AD691001

AFHRL-TR-69-103

May 1969

Interpretation and Utilization of Scores on the Air Force Officer Qualifying Test

By

Robert E. Miller

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

PERSONNEL RESEARCH DIVISION
AIR FORCE HUMAN RESOURCES LABORATORY
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas

D D C
RECEIVED
AUG 4 1969
B

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

38

ACCESSION for
CFSTI
DDC
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
DIST. AVAIL. and/or SPECIAL
1

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Composition: Hope De La Cruz
Mary L. Alvarado

AFHRL-TR-69-103

May 1969

INTERPRETATION AND UTILIZATION OF SCORES ON THE AIR FORCE OFFICER QUALIFYING TEST

By

Robert E. Miller

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

PERSONNEL RESEARCH DIVISION
AIR FORCE HUMAN RESOURCES LABORATORY
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas

FOREWORD

The Air Force Officer Qualifying Test is a product of the Personnel Research Division, Air Force Human Resources Laboratory, and is used throughout the Air Force in a variety of programs. Extraction of maximum information from test results depends on widespread dissemination to test users and other interested persons of meaningful data on the characteristics of the test. This report is intended to provide such data in a convenient form.

Research on the Air Force Officer Qualifying Test is conducted under Project 7717, Selection, Classification, and Evaluation Procedures for Air Force Personnel; Task 771706, Selection and Classification Instruments for Officer Personnel Programs.

This report has been reviewed and is approved.

F.L. McLanathan, LtCol, USAF
Chief, Personnel Research Division

ABSTRACT

This report summarizes a large body of data relevant to the proper interpretation and use of aptitude scores on the Air Force Officer Qualifying Test. Included are descriptions of the AFOQT testing program and the general characteristics of the test itself. Technical concepts are introduced by a brief explanation to assist users of AFOQT scores who are not test specialists. Technical data include an extensive sampling of validation studies covering prediction of success in pilot training, navigator training, technical training, and academic courses. Relationships to other well known tests and the Air Force structure of career areas and utilization fields are indicated. Several types of reliability data are presented, together with intercorrelations of the aptitude composites both with and without the elevating effects of overlapping subtests. The Air Force percentile scoring system is discussed in relation to the normal probability curve and the stanine scale. Score distributions are provided for officers, candidates for programs leading to a commission, basic airmen, and 12th grade males. Procedures used in standardizing new forms of the AFOQT through the Project TALENT aptitude composites are described, including operations which maintain relationships with Air Force Academy candidates and the TALENT national sample. Effects of applying minimum qualifying scores and adjustments for level of formal education at the time of testing are explained.

TABLE OF CONTENTS

		Page
I.	Introduction	1
II.	Purpose	1
III.	General Characteristics	2
IV.	The Subtests	4
V.	The Composites	5
VI.	Validity, General	5
VII.	Prediction of Performance in Pilot Training	6
VIII.	Prediction of Performance in Navigator Training	8
IX.	Prediction of Performance in Academic Courses	10
X.	Prediction of Performance in Officer Technical Courses	11
XI.	Relationship to Performance on Other Tests	13
XII.	Relationship to Career Areas and Utilization Fields	14
XIII.	Reliability and Intercorrelations	17
XIV.	Score Distributions	19
XV.	Standardization	22
XVI.	Adjustment for Educational Effects	25
XVII.	Minimum Qualifying Scores	27
References		29

LIST OF TABLES

Table		Page
1	Content and Organization of a Recent Form of the AFOQT	3
2	Relationship between AFOQT Composites and Success in Undergraduate Pilot Training	7
3	Relationship between AFOQT Composites and Success in Undergraduate Navigator Training	9
4	Relationship between AFOQT Composites and Success in Academic Courses	10
5	Relationship between AFOQT Composites and Success in Officer Technical Courses	12

List of Tables (*Continued*)

Table		Page
6	Relationship between AFOQT Composites and Other Tests	13
7	Officer Quality Distribution Statistics by Career Area and Utilization Field	15
8	Officer Quality Distribution Statistics by Academic Major Field	15
9	Percentage of Officers Entering Career Areas and Utilization Fields by Academic Major Field	16
10	Reliability of AFOQT Composites	17
11	Intercorrelation of AFOQT Composites	18
12	Intercorrelation of AFOQT Composites with Effects of Common Subtests Deleted	18
13	Proportion of AFOQT Raw Score Differences in Excess of Chance	19
14	Officer Quality Score Distributions for Candidates for Commissioning Programs and Commissioned Officers	20
15	Pilot Composite Score Distributions for Appropriate and Inappropriate Groups	21
16	Difficulty Level of AFOQT Composites	22
17	Composition of TALENT Composites Corresponding to AFOQT Composites	23
18	Correlation between AFOQT Composites and TALENT Composites	24
19	Performance of 12th Grade Males on the AFOQT	24
20	CEEB Cumulative Distributions and Distribution Statistics for the AFOQT Standardization Group	25
21	Effects of Education on AFOQT Scores	26
22	Pilot and Navigator-Technical Score Distributions for 1,000 Unselected Candidates for Pilot Training	27
23	Pilot and Navigator-Technical Score Distributions for Graduates from 1,000 Candidates for Pilot Training	28

LIST OF FIGURES

Figure		Page
1	Pilot composite and percentage of student pilots graduated versus eliminated	8
2	Navigator-Technical composite and percentage of student navigators achieving academic grade above class median	10
3	Officer Quality composite and percentage of student officers achieving academic grade above class median in OTS	11
4	Verbal composite and percentage of officers achieving final grade above class median in Personnel Officer course	12
5	Quantitative composite and percentage of Air Force Academy cadets achieving score above class median on CEEB Mathematics Aptitude Test	14
6	Air Force scoring systems in relation to the normal probability curve	20

INTERPRETATION AND UTILIZATION OF SCORES ON THE AIR FORCE OFFICER QUALIFYING TEST

I. INTRODUCTION

The purpose of this report is to provide information on the use and interpretation of scores derived from the Air Force Officer Qualifying Test (AFOQT). Such information is of particular importance to officers who use test scores in selection, classification, and assignment of personnel, and those with career counseling responsiblities. Test control officers and military psychologists are also concerned with the use and meaning of AFOQT scores.

AFOQT scores are used operationally in ways which affect the careers of officers and the composition of the Air Force. Detailed instructions in the AFOQT administrative and scoring manuals are designed to insure that scores represent accurately the aptitudes of examinees. This effort is of little avail if the scores are not properly understood and utilized. Users of the scores are not expected to be acquainted with all aspects of testing, but familiarity with pertinent manuals and directives is a minimum requirement.

It is recognized that some users of AFOQT scores are familiar with technical concepts which apply to testing, while others are not. A brief description or rationale of each concept has been included in this report, but no concept is treated exhaustively. Further information may be found in textbooks on psychological testing or statistics as applied to psychology.

This report is primarily concerned with properties of the AFOQT which are not peculiar to any particular form. Some of the data are based on one form only, but these are generalizable to other recent forms, at least in an approximate way. Many of the data have appeared in previous technical publications but have not been brought together in a single source.

II. PURPOSE

A test may be viewed as a device for the measurement of some psychological characteristic. The AFOQT is such a device for measurement of aptitudes important to various officer programs in the Air Force. It is used in the selection of candidates for most training programs leading to a commission and in the qualification of certain categories of applicants for a direct commission. It is also used in the selection of officers for pilot and navigator training and in making initial assignment recommendations for most officers entering their first tour of active duty. It has been used experimentally in the selection of astronauts.

In practice, all uses of the AFOQT involve a prediction. Personnel are selected for programs leading to a commission or to rated status on the basis that they have the personal characteristics and aptitudes necessary for a successful outcome. Prediction is implicit in career counseling also, for an assignment is expected to be satisfying to the incumbent and productive to the Air Force. By measuring the aptitudes of candidates prior to selection, the AFOQT contributes substantially to predictions on which personnel actions are based. By distinguishing between possible assignments, such as pilot or navigator training, the AFOQT accomplishes a classification function in the Air Force personnel system as well.

