DOCUMENT RESUME

ED 228 293 TM 830 213

AUTHOR Mathews, John J.; Roach, Bennie W.
TITLE Reading Abilities Tests: Development and Norming for Air Force Use.
INSTITUTION Air Force Human Resources Lab., Brooks AFB, Tex. Manpower and Personnel Div.
REPORT NO AFHRL-TR-82-26
PUB DATE Feb 83
NOTE 22p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Equated Scores; *Item Analysis; Literacy; *Occupational Tests; Reading Ability; Reading Achievement; Reading Comprehension; *Reading Tests; *Test Construction; *Test Norms; Test Reliability
IDENTIFIERS *Air Force Reading Abilities Test; Armed Services Vocational Aptitude Battery; Commercially Prepared Materials; *Parallel Test Forms; Synonyms

ABSTRACT

This report describes the development and norming of parallel forms of the Air Force Reading Abilities Test (AFRAT) which have been designed for Air Force use as replacements for commercial reading tests. A previous study on service applicants found considerably divergent reading grade levels (RGL's) from different commercial tests for subjects of the same Armed Services Vocational Aptitude Battery ability level. In addition to varying norms, the use of commercial tests has several other drawbacks, including high testing material costs and RGL norms of unknown appropriateness for military personnel. The goal of this effort was to develop reading tests with appropriate norms. Samples of airmen (n=12,938) were experimentally administered reading tests, including one or two forms of the AFRAT. After editing, analyses were done to determine the equivalence of the AFRAT forms, their relationships to other tests, AFRAT raw score to RGL equivalents, and training grade validity of AFRAT item types. The two parallel forms of the AFRAT appear to meet administrative and psychometric specifications. AFRAT appears to be a highly reliable instrument and is recommended as a replacement for commercial reading tests. (Author/PN)

AIR FORCE

HUMAN RESOURCES

LABORATORY

READING ABILITIES TESTS: DEVELOPMENT AND NORMING FOR AIR FORCE USE

By

John J. Mathews
Bennie W. Roach

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235

February 1983

Interim Report for Period 1 October 1980 – 31 December 1981

Approved for public release; distribution unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

NANCY GUINN, Technical Director
Manpower and Personnel Division

RONALD W. TERRY, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1 REPORT NUMBER
AFHRL-TR-82-26

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
READING ABILITIES TESTS: DEVELOPMENT AND NORMING FOR AIR FORCE USE

5. TYPE OF REPORT & PERIOD COVERED
Interim
1 October 1980 – 31 December 1981

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
John J. Mathews
Bennie W. Roach

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
62703F
77191808

11. CONTROLLING OFFICE NAME AND ADDRESS
HQ Air Force Human Resources Laboratory (AFSC)
Brooks Air Force Base, Texas 78235

12. REPORT DATE
February 1983

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS (of this report)
Unclassified

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
TS Study Numbers 7179
7836

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Air Force Reading Abilities Test (AFRAT)	reading abilities
Armed Services Vocational Aptitude Battery (ASVAB)	reading achievement
item analysis	reading comprehension
literacy skills	synonyms
personnel testing	test equating

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the development and norming of parallel forms of the Air Force Reading Abilities Test (AFRAT) which have been designed for Air Force use as replacements for commercial reading tests. Many Air Force organizations have been administering various commercially published reading tests to military personnel. These tests have been used for assignment to remedial training programs, as aids in counseling students, or for description of reading levels of airmen in various occupational specialties. A previous study on service applicants found considerably divergent reading grade levels (RGLs) from different commercial tests for subjects of the same Armed Services Vocational Aptitude Battery ability level. In addition to varying norms, the use of commercial tests has several other drawbacks, including high testing material costs and RGL norms of unknown appropriateness for military personnel. The goal of this effort was to develop reading tests with appropriate norms. Samples of airmen

Item 20 (Continued):

were experimentally administered several reading tests, including one or two forms of the AFRAT. After data editing, analyses were done to determine the equivalence of the AFRAT forms, their relationships to other tests, AFRAT raw score to RGL equivalents, and training grade validity of AFRAT item types. The two parallel forms of the AFRAT appear to meet administrative and psychometric specifications. AFRAT appears to be a highly reliable instrument and is recommended as a replacement for commercial reading tests.

SUMMARY

Objective

The objective of this report is to document the development and norming of parallel forms of the Air Force Reading Abilities Test (AFRAT).

