 | 05 | 04 | 04 | 74 | 16 | 04 | 06 | 63 | 25 | 05 | 07 | 37 | 43 | 07 | 13 | 27 | 50 | 14 | 09 |
| Regression | B8 | 89 | 02 | 07 | 02 | 77 | 13 | 07 | 03 | 68 | 21 | 09 | 02 | 45 | 35 | 15 | 05 | 22 | 58 | 06 | 14 |

^aTotal percent passing ET may be obtained by adding the first and third columns for each ET. Percent classified the same (pass both plus fail both) may be obtained by adding the first and second columns for each ET.

Table 12

EXPERIMENTAL TESTS DEVELOPED BY THE TEST SELECTION METHOD--CROSS SAMPLES:
 PERCENTS PASSING AND FAILING EXPERIMENTAL TESTS (ET)
 AND PERCENTS PASSING AND FAILING AFQT^a

| Cutting Method | Sample | 10th %ile | | | | 21st %ile | | | | 31st %ile | | | | 50th %ile | | | | 65th %ile | | | |
|-----------------------|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|----|----|
| | | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | Pass ET and AFQT | Fail ET and AFQT | | |
| Normative equivalence | A7 | 85 | 07 | 02 | 06 | 78 | 12 | 08 | 02 | 63 | 25 | 05 | 07 | 39 | 40 | 10 | 11 | 23 | 52 | 12 | 13 |
| Regression | A7 | 89 | 05 | 04 | 02 | 78 | 12 | 08 | 02 | 68 | 20 | 10 | 02 | 39 | 40 | 10 | 11 | 23 | 52 | 12 | 13 |
| Normative equivalence | B7 | 88 | 05 | 04 | 03 | 76 | 14 | 06 | 04 | 63 | 26 | 04 | 07 | 35 | 45 | 05 | 15 | 28 | 50 | 14 | 08 |
| Regression | B7 | 88 | 05 | 04 | 03 | 76 | 14 | 06 | 04 | 68 | 22 | 08 | 02 | 45 | 36 | 14 | 05 | 28 | 50 | 14 | 08 |
| Normative equivalence | A8 | 88 | 05 | 04 | 03 | 76 | 14 | 06 | 04 | 63 | 25 | 05 | 07 | 39 | 41 | 09 | 11 | 30 | 50 | 14 | 06 |
| Regression | A8 | 88 | 05 | 04 | 03 | 76 | 14 | 06 | 04 | 68 | 21 | 09 | 02 | 39 | 41 | 09 | 11 | 23 | 59 | 05 | 13 |
| Normative equivalence | B8 | 85 | 05 | 04 | 06 | 74 | 15 | 05 | 06 | 61 | 24 | 06 | 09 | 36 | 43 | 07 | 14 | 28 | 50 | 14 | 08 |
| Regression | B8 | 89 | 04 | 05 | 02 | 77 | 12 | 08 | 03 | 67 | 20 | 10 | 03 | 45 | 32 | 18 | 05 | 20 | 58 | 06 | 16 |

^aTotal percent passing ET may be obtained by adding the first and third columns for each ET. Percent classified the same (pass both plus fail both) may be obtained by adding the first and second columns for each ET.

Categorization of Examinees with the Short Tests

Since the EST is used to predict passing or failing AFQT at specified cutting scores, correlational validity data were supplemented by data on the extent to which men were classified in the same way by the experimental tests and by AFQT. The experimental tests were also compared with the operational EST in this respect. For experimental tests developed by the a priori method, the percentages passing; both the experimental test and AFQT, failing both tests, passing the experimental test but failing AFQT, and failing the experimental test but passing AFQT are shown in Table 9 (back samples) and Table 10 (cross samples). Tables 11 and 12 provide similar information for tests developed by test selection.

In the cross samples, the percentages classified the same by the experimental test and the AFQT ranged from 85 to 94 for experimental tests built to predict pass-fail at the 10th, 21st, and 31st percentile cuts on AFQT. Experimental tests did less well (75 - 83% classified the same) for the 50th and 65th percentile cuts on AFQT--levels of difficulty at which AFQT has proportionately fewer items. No consistent differences in correct classification were found between the normative equivalence and regression methods of establishing cutting scores. However, the limited range of scores (0 - 8) on the experimental tests precluded any very sensitive comparison between the two methods.

When the experimental tests were compared with the EST in respect to percentages classified the same as the AFQT, the short tests did as well as the 48-item EST. Table 13 presents this comparison for the four AFQT cutting scores for which EST data were available (Bayroff, Thomas, and Kehr, 1959). Data on the experimental tests are from cross samples.