Personnel actions for which AFOQT scores have relevance are not determined solely by the scores. This is made clear in regulations governing training programs. Other data which may be used formally or informally include results of physical examinations, evidence of compliance with administrative requirements, records of educational and vocational history, and evaluations by commanders or officer boards. In most cases, however, the only measure of the candidate's aptitudes for a program is his AFOQT performance. In programs where minimum qualifying scores exist, AFOQT results can be the sole basis for rejecting a candidate.

III. GENERAL CHARACTERISTICS

The AFOQT evolved from the Aircrew Classification Batteries of World War II and the Aviation-Cadet Officer-Candidate Qualifying Test of 1950. The first instrument published under the name Air Force Officer Qualifying Test appeared in 1953, but a preliminary form was prepared two years earlier. The test is revised biennially to minimize obsolescence and the possibility of compromise. Normally only one form is operational in a given program at a given time. Early forms were distinguished by a letter designation, but the fiscal year of implementation is now used to designate the form.

The AFOQT is based ultimately on analyses of tasks required of student pilots, navigators, and officers. These analyses are not accomplished anew for each form of the test, but cognizance is taken of the possibility that the most appropriate aptitudes for measurement may change over a period of time. With the advent of high performance jet aircraft this question was raised acutely regarding pilot aptitudes. However, interviews with a group of command pilots failed to disclose that a serious problem existed. Studies of test results showed that the AFOQT has substantially the same effectiveness as a predictor of training performance in both jet and piston powered aircraft.

Successive forms of the AFOQT closely resemble each other. They differ in such respects as the number of items, arrangement of subtests, administrative and scoring instructions, and conversion tables. Occasionally one subtest is replaced by another measuring the same aptitude, or a subtest may be dropped completely because of declining effectiveness. An example of a subtest dropped for lack of effectiveness is Interests. This subtest yielded four interest scores but was found to have little utility in Form G. It has not appeared in subsequent forms.

Each new form is actually an entire test battery published in five separate booklets. This design permits flexibility in the use of the test. It is necessary to administer only those booklets relevant to the specific program for which the examinee applies. Using commands, however, are encouraged to require initial administration of all booklets relevant to any program for which the examinee might conceivably apply. For most male examinees this means all five booklets. Female examinees take only Booklet 1 and the first section of Booklet 2.

In addition to the booklets, each form includes administrative and scoring manuals, keys for hand and machine scoring, and special answer sheets. For testing in the AFROTC program, answer forms are provided for use in a centralized scoring facility utilizing a video scanner and computer. Modified administrative and scoring instructions are required for use with these forms. Testing record cards and interpretive materials are prepared and updated as needed. Most AFOQT materials are controlled items and are not available for distribution outside the Air Force.

The complete AFOQT contains approximately 525 test items and requires almost six hours for administration. There are thirteen subtests into which the items are organized and from which scores can be obtained. The subtests, however, are not scored separately except for research purposes. The operational scoring keys yield five composite scores made up of sums of partly overlapping sets of subtests. These operational scores are known as the Pilot, Navigator-Technical, Officer Quality, Verbal, and Quantitative composites. An outline of the AFOQT structure in terms of items, subtests, and composites is shown in Table 1.

It is possible to form other composite scores by different groupings of subtests. This has sometimes been done to meet special needs of specific programs. Thus there has been an Airmanship composite, an Academic composite, and a Career Potential composite. None of these special composites are currently used in any program.

Each composite constitutes a measure of an aptitude area of importance to success in certain officer training programs. The selection of subtests for each composite is based on extensive studies which show that examinees who do well on specific combinations of subtests tend also to do well in certain types of training. The aptitudes required for these types of training differ from each other sufficiently to justify the use of different composites.

The various aptitude areas are not completely independent. A moderate positive relationship exists among them such that extremely high and extremely low scores on different composites do not often occur in one examinee's performance. Such differences are possible, however, and their occasional occurrence is not necessarily an indication of improper test administration or scoring.

Application of the scoring keys yields a set of raw scores which are unwieldy to handle and difficult to interpret. Raw scores are therefore converted to Air Force percentile scores by the use of conversion tables found in the scoring manual. The range of the Air Force percentile scale is from 01 to 95 in twenty steps. Such a scale permits interpretation of scores in terms of the relative standing of individual examinees on a given composite. The meaning of the 85th percentile on any composite, for example, is that the examinee's performance exceeds that of 85 percent of the examinees for whom the test is appropriate but does not exceed that of 90 percent of such examinees.

The AFOQT is constructed in such a way that a given percentile has the same meaning on successive forms of the test. In addition, it is possible to interpret differences between scores attained by different examinees on the same composite, and differences between scores of the same examinee on different composites. The latter type of interpretation is essentially diagnostic because it is concerned with strengths and weaknesses in the aptitude areas measured. Score differences, however, are often a result of chance, with the consequence that interpretations of differences may be at variance with other evaluations of relative aptitude levels. It is possible to estimate the proportion of test score differences in excess of chance.

AFOQT scores are entered in various personnel records, and examinees are generally given information on their own performance. If scores are communicated to examinees, it is important that the meaning of the scores also be communicated. A counseling responsibility is in fact implied in such communication because different examinees do not perceive their scores in the same light. A minimum qualifying score for a desired program may be all that one examinee considers necessary, while another may view the same score as a severe personal blow.

Table 1. **Content and Organization of a Recent Form of the AFOQT**

Booklet and Subtest	No. of Items	Aptitude Composite				
		Pilot	Nav-Tech.	Off. Qual.	Verbal	Quant.
Booklet 1						
Quantitative Aptitude	60		X	X		X
Booklet 2						
Verbal Aptitude	60			X	X	
Officer Biographical Inventory	100			X		
Booklet 3						
Scale Reading[a]	48		X			
Aerial Landmarks[a]	40		X			
General Science	24		X			
Booklet 4						
Mechanical Information	24	X	X			
Mechanical Principles	24	X	X			
Booklet 5						
Pilot Biographical Inventory	50	X				
Aviation Information	24	X				
Visualization of Manueuvers[a]	24	X				
Instrument Comprehension[a]	24	X				
Stick and Rudder Orientation[a]	24	X				

[a]Speeded subtests

IV. THE SUBTESTS

Although not considered separately in operational settings, the various subtests do constitute the entire content of the composites. Understanding of the composites is therefore enhanced by knowledge of the nature of the subtests, and, where possible, by a perusal of the individual items. Following is a brief description of each subtest:

Quantitative Aptitude consists of items involving general mathematics, arithmetic reasoning, and interpretation of data read from tables and graphs.

Verbal Aptitude consists of items pertaining to vocabulary, verbal analogies, reading comprehension, and understanding of the background for world events.

Officer Biographical Inventory consists of items pertaining to past experiences, preferences, and personality characteristics known to be related to success in officer training.

Scale Reading consists of items in which readings are taken of various printed dials and gauges. Many of the items require fine discriminations on nonlinear scales.

Aerial Landmarks consists of pairs of photographs of terrain as seen from different positions of an aircraft in flight. Landmarks indicated on one photograph are to be identified on the other.

General Science consists of items related to the basic principles of physical science. The emphasis is on physics, but other sciences are also represented.

Mechanical Information consists of items pertaining to the construction, use, and maintenance of machinery. Some of the items are concerned with the use of tools.

Mechanical Principles consists of diagrams of complex apparatus. Understanding of how the apparatus operates or the consequences of operating it in a specified manner is required.

Pilot Biographical Inventory consists of items pertaining to background experiences and interests known to be related to success in pilot training.

Aviation Information consists of semi-technical items related to various types of aircraft, components of aircraft, and operations involving aircraft.

Visualization of Maneuvers consists of items requiring identification of the silhouette which expresses the attitude of an aircraft in flight after executing a verbally described maneuver.

Instrument Comprehension consists of items similar to those in Visualization of Maneuvers except that the maneuvers are indicated by readings of a compass and artificial horizon.

Stick and Rudder Orientation consists of sets of photographs of terrain as seen from an aircraft executing a maneuver. The proper manipulation of the control stick and rudder bar to accomplish the maneuver must be indicated.

Each subtest is made up of test items in the numbers shown by Table 1. Most items are of the multiple choice type with four or five alternatives, but some biographical items are of the forced choice type. Items are accepted for inclusion in the AFOQT only after they have been tested in experimental booklets to determine their characteristics. About 10 percent of the items in most subtests are carried over to the next form. These anchor items make it possible to compare performance on a common set of items in groups of examinees who were administered different forms of the test. Formulas to correct for chance success are applied to composites having speeded subtests.