Background/Rationale

The Air Force has been administering various commercially published reading tests to military personnel. These tests have been used for assignment of personnel to remedial training programs, as aids in career counseling of students, or for description of reading levels of airmen in various occupational specialties. A previous study on service applicants found large divergence in reading grade levels (RGLs) estimated from different commercial tests. The evidence suggested that RGL standards differ considerably from one commercial test to another. In addition to varying norms, the use of commercial tests has several other drawbacks, including high testing material costs and RGL norms of unknown appropriateness for military personnel. The goal of this effort was to develop reading tests with appropriate norms.

Approach

A total of 12,938 airmen was administered two reading tests (e.g., either two forms of the AFRAT, or one AFRAT form and a commercial reading test). Analyses were computed to determine the equivalence of the AFRAT forms, their correlation with other reading tests, AFRAT raw score to RGL equivalents, and training grade validity of AFRAT item types. For establishing AFRAT-to-RGL equivalents, RGL standards were defined as the average RGL equivalent from several commercial reading tests.

Specifics

The AFRAT consists of 45 vocabulary items in a synonym format and 40 comprehension items consisting of one or several paragraphs followed by one or more questions. The comprehension items require either paraphrasing or making inferences from the passages. All items are multiple-choice with four alternatives with a total test time limit of 50 minutes.

Comparing AFRAT Forms A and B, the proportion of correct item responses was .85 for each form, and average-item-to-test-total correlations were similar. In addition, subtest and total-score variances for AFRAT Forms A and B were equal. These data indicated that the two forms were parallel.

The relationships of the two AFRAT forms to three commercial reading tests were moderate-to-high (correlations of approximately .60 to .67). The interrelation between the two AFRAT forms was somewhat higher (approximately .73). Since the sample population was restricted due to prior enlistment screening, the correlation between the AFRAT forms would be considerably higher if computed from a full-range sample.

Percentiles were computed for AFRAT scores and RGL scores derived from commercial tests. AFRAT forms were equated to an average RGL through use of the Air Force General Aptitude Index (AI) from ASVAB as an anchor test. Raw-score-to-RGL conversion tables for the 4th through the 12th RGL were generated for AFRAT subtest and total scores.

The median coefficient of correlation of AFRAT with technical training grades in 19 Air Force specialty groups was .40. In 16 of the 19 groups, this correlation coefficient was greater than .30.

Conclusions/Recommendation

Forms A and B of the AFRAT were found to be parallel. The computed percentile and RGL norms should be appropriate for enlistees. The AFRAT was found to correlate quite well (approximately .60 or higher) with three commercial reading tests. A preliminary analysis indicated that the AFRAT would be a valid predictor of performance in technical training.

It is recommended that the AFRAT replace the various commercial tests now being used as the operational test to screen enlistees for marginal or inadequate reading ability.

PREFACE

This study was completed under Task 771918, Selection and Classification Technologies, which is part of a larger effort in Force Acquisition and Distribution. It was subsumed under work unit number 77191808, "Reading-related Problems in the Air Force." This work unit was established in response to Request for Personnel Research (RPR 76-25) submitted by the Air Force Manpower and Personnel Center (AFMPC/MPCYP) Maj John Welsh, Requirements Manager—entitled "Development, Validation, and Standardization of a Reading Ability Test for Air Force Personnel."

The authors wish to express their appreciation to Tammy Hilbert and Roy Chollman of the Air Force Human Resources Laboratory for their assistance during this project.

TABLE OF CONTENTS

		Page
I.	Introduction	7
II.	Method	7
	Design Goals for AFRAT Forms	7
	Reading Measurement Instruments	7
	Samples	8
	Analytic Methods	9
III.	Results and Discussion	9
	AFRAT Internal Analyses	9
	Test Intercorrelations	11
	AFRAT Norming	12
	Technical Training Validation	17
IV.	Conclusions and Recommendations	18
	References	19

LIST OF TABLES

Table		Page
1	Test Lengths and Times for AFRAT Forms	8
2	Distribution of Difficulties (P) for AFRAT Forms A and B Items	9
3	Distribution of AFRAT Forms A and B Item-Test Correlations	10
4	Mean AFRAT Difficulties for Army Samples	10
5	AFRAT Reliabilities for Air Force Subgroups	11
6	Test Intercorrelations	11
7	Intercorrelations of Like-Named Subtests	12
8	AFRAT Forms A and B Means, and Standard Deviations (SD)	12
9	AFRAT Means and Standard Deviations (SD) for Army Samples	13
10	Equipercentile Equating of AFRAT Forms A and B for Air Force Samples	13
11	Equipercentile Equating of AFRAT Scores for Army Samples	14
12	Equipercentile Calibration of AFRAT Form A-B Average Scores to ASVAB General Composite	14
13	Equipercentile Calibration of Reading Tests to General Composite Percentiles	15
14	AFRAT Forms A and B Total Score Conversions to Reading Grade Level (RGL)	16
15	AFRAT Vocabulary and Comprehension Reading Grade Level (RGL) Conversions	17
16	AFRAT Form X Validities for Technical Training Grades	18