Table 13

AVERAGE PERCENTAGES WHICH EXPERIMENTAL TESTS AND THE EST
CLASSIFIED THE SAME AS AFQT AT FOUR POINTS OF CUT

| | AFQT Cutting Score | | | |
|--------------------------|--------------------|-----------|-----------|-----------|
| | 10th %ile | 21st %ile | 31st %ile | 65th %ile |
| 8-Item Experimental Test | 92 | 90 | 88 | 78 |
| 48-Item EST | 92 | 89 | 88 | 82 |

CONCLUSIONS

Results tended to support the feasibility of using specially constructed, very short tests to categorize examinees as passing or failing the AFQT. Specifically, the effectiveness of the hypothetical tests constructed in the present study was considered to justify continued efforts to develop very short tests to serve as subsequent forms of the Enlistment Screening Test.

However, the tests analyzed here were constructed using item data available from administration of the much longer tests (standardization forms of AFQT 7 and 8) in which the items appeared. Final determination of the effectiveness of 8-item tests should be made after independent administration in the form in which the tests would be used operationally. If, under these circumstances, the tests are found to be adequately predictive of passing or failing AFQT at specified qualifying scores, the development of 8-item tests to serve as operational EST forms will be undertaken.

REFERENCES

Selected Publications of the US Army Personnel Research Office Dealing with INPUT QUALITY Research

- Bayroff, A. G. Methods for improving enlisted input--Status Report, 30 June 1962. Technical Research Report 1125. June 1962.
- Bayroff, A. G. Successive AFQT forms--comparisons and evaluations. Technical Research Note 132. May 1963.
- Bayroff, A. G. and Anderson, A. A. Development of Armed Forces Qualification Test 7 and 8. Technical Research Report 1132. May 1963.
- Bayroff, A. G., Heermann, E. F., and Anderson, A. A. Screening devices for selective service registrants who fail AFQT 7 and 8. Technical Research Report 1130. January 1963.
- Bayroff, A. G., Morton, M. A., Anderson, A. A., and Hilligoss, R. E. Item analysis and selection of items for standardization forms of AFQT 7 and 8. Research Memorandum 60-10. April 1960.
- Bayroff, A. G., Seeley, L. C., and Anderson, A. A. Development of the Army Qualification Battery, AQB-1. Technical Research Report 1117. October 1959.
- Bayroff, A. G., Seeley, L. C., and Anderson, A. A. Relationship of AFQT to rated training performance. Technical Research Note 106. February 1960.
- Bayroff, A. G., Thomas, J. A., and Kehr, Carol J. Evaluation of EST for predicting AFQT performance. Technical Research Report 1114. February 1959.
- Bolanovich, D. J., Harper, B., Birnbaum, A. H., and Lovelace, N. R. Development and standardization of short forms of Armed Forces Qualification Test. Technical Research Report 934. April 1952.
- Bolanovich, D. J., Mundy, J. P., Burke, Laverne K., Falk, Gloria H., and Marks, M. R. Development of the Armed Forces Qualification Test, Forms 3 and 4. Technical Research Report 1078. October 1953.
- Morton, Mary A., Goldstein, L. G., Houston, T. J., and Bayroff, A. G. Predicting proficiency of enlisted men of limited ability. Technical Research Report 1099. February 1957.

Morton, Mary A., Houston, T. J., and Bayroff, A. G. Development of Enlistment Screening Test, Forms 3 and 4. Technical Research Report 1102. May 1957.

Mundy, J. P., Goldstein, L. G., and Bayroff, A. G. Development of the Armed Forces Qualification Test, Forms 5 and 6. Technical Research Report 1101. April 1957.

Mundy, J. P., Meyer, L. A., Leedy, H. B., Sternberg, J., Goldstein, L. G., and Bayroff, A. G. Development of instruments for screening AFQT-5 and -6 failures. Technical Research Report 1104. May 1957.

Seeley, L. C. and Anderson, A. A. Development of the Army Qualification Battery, Forms 2 and 3. Technical Research Report 1131. May 1963.

Seeley, L. C., Morton, Mary A., and Anderson, A. A. Exploratory study of a sequential item test. Technical Research Note 129. December 1962.

Uhlener, J. E. Development of Armed Forces Qualification Test and predecessor Army screening tests, 1946-1950. (3d printing) Technical Research Report 976. November 1952.

Additional Reference

Cronbach, L. J. and Warrington, W. G. Efficiency of multiple choice tests as a function of spread of item difficulties. *Psychometrika*, 1952, 17, 127-147.