Technical data of several types have been collected on AFOQT subtests and items. Included are data on reliability, validity, internal consistency, intercorrelations, and difficulty. Most of these data have been published elsewhere. They are not included in this report because it is not desired to encourage interpretation of subtests or items. Such interpretations are usually misleading because individual subtests and items are insufficiently stable for practical use. Only the composites possess the properties required of interpretable test data.

V. THE COMPOSITES

Table 1 and the description of the subtests suffice to describe the content of the composites. There are also general characteristics applicable to each composite and recommended uses for each. The recommended uses are based on empirical data in as many instances as possible, but some are based on logical analysis.

The Pilot composite is designed to predict success in undergraduate pilot training. The specific measure of performance used in developing this composite was elimination from training by reason of flying deficiency. Examinees with high Pilot scores may be expected to possess in sufficient degree the aptitudes necessary for successful completion of training. Those with low scores represent a serious risk of elimination. Success in pilot selection requires that these expectations be generally confirmed by experience. The Pilot composite does not distinguish between aptitudes for flying different types of aircraft.

The Navigator-Technical composite is designed to predict success in undergraduate navigator training and in training programs emphasizing mechanical and engineering concepts. Examples of such programs are officer technical courses in the areas of communications, electronics, armament, aircraft maintenance, photography, cartography, meteorology, and technical intelligence. This composite also has relevance for success in pilot training. In many types of aircraft the pilot must additionally function as navigator.

The Officer Quality composite is a measure of learning ability or academic aptitude, coupled with a biographical inventory. Examinees with high Officer Quality scores may be expected to do well in any training program having appreciable academic content. Examples are the academic phases of Officer Training School (OTS) and the Air Force Academy, and the academic curriculum associated with the AFROTC program. Officer Quality is a predictor of academic averages, specific course grades in a variety of fields, and certain nonacademic performance measures obtained in educational settings.

The Verbal composite contains four types of items which in early AFOQT forms constituted four short subtests. These have now been consolidated into one. The Verbal composite is designed to predict success in training programs which emphasize linguistic skills. Examples are in the areas of administrative services, personnel administration, public information, education and training, psychological warfare, and historical activities.

The Quantitative composite is composed of a single subtest into which three former short subtests were consolidated. This composite is predictive of success in training courses which emphasize mathematical ability. Examples are programs in statistical services, accounting, auditing, disbursing, and supply.

VI. VALIDITY: GENERAL

The indispensible property of a test is validity. Validity is commonly defined either as the extent to which a test measures what it is supposed to measure, or the extent to which whatever it measures is known. Several types of validity are recognized. For aptitude tests such as the AFOQT, the most relevant type is predictive validity. This is demonstrated by administering the test to a group of examinees prior to their admission to a training program, collecting data on the outcome of training when these become available, and expressing the relationship between test scores and outcome in some way. The usual method of expressing the relationship is by a statistic known as the correlation coefficient.

Since nearly all testing is done on samples of some population, rather than on the entire population, the results are somewhat peculiar to the samples. It may be that an obtained correlation coefficient is merely a function of chance factors affecting the composition of the sample. Such a correlation is effectively equal to zero and indicates an absence of relationship in the population. Methods exist for determining the probability that an obtained correlation could arise by chance. The generally accepted convention is that when the probability is .05 or less, the correlation is said to be statistically significant. If the sample is large, a very small correlation can be statistically significant.

When applied to the relationship between test scores and an independently measured criterion of performance, such as course grades, a correlation coefficient becomes a validity coefficient. Even low validity coefficients, if statistically significant, represent a relationship between test scores and outcome of

training such that a better prediction of outcome is possible with the scores than without them. This improvement often has practical value, and its extent can be quantitatively expressed.

A high validity coefficient, however, is more desirable than a low one because it represents a stronger relationship and more accurate prediction. The reduction in errors of prediction as the correlation increases is nonlinear and becomes rapid only as the correlation becomes fairly high. There is no specific value to define the lower limit of a high correlation, but the closer it approaches +1.00 or -1.00 the higher it is. For predictive purposes, a negative correlation is as useful as a positive one of equal absolute value, but a negative correlation is likely to be more difficult to understand.

In the prediction of academic grades, where predictive validities tend to be higher than in other situations, a validity coefficient of .50 might be considered exceptionally good for a single test. Higher validities can often be obtained from a combination of several carefully selected tests which are differentially weighted to provide maximum prediction of the criterion. Combinations which include AFOQT scores have attained validities as high as .74 in predicting academic grades of Air Force Academy cadets, but this validity applies to the combination and not to the AFOQT alone.

VII. PREDICTION OF PERFORMANCE IN PILOT TRAINING

Table 2 presents validity coefficients of the AFOQT for prediction of the outcome of undergraduate pilot training. Validities of all composites for which relevant data exist are included, but the Pilot composite is the only one designed specifically to predict any of these criteria. Data from several sources of commission are provided. The AFROTC source is limited to those who participated in the light plane Flying Instruction Program while in college. The table shows the number of cases (N) in each group and the total elimination rate for each group. Blank cells represent absence of data or insufficient data for stable computations. Statistically significant validities are indicated by asterisks.

The table shows two distinct types of criteria of success in pilot training. The first three criteria belong to one type and consist of numerical grades for various aspects of training. The remaining criteria are dichotomies between graduation and elimination from the program for some specified reason. Correlations with the dichotomies are of a special type known as biserials. A biserial coefficient estimates what the correlation would be if the criterion were not dichotomized. It is apparent that the criteria are far from equally predictable. This is to be expected because they are not closely related to each other. The mean correlation between the three numerical grades, for example, is .42.

The final Pilot composite column in the table contains a corrected form of the Pilot data from the Total column. The correction is for a restriction in the range of Pilot scores entering into the validation study. Since all cases in the study must have test scores and criterion measures, it follows that examinees with scores too low to qualify for training could not be included. The absence of these cases limits the variability of scores and depresses the validity coefficients. Methods exist to correct for this effect under several different circumstances. Here the correction is applied only to the Pilot composite as the composite of greatest interest.

Properly corrected coefficients do not exaggerate the validity of a test. Rather, they provide the best estimate of it. This is because the test is applied to all applicants, including those who do not qualify, and its effectiveness should be evaluated on all cases to which it is applied. All Pilot composite validities in the table except the corrected ones are to some extent underestimates. Validities of the other composites are probably underestimates also. Corrected validities are not often computed because of difficulties in meeting the assumptions underlying the correction process.

The various sources of commission yield somewhat different validity coefficients. Many of the differences are too small to be meaningful in practice. Nevertheless, the best estimate of validity in a group of examinees from the same source of commission is probably the validity computed specifically on that source. Validities based on the total group are best used for mixed sources or sources not otherwise represented in the table.

To facilitate interpretation of validity coefficients, Figure 1 has been provided as a graphic expression of a validity from Table 2. The figure shows the percentage of student pilots from all sources combined who are expected to graduate from pilot training at various pilot composite percentile levels. In this figure, the percentage values are those to be expected theoretically, based on the corrected empirical validity of .40 and the elimination rate of 21 percent in the qualified group. This amounts to an expected elimination rate of about 30 percent in the qualified and unqualified groups combined.