READING ABILITIES TESTS: DEVELOPMENT AND NORMING FOR AIR FORCE USE

I. INTRODUCTION

Many Air Force organizations have been administering various commercially published reading tests to military personnel. These tests are used for assignment of personnel to remedial reading training programs, as aids in career counseling of students, or for description of reading levels of airmen in various occupational specialties. The Tests of Adult Basic Education (TABE) (CTB/McGraw-Hill, 1976) comprise the reading test instrument most frequently used in the Air Force.

One of the problems resulting from the use of different reading tests in the Air Force is the noted variation in computed reading grade levels (RGLs) for individuals with similar levels of intellectual functioning. A study on service applicants (Mathews, Valentine, & Sellman, 1978) found considerably divergent RGLs from different commercial tests for subjects of the same Armed Services Vocational Aptitude Battery (ASVAB) ability level. In addition, results indicated that the ASVAB General (called General-Technical by some military services) composite correlated as highly with some reading tests as those reading tests correlated with each other. Based on these results, the use of ASVAB to estimate reading ability of groups was considered. However, there are some problems associated with using ASVAB composites to measure reading ability of individuals. These composites contain several short subtests covering different ability factors. The General composite includes Arithmetic Reasoning (AR), in addition to the verbal subtests of Word Knowledge (WK) and Paragraph Comprehension (PC). Most women perform slightly better than do men on verbal tests; however, they generally do somewhat less well on AR than do men. When the General composite is used to gauge reading ability of women, underestimation will result in the majority of cases. For individual measurement, therefore, a more content specific and reliable measure of reading than that based on ASVAB was desired.

The use of commercial tests has several additional drawbacks, including high testing material costs and RGL norms of unknown appropriateness for military personnel. To resolve these problems, it was decided that a reading test should be developed specifically for Air Force use. The objective of this report is to provide a description of the development and norming of the Air Force Reading Abilities Test (AFRAT) that was designed to standardize the assessment of reading ability of Air Force personnel and to replace the commercial reading tests that have been used throughout the Air Force.

II. METHOD

Design Goals for AFRAT Forms

The following general goals were pursued in developing reading tests:

1. Vocabulary and comprehension sections, as found in most commercial reading tests, were designed.

2. Comprehension passages were written with expository prose.

3. Comprehension questions covered factual matter that was unlikely to be answered correctly based solely on prior knowledge.

4. Vocabulary words were selected which might likely be encountered in a work environment. Esoteric adverbs and adjectives were avoided to keep the test from being overly academic in nature.

5. The test was designed to be as reliable as possible but to require less than one hour of testing time.

Reading Measurement Instruments

The following reading tests were used in this study.

AFRAT Form X. An experimental form of AFRAT was constructed based on available items from obsolete Air Force classification tests. This test was used to obtain initial estimations of the construct and predictive validity of the item types. Due to the limited pool of items, the difficulty of AFRAT Form X items varied considerably from very easy to very hard.

AFRAT Forms A and B. Two parallel AFRAT forms (A and B) were developed. The second form allows for retesting after remedial training. Items were selected from a pool assembled specifically to specifications for AFRAT. The AFRAT consists of vocabulary items in a synonym format and comprehension items consisting of one or several paragraphs followed by one or more questions. The comprehension items require either paraphrasing or making inferences from the passages. AFRAT Forms A and B each contain 45 vocabulary and 40 comprehension items, with a total test limit of 50 minutes (see Table 1). All items are multiple choice with four alternatives. The tests were targeted at the 8th RGL as measured by the Adult Basic Learning Examination (see the following paragraphs). Although AFRAT Forms A and B were to be peaked at a difficulty level corresponding to the 8th RGL, the desired norms would span from the 5th through the 12th RGL.

Table 1. **Test Lengths and Times for AFRAT Forms**

	AFRAT A–B		AFRAT X	
Scale	No. Items	Time (Minutes)	No. Items	Time (Minutes)
Vocabulary	45	15	50	10
Comprehension	40	35	42	25
Total	85	50	92	35

Gates-MacGinitie Reading Tests (Survey D). Included in these tests are a 50-item vocabulary section and a 42-item comprehension section, with a combined testing time of 40 minutes. The vocabulary items require the selection of synonyms for single words. The comprehension items consist of questions about single paragraphs. Vocabulary and comprehension RGLs of 1.0 to 11.9 are reported (Gates & MacGinitie, 1972).