Table 2. **Relationship between AFOQT Composites and Success in Undergraduate Pilot Training[a]**

Criterion	Pilot Composite					Nav-Tech Composite				Off Qual Composite			
	AFROTC	OTS	AFA	Total	Total Corrected	AFROTC	OTS	AFA	Total	AFROTC	OTS	AFA	Total
Academic Grade	.41*	.55*	.34*	.41*	.48*	.58*	.55*	.36*	.47*	.50*	.48*	.17*	.42*
Flying Grade	.18*	.33*	.25*	.28*	.33*	.22*	.19*	.17*	.20*	.13*	.17*	.03	.10*
Military Grade	.13*	.15	.13*	.15*	.18*	.16*	.01	.14*	.16*	.20*	.20*	.11	.21*
Academic Elim.	.34*			.40*	.53*	.59*			.56*	.52*			.55*
Flying Deficiency Elim.	.28*	.20*	.08	.22*	.36*	.10	.01	.08	.11*	-.01	-.10	-.03	.04
Fear of Flying Elim.	.23*	.10		.25*	.42*	.16	.08		.27*	.02	-.29		.11
Self Initiated Elim.	.29	.29		.20*	.37*	.17	.22		.19*	.10	-.07		.11
Medical Elim.	-.02	.26		.13	.24*	-.12	-.03		.00	-.06	.19		.16*
Total Elim.	.28*	.22*	.03	.26*	.40*	.16*	.06	.09	.18*	.05	-.10	.10	.12*
N	522	197	305	1,500	1,500	522	197	305	1,500	522	197	305	1,500
Elim. Rate	.23	.34	.07	.21	.21	.23	.34	.07	.21	.23	.34	.07	.21

[a]Based on student pilots trained in T-38 aircraft, classes 63A through 65B. Asterisks represent statistically significant correlations.

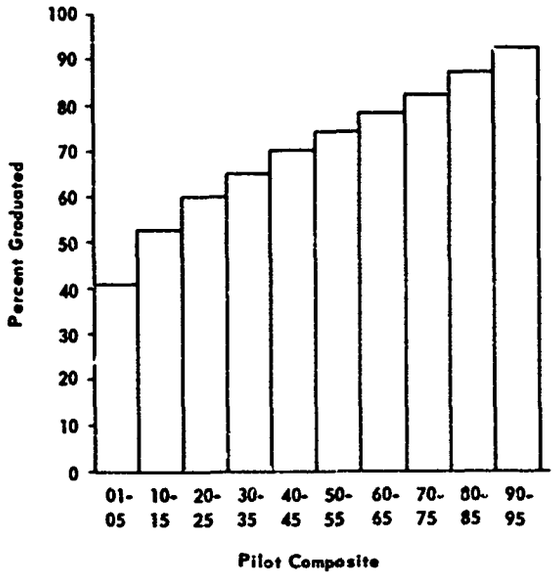

Fig. 1. **Pilot composite and percentage of student pilots graduated versus eliminated.**

Figure 1 shows that the percentage of graduates increases appreciably as test scores increase. This trend illustrates the validity of the Pilot composite. The figure is essentially similar to an expectancy table such as is used in educational counseling to show that students with low test scores may be successful but are not as likely to succeed as those with high scores.

There is an additional meaningful way to express the validity of the Pilot composite. This is in terms of dollar savings to the pilot training program. Data on the number of examinees tested in a recent fiscal year, the validity of the test, and the elimination rate among the selectees permit an estimate that there were 365 examinees disqualified by the Pilot composite who would have been eliminated had they entered training. At an estimated average cost per eliminee of $24,000, the total savings in one year from application of the Pilot composite is found to be $8,760,000. The average cost figure in this computation is subject to rapid obsolescence and is probably an underestimate.

The AFOQT has been used to predict success in pilot training in other countries. Efforts to do this with direct translations into the language of the country are unsatisfactory because the test is in many ways inappropriate to the foreign culture. A more thorough adaptation of the test may be fairly successful. Modified Pilot composite validities for predicting ratings by flying instructors have been reported from Spain and Norway. The coefficients were .52 and .53, respectively, in samples large enough for these coefficients to be statistically significant.

VIII. PREDICTION OF PERFORMANCE IN NAVIGATOR TRAINING

Table 3 presents AFOQT validity data for the prediction of performance in undergraduate navigator training. Data for this table came from the same study as the data in Table 2, and they are organized in an analogous manner. In this instance, the Total group contains 617 Aviation Cadets in addition to other sources, and it is these Cadets who account largely for the washing out of some validities in the Total group. A correction for range restriction is applied in the Total group to the Navigator-Technical composite. The mean correlation among the three course grades is .46.

Figure 2 is provided to show graphically the validity of the Navigator-Technical composite for the prediction of academic grades in undergraduate navigator training. The figure shows the percentage of students attaining grades above the median of their class at various Navigator-Technical percentile levels. The percentages are theoretical but are computed from the corrected empirical validity coefficient of .42. Figure 2 is based on nearly the same validity as Figure 1 and approximates what Figure 1 would look like with a 50 percent pilot elimination rate.

Table 3. **Relationship between AFOQT Composites and Success in Undergraduate Navigator Training[a]**

Criterion	Pilot Composite				Nav-Tech Composite					Off Qual Composite			
	AFROTC	OTS	AFA	Total	AFROTC	OTS	AFA	Total	Total Corrected	AFROTC	OTS	AFA	Total
Academic Grade	.34*	.31*	-.09	.27*	.47*	.50*	.22	.35*	.42*	.38*	.42*	.46*	.34*
Flying Grade	.15*	.10*	.03	.12*	.24*	.26*	.38*	.22*	.27*	.12*	.18*	.39*	.13*
Military Grade	.16*	.07	-.19	.10*	.21*	.16*	.01	.14*	.17*	.19*	.20*	.22	.18*
Academic Elim.	.27*	.30*		.23*	.32*	.38*		.24*	.36*	.28*	.21*		.16*
Flying Deficiency Elim.	.07	.16		.13*	.16	.26*		.16*	.26*	.18	.20		.13*
Fear of Flying Elim.				.11				.07	.25*				.10
Self Initiated Elim.	.21	.36*		.01	-.03	.19		-.16*	-.30*	-.10	-.07		-.05
Military Elim.				-.23*				-.29*	-.29*				-.09
Total Elim.	.17*	.22*		.07*	.20*	.25*		-.02	-.02	.15*	.14*		.04
N	700	570	32	2,132	700	570	32	2,132	2,132	700	570	32	2,132
Elim. Rate	.07	.11	.00	.14	.07	.11	.00	.14	.14	.07	.11	.00	.14

[a]Based on student navigators in classes 63-03 through 65-04. Asterisks represent statistically significant correlations.

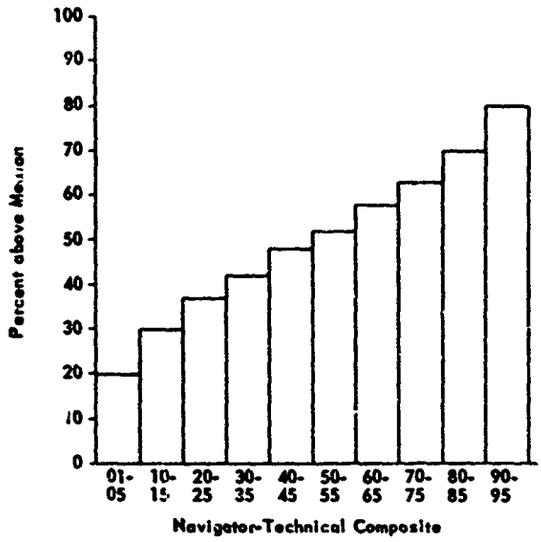

Fig. 2. **Navigator-Technical composite and percentage of student navigators achieving academic grade above class median.**

IX. PREDICTION OF PERFORMANCE IN ACADEMIC COURSES

Table 4 presents validity coefficients of AFOQT composites for the prediction of a variety of academic performance measures obtained in Air Force settings. The measures include over-all averages, final course grades, and a few nonacademic measures gathered at the Air Force Academy. The table indicates the source of each measure and the number of cases on which it is based. The fourth column shows Officer Quality validities corrected for range restriction where the assumptions could be met. Motivational Elimination is a dichotomy predicted by biserial correlations.

Table 4. **Relationship between AFOQT Composites and Success in Academic Courses[a]**

Criterion	Pilot	Nav-Tech	OQ	OQ Corrected	Verbal	Quant	N	Source
Academic Average			.52*	.57*			90	OTS Class 60A
Over-all Average		.15	.35*	.39*			90	OTS Class 60A
Academic Average, 4 years	.17*	.31*	.33*	.37*	.25*	.31*	971	15 AFROTC Dets, 1957-61
Academic Average	.17*	.35*	.45*		.30*	.45*	495	AF Academy Class 64
Chemistry 102	.02	.30*	.38*		.14*	.40*	224	AF Academy Class 62
English 102	-.10	.01	.14*		.08	.12	239	AF Academy Class 62
Geography 102	.01	.18*	.30*		.17*	.14*	261	AF Academy Class 62
Graphics 102	.43*	.57*	.51*		.32*	.54*	176	AF Academy Class 61
History 102	-.14*	.01	.27*		.18*	.08	216	AF Academy Class 62
Mathematics 102	.06	.23*	.17*		-.05	.26*	260	AF Academy Class 62
Military Science 101	.08	.17*	.25*		.26*	.18*	176	AF Academy Class 61
Philosophy 101	.11	.26*	.35*		.27*	.28*	133	AF Academy Class 61
Physics 201-202	.25*	.49*	.47*		.24*	.56*	222	AF Academy Class 59
Psychology 201-202	.19*	.28*	.40*		.39*	.28*	222	AF Academy Class 59
Electrical Engineering 302	.20*	.40*	.37*		.23*	.43*	173	AF Academy Class 59
Engineering Drawing 300	.40*	.51*	.31*		.09	.29*	144	AF Academy Class 62
Mechanics 302	.01	.26*	.23*		.03	.37*	172	AF Academy Class 59
Cadet Effectiveness Rating	-.06	-.06	-.01		-.11*	-.08	495	AF Academy Class 64
Extracurricular Activities	-.09*	-.09*	-.09*		-.09*	-.07	495	AF Academy Class 64
Nonacademic Average	-.09*	-.10*	-.06		-.13*	-.10*	495	AF Academy Class 64
Motivational Elimination			.28*		.24*	.20*	960	AF Academy Class 71

[a]Asterisks represent statistically significant correlations.

Many of the Air Force Academy data are available for more than one class. Where this is true, data are reported only for the most recent class. Course numbers are provided to show the class year in which the course is normally taken. The lower numbers indicate the earlier class years. Unless otherwise indicated, all Academy criteria are from the fourth class (freshman) year. For the upper class years, the period over which predictions are made must obviously be longer, extending to three years or more.

The principal value of presenting validities for specific course grades at the Air Force Academy is that these validities can be generalized within limits. Validities should be somewhat similar for courses with similar content in other educational institutions. However courses having the same name in different institutions may have markedly different content. Also, shifting validities for the same course in successive Academy classes suggest a further limitation on generalizability. Such shifts were observed frequently in early classes.

Figure 3 illustrates an Officer Quality validity coefficient from Table 4. The figure shows the percentage of student officers expected to attain an academic average above the class median in OTS at various Officer Quality percentile levels. The figure is constructed in the same manner as Figure 2 and is based on the corrected empirical validity coefficient of .57.

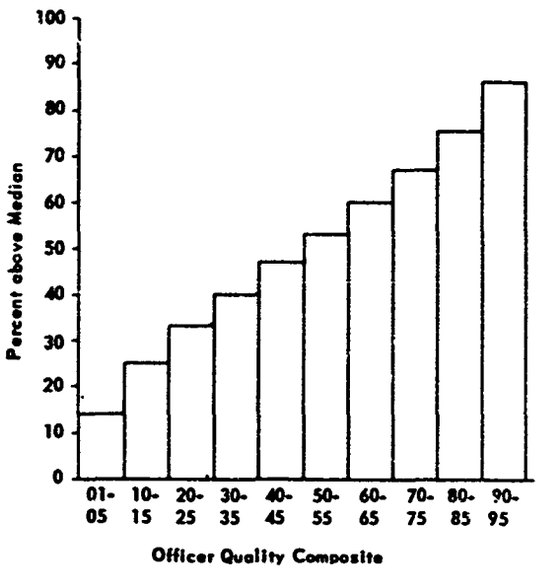

Fig. 3. **Officer Quality composite and percentage of student officers achieving academic grade above class median in OTS.**

X. PREDICTION OF PERFORMANCE IN OFFICER TECHNICAL COURSES

AFOQT scores are used more informally in assignment of officers to technical courses than in selection for flying training or programs leading to a commission. This is because no minimum qualifying score exists on any composite for admission to any technical school. The Navigator-Technical, Verbal, and Quantitative composites are likely to be good indicators of success in technical courses, but they should be considered in relation to a course assignment only when they are known to be valid for the particular course in question.

Table 5 shows validities for various officer technical courses. Some courses are shown with course numbers for unambiguous identification. Data for courses lacking numbers are from earlier studies and should be interpreted with caution. Validities for these courses may be suggestive of current validities, but only where it is known that the course content has not undergone basic changes.

Table 5. **Relationship between AFOQT Composites and Success in Officer Technical Courses[a]**

Criterion	Pilot	Nav-Tech	Off Qual	Verbal	Quant	N
Aircraft Maintenance OB 4341	.46*	.58*	.58*	.35*	.55*	164
Air Police OB 7721	.04	.29*	.31*	.15	.31*	97
Air Transportation OB 6021	.17	.24*	.29*	.13	.33*	76
Communications OB 3031	.50*	.56*	.55*	.39*	.50*	84
Personnel OB 7321	.23*	.43*	.48*	.36*	.45*	116
Supply OB 6421	.22*	.46*	.52*	.38*	.50*	125
Surface Transportation OB 6031	.18	.40*	.42*	.26*	.34*	70
Aircraft Controller			.41*			160
Air Electronics			.44*			289
Air Intelligence			.45*	.47*		177
Armament			.63*			169
Budget and Fiscal			.38*		.39*	147
Classification and Assignment			.36*			197
Electronics Countermeasures		.48*	.37*			188
Ground Electronics			.40*			671
Photo-Radar Interpretation			.53*			63
Statistical Services			.34*			99

[a]Based on validation studies performed between 1951 and 1960. Asterisks represent statistically significant correlations.

Figure 4 shows the percentage of students in the Personnel Officer course, OB7321, who are expected to exceed the class median on the final course grade at various levels of the Verbal composite. The figure is based on the empirical validity coefficient of .36 in Table 5. Correction of this coefficient for range restriction was not attempted because there is no specific minimum qualifying score to cut off the bottom of the score distribution.

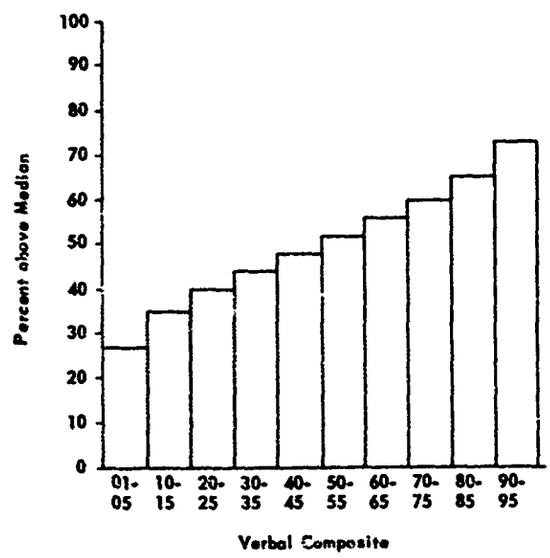

Fig. 4. **Verbal composite and percentage of officers achieving final grade above class median in Personnel Officer course.**

XI. RELATIONSHIP TO PERFORMANCE ON OTHER TESTS

It is helpful in test interpretation to understand the relationships between the test being interpreted and other tests with well known properties. Relationships between two tests are usually expressed by the correlation between their scores. Such correlations can be interpreted as validities in which the criterion for one test is the score on the other. If the tests are administered at approximately the same time, the validity expressed is known as concurrent validity. It does not necessarily imply predictive validity.

High correlations between tests can be taken to mean that the tests are measuring approximately the same psychological attribute, even though the names of the tests may not suggest that this is so. Low correlations indicate that the tests are measuring something different. Intermediate correlations show that the tests are measuring the same attribute or covarying attributes to some degree. A study of the interrelationships among tests can thus shed light on the psychological characteristics which they measure. The relationship between a test and a hypothesized psychological characteristic represents still another kind of validity, known as construct validity.

Table 6 presents correlations between AFOQT composites and several other tests. The sample sizes and sources of the data are also shown. Because of the temporal relationships involved, the coefficients represent concurrent validities. They also represent construct validities because they support such expectations as that the AFOQT Verbal composite should correlate highly with the CEEB Verbal Aptitude Test. However, the tests were not administered together to provide systematic evidence for any hypothetical construct.