Tests of Adult Basic Education (TABE) Level D. This instrument consists of a 50-item vocabulary section and a 45-item comprehension section, with a combined time of 50 minutes for testing. The vocabulary items question the meaning of words in phrases. The comprehension items consist of questions about passages containing one or several paragraphs. Vocabulary, comprehension, and combined RGLs of from 5.0 to 12.9 are reported (CTB/McGraw-Hill, 1976).

Adult Basic Learning Examination (ABLE) Levels I–III. This instrument includes a 50-item vocabulary section and a 58-item comprehension section taking approximately 50 minutes of testing time (this varies by level). The vocabulary items ask about the meaning of words in phrases. The comprehension items consist of questions about single paragraphs. Vocabulary, reading, and problem solving sections were used to calibrate the ASVAB General Composite to ABLE in an unpublished Army study completed in 1980. ABLE gives RGLs from 3.0 to 12.9 (Karlsen, Madden, & Gardner, 1971).

Samples

A total of 6,555 subjects tested from May to July 1981, except as noted, formed the following seven samples:

1. 625 Air Force trainees given AFRAT Forms A and B.
2. 820 Air Force trainees given ABLE II and AFRAT Form A (N = 413) or AFRAT Form B (N = 407).
3. 946 Air Force trainees given Gates-MacGinitie and AFRAT Form A (N = 454) or B (N = 492).
4. 883 Air Force trainees given AFRAT Form X and AFRAT Form A (N = 459) or AFRAT Form B (N = 424).

5. 3,274 Air Force trainees given TABE and AFRAT Form A (N = 1951, composed of subjects from samples 1 – 4) or AFRAT Form B (N = 1948, composed of subjects from samples 1 – 4 less 625 subjects given both forms).

6. 1,049 Army trainees given AFRAT Form A (N = 491) or AFRAT Form B (N = 558).

7. 2,232 Air Force trainees given AFRAT Form X in 1978.

In addition, data based on about 1,100 Army trainees given ABLE I, II, or III in 1980 and 2,033 service applicants given Gates-MacGinitie in 1978 (Mathews, Valentine, & Sellman,1978) were used in developing norms. These two tests and the TABE are widely used by the armed services.

Analytic Methods

An item analysis program (Koplyay, 1981) was used to compute the following internal AFRAT and test summary statistics: Difficulty (proportion answering each item correctly), item biserial (correlation of item with test scale), internal consistency reliability (Kuder-Richardson Formula 20), test mean, and standard deviation. Means for Army samples were adjusted in order to control for test score differences resulting from sampling fluctuations. This was accomplished by using regression equations (Guilford & Fruchter, 1978) for predicting AFRAT scores based on the relationship of AFRAT forms with the ASVAB General composite.

Construct and predictive validities of AFRAT forms were assessed through Pearson correlation coefficient (r) values, which were computed among tests. Predictive validities were obtained by correlating AFRAT Form X scores with technical training grades for subsamples. Fisher's r to z transformations were used to average the r values, across combined samples (Guilford & Fruchter, 1978). The technical training validation was only a preliminary analysis as a more comprehensive study will be done on AFRAT Forms A and B when sufficient criterion data are available.

Percentile norms were obtained for AFRAT forms, and AFRAT Forms A and B were placed on the same scale through equipercentile equating (Angoff, 1971). This same procedure was used to equate AFRAT to TABE RGLs. AFRAT Forms A and B were also equated to ABLE and Gates-MacGinitie RGL scales through the use of the ASVAB General composite as a common anchor test. This is the Angoff (1971, p. 576) Design III where all groups take the common anchor test, and each group takes one of the reading tests.

III. RESULTS AND DISCUSSION

AFRAT Internal Analyses

Table 2 gives the item difficulties for AFRAT Forms A and B based on Air Force trainees given both tests (Sample 1). These alternate forms appear to be of parallel difficulty, with fairly similar means and distributions. The bulk of the items are quite easy with means around .82 (not corrected for guessing). In comparison, the TABE items had an average difficulty of .84 for the same sample (N = 625).

Table 2. **Distribution of Difficulties (P) for AFRAT Forms A and B Items *(N = 625)***

	Vocabulary		Comprehension	
Difficulty[a]	A	B	A	B
90–99	18	16	8	15
80–89	12	11	19	15
70–79	8	9	8	5
60–69	3	5	3	2
59 and less	4	4	2	3
Total	45	45	40	40
Average P	.798	.796	.828	.833

[a]Decimal points omitted for readability.

The item-test biserial correlations (r_{bis}) are moderate-to-high for virtually all items, with means of the r values for subtests of .60 to .65 and an item r range of .29 to .89. Again, the AFRAT forms appear parallel (see Table 3).

Table 3. **Distribution of AFRAT Forms A and B Item–Test Correlations (*N* = 625)**

	Vocabulary		Comprehension	
r_{bis}[a]	A	B	A	B
70–99	9	11	13	15
50–69	28	19	16	17
30–49	8	14	11	8
29 and less	0	1	0	0
Total	45	45	40	40
Average r_{bis}	.598	.597	.613	.649

[a]Decimal points omitted for readability.

An estimate of mean AFRAT item performance for subjects equal in average ability to that of the normative population for ASVAB can be obtained from the data collected on Army trainees (Sample 6). Army samples given AFRAT had an average ASVAB General composite score of about 50 percentile, 50.6 for AFRAT A sample and 49.7 for AFRAT B sample. Mean AFRAT difficulties (P) for these subjects (sample 6) are given in Table 4. Because the lowest ability subjects are excluded from service, the distribution of scores would differ in an applicant or normative sample. The average P value was .69 for the Army samples compared to the P value of .82 for Air Force samples. Since ASVAB selection tests have P values of about .70 (Ree, Mullins, Mathews, & Massey, 1982), AFRAT seems to be comparable in mean difficulty to these tests.

Table 4. **Mean AFRAT Difficulties for Army Samples**

	AFRAT Form		
	A	B	
Scale	(N = 491)	(N = 558)	Average
Vocabulary	.69	.66	.68
Comprehension	.70	.70	.70
Total	.70	.68	.69

AFRAT internal consistency reliability coefficients are shown in Table 5 for subgroups of Air Force samples. These data are based on all female and all Black trainees and on representative subsamples of male and Caucasian trainees. The average reliabilities were .92 for AFRAT Form A and .91 for AFRAT Form B. These values are quite high considering that reliability is maximized when item variance is large (i.e., when item difficulties are moderate).

Table 5. AFRAT Reliabilities for Air Force Subgroups

Form		Subgroup			
		Black	Caucasian	Female	Male
AFRAT A	N	520	520	731	731
	Rel	.92	.92	.89	.92
AFRAT B	N	540	540	736	736
	Rel	.92	.90	.87	.94

Note. Internal consistency reliabilities (Rel) based on formula KR-20.

Reliabilities were not as high for female samples, .89 for AFRAT Form A and .87 for AFRAT Form B. This is most likely due to significantly lower score variance for women compared to men ($F = 1.6$, $p < .01$ on AFRAT Form A and $F = 2.3$, $p < .01$ on AFRAT Form B). At least two plausible explanations for the gender difference in score variance exist. First, the mean AFRAT scores were 2.5 points higher for women than men, thus restricting the range. Second, some previous studies of aptitude/achievement tests have revealed higher male variance on a number of tests (Jensen, 1980).

Test Intercorrelations

Table 6 shows the intercorrelations for tests given to Air Force subjects in sample 5. These r values have not been corrected for restriction in range from selection on the ASVAB since it is doubtful that assumptions required to make corrections can be met. Despite the attenuation, the alternate AFRAT forms correlated .73. The degree of restriction in these r values is illustrated by visually comparing the r of .57 in Table 6 between Gates-MacGinitie and ASVAB General to the r of .76 obtained between the same two measures in a study using service applicants (Mathews, Valentine, & Sellman, 1978). The average r values for other tests was .65 for AFRAT A and .63 for AFRAT B. These AFRAT forms correlated somewhat more highly with other reading tests than did the TABE. The average r values for AFRAT and TABE were .65 and .57 with Gates-MacGinitie, respectively, and .62 and .50 with ABLE, respectively. The two AFRAT forms correlated to the same degree with TABE as they did with the GM and ABLE (average r = .64) with both AFRAT and TABE.

Table 6. Test Intercorrelations
(Samples 1-4, N Values Range from 407 to 3,274)

Test	AFRAT A	AFRAT B	AFRAT X	TABE	GM	ABLE	GT
AFRAT A	1.00	.73	.63	.67	.66	.64	.63
AFRAT B		1.00	.65	.61	.64	.60	.63
AFRAT X			1.00	.56	a	a	.63
TABE				1.00	.57	.50	.60
Gates-MacGinitie					1.00	a	.57
ABLE II						1.00	.49
General (ASVAB)							1.00

[a]Due to sampling constraints these intercorrelations are unavailable.