Table 6. **Relationship between AFOQT Composites and Other Tests[a]**

Test	Pilot	Nav-Tech	OQ	Verbal	Quant	N	Source
CEEB Verbal Aptitude	.25*	.30*	.52*	.71*	.29*	616	AF Academy Class 64
CEEB English Composition	.14*	.21*	.40*	.46*	.31*	616	AF Academy Class 64
CEEB Math Aptitude	.27*	.59*	.50*	.28*	.72*	616	AF Academy Class 64
CEEB Intermediate Math	.27*	.47*	.42*	.19*	.60*	616	AF Academy Class 64
ETS High School Rank	-.04	.12*	.26*	.14*	.24*	616	AF Academy Class 64
Calif. Reading, Vocabulary			.51*	.61*	.26*	444	OTS Classes 66E-G
Calif. Reading, Comprehension			.65*	.57*	.57*	444	OTS Classes 66E-G
Calif. Reading, Total			.68*	.66*	.51*	444	OTS Classes 66E-G
Davis Reading, Level			.46*	.56*	.26*	440	OTS Classes 66E-G
Davis Reading, Speed			.57*	.65*	.28*	440	OTS Classes 66E-G
Vocabulary Test G-T	.05	.12*	.40*	.57*	.20*	722	AF Academy Class 63
Survey of Study Habits and Attitudes	.03	.09	.18*	.09	.27*	414	AF Academy Class 62
AFROTC Pre-Enrollment Test			.82*	.68*	.72*	387	OTS Classes 66E-G
Physical Aptitude Examination	-.06	-.09*	-.09*	-.12*	-.09*	616	AF Academy Class 64

[a]Asterisks represent statistically significant correlations.

Most tests in Table 6 are well known commercial tests for selection and counseling purposes. The College Entrance Examination Board (CEEB) tests are used in a national program of testing for admission to college. ETS High School Rank is an adjusted and standardized form of the high school average. The AFROTC Pre-Enrollment Test is an operational Air Force test used in the AFROTC program as a screening device for Officer Quality. The Physical Aptitude Examination is an Air Force Academy selection test involving performances demonstrating physical strength and skill.

Figure 5 illustrates the relationship between the AFOQT Quantitative composite and the CEEB Mathematics Aptitude Test. The figure utilizes the empirical correlation of .72 between these two tests and expresses the percentage of examinees who attain a CEEB mathematics aptitude score above the class median at various AFOQT Quantitative composite levels. Because of the high correlation and similar content, the relationship demonstrated is one of equivalence. Equivalence also exists between the CEEB Verbal Aptitude Test and the AFOQT Verbal composite.

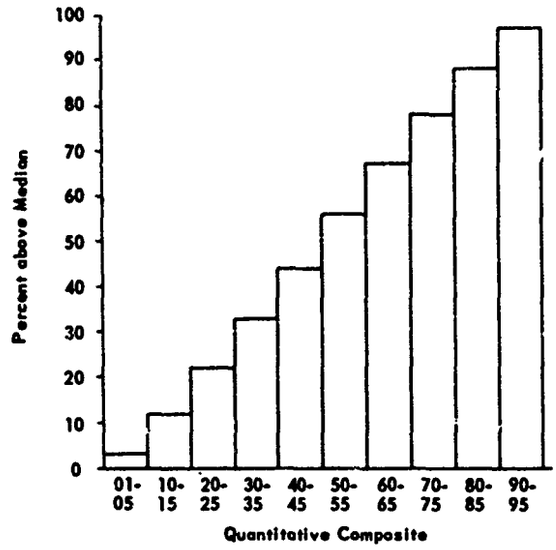

Fig. 5. **Quantitative composite and percentage of Air Force Academy cadets achieving score above class median on CEEB Mathematics Aptitude Test.**

A factor of crucial importance in nearly all training programs and most duty assignments is reading comprehension. It is therefore of interest to compare Officer Quality scores with scores on a reading test. The Comprehension scale of the California Reading Test was chosen for this purpose. Grade levels on this scale were estimated from Officer Quality scores in a sample of 444 OTS students. It was found that the 50th percentile on the Officer Quality composite corresponds to a reading comprehension grade level of 14.4. At the 25th percentile the corresponding value is 13.4, and at the 75th percentile it is 15 1. These results refer to the sample as a whole and do not necessarily describe individual cases.

XII. RELATIONSHIP TO CAREER AREAS AND UTILIZATION FIELDS

Air Force tests are not ordinarily used to predict performance on the job. Performance is considered to be a function of training. Moreover, tests frequently do not predict on-the-job performance very well. This can be attributed in many instances to unreliability or irrelevance of the criterion. Officer Effectiveness Reports (OERs) can not be well predicted by tests, and the ultimate criteria of combat performance are even more difficult to predict. Validities of about .10 have been reported for Officer Quality as a predictor of OERs. This validity would be significant only in large samples.

It is nevertheless possible to detect relationships in the form of differences between career areas and utilization fields in test performance. These differences become apparent when comparisons are made of score distributions for the various areas and fields. The commonly used statistics for such comparisons are the mean and a measure of variability known as the standard deviation. Differences between selected career areas and utilization fields in terms of Officer Quality percentile distributions are presented in Table 7. The table is based on reported assignments of OTS graduates.

Differences between career areas and utilization fields in terms of score distributions can be partially accounted for by differences between major academic fields. Currently, all officers are required to be college graduates at the time of commissioning. Because of the diversity of educational influences in the many colleges from which officers are drawn, one can expect AFOQT score distributions to vary both with the college and the major field of study. There are known to be colleges having AFROTC detachments whose distributions of Officer Quality scores do not even overlap.

Differences between major fields of study with respect to Officer Quality distributions are shown in Table 8. The table is organized in the same manner as Table 7. It is based on subsamples of the sizes shown from a total of 6,797 examinees who were tested in 1968 for all programs except AFROTC. Some of the score distributions are unusually high. This is a consequence of selective effects generated in the more demanding academic fields.

Table 7. **Officer Quality Distribution Statistics by Career Area and Utilization Field[a]**

Career Area or Utilization Field	N	Mean	Standard Deviation	Percent of Cases at or above 50th Percentile
Operations Area	541	55.2	21.6	57.5
Pilot	204	59.2	20.6	68.1
Navigator-Observer	257	53.0	21.7	51.4
Aircraft Control	59	46.9	20.6	47.5
Scientific and Development Engineering Area	261	72.0	19.3	85.8
Weather	164	72.6	18.1	87.2
Scientific	44	73.4	18.6	88.6
Electronics and Maintenance Engineering Area	571	67.1	21.2	77.1
Communications-Electronics	123	69.4	20.6	80.5
Avionics	281	65.8	21.4	75.4
Civil Engineering Area	39	66.4	20.0	76.9
Materiel Area	222	53.5	19.6	55.9
Supply Services	157	51.4	18.3	51.6
Comptroller Area	59	57.3	22.0	64.4
Personnel Resources Management Area	319	54.8	20.7	58.3
Information Area	93	49.9	21.5	52.7
Intelligence Area	44	74.8	17.2	93.2
Security Police Area	150	47.8	20.2	45.4

[a]Based on subsamples of OTS graduates in 1963 and 1964.

Table 8. **Officer Quality Distribution Statistics by Academic Major Field[a]**

Major Field	N	Mean	Standard Deviation	Percent of Cases at or above 50th Percentile
Electrical Engineering	523	74.9	25.1	85.1
Mechanical Engineering	370	69.4	26.0	77.8
Civil Engineering	96	66.1	30.0	72.9
Other Engineering	98	62.9	31.6	64.3
Physics	144	79.8	23.7	86.1
Chemistry	168	69.5	26.5	78.6
Biology	225	50.9	30.6	55.6
Mathematics	329	69.5	27.1	79.3
Business Administration	597	38.8	29.0	37.2
Social Science	77	38.0	30.5	36.4
Education	70	33.6	28.9	35.7
Unspecified or Unknown	473	46.1	31.0	48.4

[a]Based on subsamples of 6,797 examinees tested in 1968 for all programs except AFROTC.