Table 7 gives intercorrelations of similar subtests across reading tests. Among vocabulary subtests, the highest r, .68, was between the two AFRAT forms. For comprehension subtests, the r between AFRAT Forms A and B, .62, was again the highest. Correlations among comprehension tests were generally lower than the r values among vocabulary tests. This would be indicative of more unique variance within the different comprehension tests than within the different vocabulary tests.

Table 7. **Intercorrelations of Like-Named Subtests (N Values = 407 to 3,274)**

Test	AFRAT A	AFRAT B	AFRAT X	TABE	G-M	ABLE
Vocabulary						
AFRAT A	1.00	.68	.53	.57	.67	.62
AFRAT B		1.00	.52	.48	.64	.52
AFRAT X			1.00	.41	a	a
TABE				1.00	.55	.41
Gates-MacGinitie					1.00	a
ABLE 11						1.00
Comprehension						
AFRAT A	1.00	.62	.49	.50	.40	.37
AFRAT B		1.00	.53	.46	.38	.46
AFRAT X			1.00	.48	a	a
TABE				1.00	.37	.28
Gates-MacGinitie					1.00	a
ABLE 11						1.00

[a]Due to sampling constraints these intercorrelations are unavailable.

AFRAT Norming

Descriptive statistics for AFRAT Forms A and B are listed for Sample 1 in Table 8. AFRAT means and standard deviations for Army samples are given in Table 9. Adjusted means are also shown based on regression to compensate for ability differences on the ASVAB General composite. These differences noted earlier are caused by sample fluctuations. These means, 58.6 for AFRAT Form A and 58.1 for AFRAT Form B, should be representative since these samples had the same average ability as the normative population. However, as previously mentioned, the distribution of scores in the general population would differ.

Table 8. **AFRAT Forms A and B Means, and Standard Deviations (SD)** ***(N = 625)***

Test	Mean	SD
AFRAT A		
Vocabulary	35.9	5.5
Comprehension	33.1	5.8
Total	69.0	10.2
AFRAT B		
Vocabulary	35.8	5.3
Comprehension	33.3	5.3
Total	69.1	9.4

Table 9. **AFRAT Means and Standard Deviations (SD) for Army Samples (N = 491 and N = 558)**

Scale	AFRAT A		AFRAT B	
	Mean	SD	Mean	SD
Vocabulary	31.1	9.5	29.7	9.6
Comprehension	28.0	8.3	27.8	8.8
Total	59.1	16.8	57.5	17.2
Adjusted Total[a]	58.6		58.1	

[a] Adjusted via regression for ability differences on ASVAB General Composite due to sampling.

AFRAT Forms A and B were equated using the equipercentile method with Air Force samples. Because the forms appear parallel, the raw scores were combined to compute percentiles and to give a single, more stable conversion table (see Table 10). At every percentile point, raw scores for AFRAT Forms A and B are within one point of the average raw score.

Table 10. **Equipercentile Equating of AFRAT Forms A and B for Air Force Samples**

Percentile	AFRAT A Raw	AFRAT B Raw	A-B Avg[a] Raw	Percentile	AFRAT A Raw	AFRAT B Raw	A-B Avg Raw
1	40	41	40	30	66	67	66
2	45	46	46	32	67	67	67
3	48	49	48	34	68	68	68
4	49	51	50	36	69	68	68
5	51	53	52	38	69	69	69
6	52	54	53	40	70	69	70
7	54	55	54	42	70	70	70
8	55	56	55	44	71	70	71
9	56	57	56	46	71	71	71
10	57	58	57	48	72	71	72
12	58	59	59	50	72	72	72
14	59	60	60	55	73	73	73
16	60	62	61	60	75	73	74
18	61	63	62	65	76	74	75
20	62	63	62	70	77	75	76
22	63	64	63	75	78	76	77
24	64	65	64	80	79	77	78
26	65	65	65	85	80	78	79
28	66	66	66	90	81	79	80
				95	83	81	82

[a] All entries have been rounded to integer form.

The AFRAT is negatively skewed (i.e., the raw score distribution is skewed to the left), which is appropriate for a test designed to identify low-performing subjects. The AFRAT median score (50th percentile) was 72, compared to a mean of about 69 (from Table 8). A higher median than mean is characteristic of negatively skewed tests.

AFRAT percentiles for Army samples are listed in Table 11. The median score was about 62, compared to a mean of 58.