Table 9 shows the degree of concentration of specific academic fields in specific career areas and utilization fields. The table indicates that no academic field is channeled exclusively into a single utilization field, and that no utilization field absorbs any academic field to the exclusion of all others. Some utilization fields include officers with very heterogeneous academic backgrounds. Where there is an academic field related to a utilization field, however, most officers in the utilization field have the related academic background. Table 9 illustrates the use of educational data in making officer assignments.

Table 9. **Percentage of Officers Entering Career Areas and Utilization Fields by Academic Major Field[a]**

Career Area or Utilization Field	N	Major Field											
		Elect. Engr.	Mech. Engr.	Civil Engr.	Other Engr.	Physics	Chemistry	Biology	Math.	Bus. Admin.	Social Science	Educ.	Unspec. or Unknown
Operations Area	1,883	1.0	1.8	0.5	2.2	3.3	3.6	4.2	11.7	17.4	19.1	7.0	28.2
Pilot	383	2.1	3.9	1.0	2.9	2.9	2.1	5.0	8.1	21.9	19.1	7.0	23.0
Navigator-Observer	383	1.0	2.1	1.0	1.0	2.1	6.0	3.9	11.0	19.1	20.1	8.1	25.3
Aircraft Control	59	0.0	0.0	0.0	0.0	1.7	0.0	1.7	1.7	62.7	10.2	5.1	16.9
Scientific and Devel. Engr. Area	231	24.2	10.0	0.0	4.3	19.0	4.3	0.0	26.0	0.4	1.3	1.3	9.2
Weather	96	0.0	0.0	0.0	3.1	10.4	10.4	0.0	56.2	1.0	1.0	2.1	14.6
Scientific	44	0.0	0.0	0.0	2.3	77.3	0.0	0.0	9.1	0.0	4.5	2.3	4.5
Electron. and Maint. Engr. Area	826	2.1	5.2	0.0	4.0	4.1	10.3	6.3	18.4	11.0	9.4	5.0	24.2
Communications-Electronics	413	4.1	1.9	0.0	2.9	5.1	8.0	7.0	31.0	7.0	7.0	5.1	22.0
Avionics	137	0.0	3.6	0.0	7.3	7.3	35.8	10.9	13.1	0.7	3.6	0.7	16.1
Civil Engineering Area	169	0.6	7.1	52.1	3.6	0.6	0.0	0.6	0.6	0.6	1.2	1.2	31.8
Materiel Area	838	0.0	0.0	0.0	0.7	0.0	0.4	1.0	1.0	59.7	22.4	1.4	13.4
Supply Services	658	0.0	0.0	0.0	0.3	0.0	0.3	1.1	1.1	59.0	24.0	1.1	13.1
Comptroller Area	50	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	78.0	4.0	0.0	14.0
Personnel Resources Mgt. Area	1,269	0.0	0.0	0.4	0.4	0.4	0.4	1.0	0.4	26.0	26.0	9.0	36.0
Information Area	93	0.0	0.0	0.0	0.0	0.0	0.0	2.2	0.0	7.5	9.7	3.2	77.4
Intelligence Area	165	0.0	0.0	0.0	0.0	0.0	1.2	1.2	1.2	4.2	38.8	4.2	49.2
Security Police Area	150	0.0	0.0	0.0	0.0	0.0	1.3	1.3	0.0	5.3	66.7	4.0	21.4

[a]Based on subsamples of OTS graduates in 1963.

XIII. RELIABILITY AND INTERCORRELATIONS

Reliability is a term covering several different but related testing concepts pertaining to the consistency with which a test yields measurements. Each concept has experimental procedures associated with it for determining reliability in a specific sense. One of these concepts is concerned with the equivalence of measurements. Equivalence is shown either by administering alternate forms of a test to a group of examinees on a single occasion and correlating the two sets of scores, or by splitting one form of a test into segments which can be treated as alternate forms. A refinement of the latter method is to split the test into its constituent items and to analyze these into reliable and unreliable components.

Another concept of reliability is concerned with stability of measurements. Stability is determined by administering a test form to a group of examinees on two occasions and correlating the resulting sets of scores. The most stringent test of reliability is to administer one form to a group of examinees and, on a later occasion, to administer an alternate form and correlate the scores. This method yields a coefficient of stability and equivalence. Such a coefficient is characteristically lower than that obtained by other methods.

Reliability data are of great value at certain stages in the development of a new test because they give indications of whether a test or subtest is worth further development. In test interpretation, reliability data are useful mainly in clarifying limits beyond which there is no evidence to support the interpretation. Reliability data also determine the limits of validity. Like validity, reliability decreases as the range of test scores is restricted. Undistorted measures of reliability can be obtained only from samples for which the test is wholly appropriate.

Not all concepts of reliability are applicable to all tests. Using only the appropriate methods, AFOQT subtest reliabilities were computed on samples of over 400 student officers. Based on these data, composite reliabilities were computed by the Wherry and Gaylord formula for the reliability of a composite from its components. The results are presented in Table 10 as coefficients of equivalence, but for composites containing speeded subtests they are not pure examples of this type of reliability. The coefficients of stability and equivalence in the same table represent correlations between scores on one form of the AFOQT and a different form administered about three years later to a sample of 415 AFROTC cadets.

Table 10. **Reliability of AFOQT Composites[a]**

Composite	Coefficient of Equivalence	Coefficient of Stability and Equivalence	Standard Error of Measurement
Pilot	.91	.71	6.7
Navigator-Technical	.95	.90	4.5
Officer Quality	.94	.84	3.3
Verbal	.89		2.8
Quantitative	.93		1.8

[a]Based on various groups specified in the text. Sample sizes are 415 or more.

Table 10 also contains a different type of reliability data. This is a measure of precision known as the standard error of measurement. It is actually an estimate of the variability in a distribution of test scores obtained from repeated applications of the test to an examinee. It expresses by how much an examinee's score may be expected to vary on repeated testing. The interpretation is that the score will lie within one standard error of the true score, taken as the average on repeated testing, on approximately two occasions out of three, and within three standard errors on virtually every occasion. Standard errors in Table 10 are in raw score form.

By indicating the precision of measurement, the standard error provides a basis for confidence in whether different scores for two examinees on the same composite represent an actual difference in aptitude or the same aptitude save for unreliability of measurement. A related question for which the standard error has relevance is whether different scores for the same examinee on different composites represent actual differences in aptitude. This question can be approached in another manner with the aid of the reliability coefficients and intercorrelations of the AFOQT composites. The intercorrelations are shown in Table 11.

Table 11. **Intercorrelation of AFOQT Composites[a]**

Composite	Pilot	Nav-Tech	Off Qual	Verbal
Navigator-Technical	.69			
Officer Quality	.38	.66		
Verbal	.23	.37	.71	
Quantitative	.44	.81	.74	.38

[a]Based on 39,545 examinees tested in 1967 for all programs except AFROTC.

Whether high or low intercorrelations of composites are desired depends on their purpose. For the AFOQT it is desired that the intercorrelations be low because the composites are not intended to measure the same aptitudes. On the other hand, composites with subtests in common will tend to correlate substantially just because of these common elements. Five of the ten correlations in the table are between composites having subtests in common. These correlations are moderately high. The remaining correlations are sufficiently low to support the statement that the composites are not measuring the same aptitudes to any marked extent.

Special methods exist for obtaining coefficients between a part and a remainder, and between variables from which the effects of one or more other variables have been excluded. Using these methods, the intercorrelations of the AFOQT composites were recomputed with the effects of overlapping subtests deleted. The results are shown in Table 12. These are not necessarily correlations between composites as they are actually constituted, but they express the degree of independence of the composites without the elevating effects of their common elements. The deletion results in a drop in mean intercorrelation from .57 to .35.

Table 12. **Intercorrelation of AFOQT Composites with Effects of Common Subtests Deleted[a]**

Composite	Pilot	Nav-Tech	Off Qual	Verbal
Navigator-Technical	.36			
Officer Quality	.38	.15		
Verbal	.23	.37	.35	
Quantitative	.44	.56	.26	.38

[a]Correlations computed from basic data in Table 11.

Using the data in Table 11 and the Wherry and Gaylord reliabilities of the composites, it is possible to estimate the proportion of score differences in excess of chance between any two composites. The proportions are given in Table 13. An illustration of interpretation of this table is that obtained raw score differences between the Pilot and Navigator-Technical composites represent actual differences in aptitude levels in 34 instances out of 100. While it is desired that the proportions be as high as possible, the proportions in the table are sufficient to permit cautious use of the test in this way. The minimum value for a useful proportion is about .25.