Table 11. **Equipercentile Equating of AFRAT Scores for Army Samples**

Percentile	AFRAT A – B Average[a]	Percentile	AFRAT A – B Average
1	21	30	51
2	22	32	52
3	23	34	53
4	24	36	55
5	25	38	56
6	26	40	57
7	28	42	58
8	29	44	59
9	31	46	60
10	33	48	61
12	35	50	62
14	38	55	64
16	39	60	66
18	41	65	69
20	43	70	71
22	45	75	72
24	46	80	75
26	48	85	76
28	50	90	78
		95	81

[a]All entries have been rounded to integer form.

Table 12 contains an equipercentile calibration of AFRAT scores to ASVAB General (or General-Technical) composite percentiles based on combined Air Force and Army subjects (Samples 5 and 6). The General composite is the ASVAB measure which has been found to correlate highest with reading tests (Mathews et al., 1978).

Table 12. **Equipercentile Calibration of AFRAT Form A-B Average Scores to ASVAB General Composite**

General Composite	AFRAT Raw Score
10	18
15	20
20	23
25	34
30	45
35	55
40	58
45	62
50	65
55	68
60	71
65	73
70	75
75	76
80	77
85	79
90	81
95	83

Equipercentile calibrations of other reading tests to ASVAB general percentiles are shown in Table 13. The data on ABLE and Gates-MacGinitie are based on previous studies (see "*Samples*" subsection), and the TABE data are from Air Force Sample 5 in this study.

Table 13. **Equipercentile Calibration of Reading Tests to General Composite Percentiles**

ASVAB General	ABLE[a] Grade L	TABE RGL	G-M[b] RGL	Average RGL
10	5.4	–	4.0	4.7
15	6.3	–	5.9	6.1
20	7.0	6.9	6.9	6.9
25	7.6	7.7	7.2	7.5
30	8.0	8.7	7.9	8.2
35	8.4	9.7	8.9	9.0
40	8.7	9.9	9.4	9.3
45	9.0	10.1	9.9	9.7
50	9.4	10.6	10.4	10.1
55	9.8	11.0	10.9	10.6
60	10.4	11.4	11.2	11.0
65	10.7	11.8	11.5	11.3
70	11.1	12.2	11.9[c]	11.7
75	11.5	12.5	–	12.0
80	11.7	12.8	–	12.2
85	12.0	12.9[a]	–	12.4
90	12.3	–	–	12.6
95	12.7	–	–	12.9

[a]Based on data from Army subjects tested in 1980.
[b]Based on renorming of data from 1978 study.
[c]Maximum RGL from normative tables is 11.9.

It is apparent that there are substantial differences in grade level norms among the commercial reading tests. At some specific levels, at least one grade separates each of the reading tests from another. Without substantial evidence as to which test yields the most accurate RGL conversions, a good estimate should be obtained by averaging the RGLs across the commercial tests. The column on the right side of Table 13 gives this average.

Equipercentile conversions of average RGL for each AFRAT total raw score point are shown in Table 14. Separate RGL conversions for AFRAT Vocabulary and Comprehension subscores are listed in Table 15.

Table 14. **AFRAT Forms A and B Total Score Conversions to Reading Grade Level (RGL)**

AFRAT Total	Average RGL	AFRAT Total	Average RGL
1-15	4.0	51	8.6
16	4.2	52	8.7
17	4.4	53	8.8
18	4.7	54	8.9
19	5.5	55	9.0
20	6.1	56	9.1
21	6.5	57	9.2
22	6.7	58	9.3
23	6.9	59	9.4
24	6.9	60	9.5
25	6.9	61	9.6
26	7.0	62	9.7
27	7.1	63	9.8
28	7.1	64	10.0
29	7.1	65	10.1
30	7.2	66	10.3
31	7.3	67	10.5
32	7.3	68	10.6
33	7.4	69	10.7
34	7.5	70	10.8
35	7.6	71	11.0
36	7.7	72	11.1
37	7.8	73	11.3
38	7.8	74	11.5
39	7.9	75	11.7
40	7.9	76	12.0
41	8.0	77	12.2
42	8.0	78	12.4
43	8.1	79	12.4
44	8.1	80	12.5
45	8.2	81	12.6
46	8.2	82	12.7
47	8.3	83	12.9
48	8.4	84	12.9
49	8.5	85	12.9
50	8.5		

Table 15. **AFRAT Vocabulary and Comprehension Reading Grade Level (RGL) Conversions**

Vocabulary Score	Average RGL	Comprehension Score	Average RGL
1-7	4.0	1-7	4.0
8	4.0	8	4.4
9	4.4	9	5.6
10	5.0	10	6.4
11	6.4	11	6.8
12	7.0	12	6.8
13	7.2	13	6.9
14	7.4	14	7.0
15	7.5	15	7.1
16	7.6	16	7.2
17	7.8	17	7.3
18	7.9	18	7.4
19	8.2	19	7.5
20	8.3	20	7.6
21	8.4	21	7.8
22	8.4	22	8.0
23	8.5	23	8.2
24	8.5	24	8.4
25	8.6	25	8.7
26	8.8	26	8.9
27	8.9	27	9.1
28	9.0	28	9.3
29	9.1	29	9.5
30	9.2	30	9.6
31	9.4	31	9.9
32	9.6	32	10.2
33	9.9	33	10.5
34	10.3	34	10.7
35	10.6	35	11.1
36	10.9	36	11.6
37	11.2	37	12.2
38	11.4	38	12.4
39	11.8	39	12.7
40	12.2	40	12.9
41	12.5		
42	12.6		
43	12.8		
44	12.9		
45	12.9		

Technical Training Validation

In order to get an initial estimate of the predictive validity of the item types in AFRAT, Form X was administered to approximately 3,000 airmen. Technical training grades were subsequently obtained for those in common Air Force Specialty Code (AFSC) groups. Validities for AFRAT Form X in 19 AFSC groups (total N = 2,253) are listed in Table 16. The median r with training grades was .40. Validities were generally higher for Comprehension than for Vocabulary. This is to be expected due to selection on the ASVAB General composite which has more vocabulary than reading comprehension content. This would severely restrict r value involving a

vocabulary test given after qualifying on ASVAB. A more complete validation study involving AFRAT Forms A and B will be accomplished when criterion data are obtained for sufficiently large samples.

Table 16. **AFRAT Form X Validities[a] for Technical Training Grades**

AFSC Code[b]	N	Vocabulary	Comprehension	Total
276	42	.36	.31	.39
304	91	.52	.58	.61
326	57	.27	.37	.36
423	178	.26	.40	.38
426	151	.40	.29	.43
431	217	.26	.35	.35
461	84	.32	.49	.45
462	48	.21	.48	.44
54X	66	.31	.41	.41
55X	69	.17	.09	.13
571	67	.33	.41	.40
605	50	.10	.22	.19
631	84	.27	.50	.47
645	148	.28	.40	.38
702	376	.24	.31	.31
732	38	.15	.40	.34
811	294	.39	.43	.45
902	134	.35	.38	.40
922	38	.34	.14	.26

[a] Not corrected for range restriction.
[b] AFSC code is used to identify clusters of highly similar jobs.

IV. CONCLUSIONS AND RECOMMENDATIONS

Two parallel forms of the AFRAT have been developed and calibrated to three commonly used reading tests and appear to meet administrative and psychometric specifications. All items correlate positively with total test score and are in an appropriate range of difficulty (from average to very easy) for use in detecting reading deficiency.

The AFRAT appears to be a highly reliable instrument (internal consistency coefficients of .92 for Form A and .91 for Form B). The two AFRAT forms appear parallel based on similar distributions of item difficulty and criterion correlation values and statistically equivalent means and variances. AFRAT correlated .60 or higher with each of the three commercial tests.

Interpretation of AFRAT scores is provided by percentile norms and calibration to an average RGL based on the commercial tests. A calibration is also presented with ASVAB General percentile scores. A preliminary analysis indicated the AFRAT would be a valid predictor of technical training performance.

It is recommended that AFRAT Forms A and B replace commercial reading tests for use in screening enlistees for marginal or inadequate reading ability.

REFERENCES

Angoff, W. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971, 508-600.

Gates, A.I., & MacGinitie, W.H. *Gates-MacGinitie Reading Tests Technical Manual.* New York: Teachers College Press, 1972.

Guilford, J.P., & Fruchter, B. *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill, 1978.

Jensen, A.R. *Bias in mental testing.* London: Methuen, 1980.

Karlsen, B., Madden, R., & Gardner, E. *ABLE Handbook.* New York: Harcourt Brace Jovanovich, 1971.

Koplyay, J.B. *Item analysis program (IAP) for achievement tests.* AFHRL-TP-81-22, AD-A107 884. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, October 1981.

Mathews, J.J., Valentine, L.D., Jr., & Sellman, W.S. *Prediction of reading grade levels of service applicants from Armed Services Vocational Aptitude Battery (ASVAB).* AFHRL-TR-78-82, AD-A063 656. Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1978.

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

Tests of Adult Basic Education Examiner's Manual: Level D. Monterey, CA: CTB/McGraw-Hill, 1976.