Raw score means and standard deviations of the composites are included in Table 13. These are estimated from published conversion tables and are strictly applicable only to Form 68, but other recent forms yield fairly similar data. Where raw composites are added together to yield a simple sum for use in qualifying examinees, the weight of each composite in the total is proportional to its standard deviation. Usually however, such sums are based on percentiles as a matter of convenience. In this case, all composites are weighted about equally because in unselected samples all means in percentile form are near 50 and all standard deviations are near 30.

Table 13. **Proportion of AFOQT Raw Score Differences in Excess of Chance[a]**

Composite	Pilot	Nav-Tech	Off Qual	Verbal	Mean	SD
Pilot					115.5	22.4
Navigator-Technical	.34				115.5	20.4
Officer Quality	.46	.38			114.5	13.6
Verbal	.46	.46	.31		40.5	8.6
Quantitative	.45	.28	.34	.43	39.5	6.8

[a]Proportions estimated from coefficients of equivalence in Table 10 and intercorrelations in Table 11.

If weights other than those determined by the standard deviations are desired, these can be established by multiple linear regression analysis. Where data are insufficient for this analysis, recourse may be had to professional judgment. In this case, however, it is impossible to specify precisely how the weights were derived, and it has frequently been shown that such weights do not yield optimal prediction of a criterion. The application of weights which are not determined by the distributions themselves introduces several extra steps in the scoring process which are best avoided in a decentralized testing program.

XIV. SCORE DISTRIBUTIONS

If any AFOQT composite is administered to a large number of examinees for whom it is appropriate, the raw score most frequently encountered will be near the mean of the group, and the least frequently encountered raw scores will be at the extremes. If raw scores are shown on the horizontal axis and frequencies on the vertical axis, a figure is generated which closely approximates Figure 6. Figure 6 is the normal probability curve and is defined by an equation. Many sets of psychological and biological data assume the form of this curve, and it is therefore a useful model for representing such data. Properties of the data can be understood from the known properties of the curve.

In a normal distribution, the mean score is so located that half the cases lie above it. Hence it can also be taken as the median score. The partition of the distribution at this point is shown in Figure 6. Other partitions are shown at one, two, and three standard deviations above and below the mean, and the percentages of the total area under the curve and between the partitions are indicated. These percentages also represent the proportions of the total number of cases in the distribution lying within these areas.

There are definite mathematical relationships between these properties of the curve and the percentile scale used for the AFOQT. The percentile scale is shown below the curve in Figure 6. Each interval of the scale includes 5 percent of the area under the curve. The intervals are spaced more closely near the mean to preserve this relationship. Contrary to the case of raw score distributions, a distribution of percentile scores has a rectangular shape with the same frequency at each interval.

AFOQT scores were formerly expressed as stanines. This term refers to a scale belonging to a class known as standard score scales. Stanines serve, as do percentiles, to permit meaningful interpretation of test performance. Though no longer used, stanines are still frequently encountered in personnel records. The stanine scale is included in Figure 6 to illustrate its relationships to the percentile scale and the standard deviation of the raw score distribution. Frequencies in the intervals of the stanine scale are unequal.

The AFOQT is an appropriate test for officers and candidates for programs leading to a commission. It is only in these groups and others with approximately the same aptitude distributions that the distribution of AFOQT percentiles has a rectangular form. The appropriateness of the Officer Quality composite for candidate and officer groups representing all sources of commission combined is shown in Table 14. The rectangular form is shown by the presence of roughly 5 percent of the cases at each percentile level. The officer group, however, has a greater concentration of scores in the upper ranges. This feature illustrates the difference between unselected examinees and examinees who have attained commissioned status.

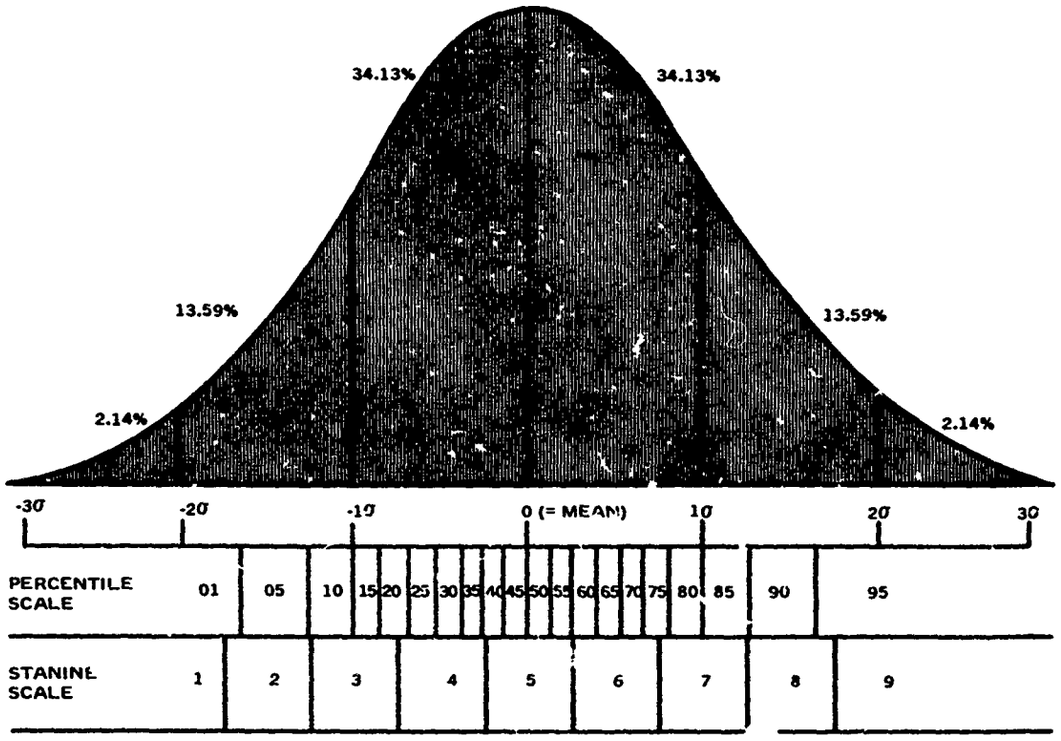

Fig. 6. **Air Force scoring systems in relation to the normal probability curve.**

Table 14. **Officer Quality Score Distributions for Candidates for Commissioning Programs and Commissioned Officers[a]**

Percentile	Percent of All Candidates at Each Percentile	Percent of Officers at Each Percentile	Percent of Qualified OTS Candidates at Each Percentile
95	3.2	9.4	5.8
90	6.2	8.1	5.9
85	6.9	5.9	7.0
80	6.5	6.3	6.4
75	4.6	6.8	6.4
70	4.7	5.4	6.4
65	4.8	5.1	5.9
60	5.2	5.1	7.1
55	3.6	5.1	5.6
50	3.6	4.8	5.3
45	3.7	5.0	5.8
40	5.4	4.3	8.0
35	5.2	4.2	6.9
30	5.1	4.8	8.2
25	5.0	4.5	9.3
20	4.3	4.2	
15	4.8	5.0	
10	5.9	5.9	
05	4.7		
01	6.6		

[a]Sample sizes are 40,302 for all candidates, 36,625 for officers, and 4,239 for qualified OTS candidates.

The third group in the table consists of examinees at an intermediate stage of selection. These are qualified candidates for OTS. For this group and the officer group, no scores are shown below those which are minimally qualifying. Some cases were found in the raw data below minimum levels, but these were ignored for purposes of the table. The three groups in the table are independently defined. They do not represent the progression of any single group through the selection process to a commission.

Differences in score distributions for appropriate and inappropriate groups are shown in Table 15 for the Pilot composite. This composite is appropriate for the Academy and AFROTC groups, and their score distributions have the rectangular form. In the basic airman group, nearly half the cases fall in the bottom percentile. This group did not contain examinees with Armed Forces Qualifying Test (AFQT) percentiles below the 21st. An even greater skewness would be seen if the full range of AFQT scores were included. The observed skewness is typical of distributions where the test is too difficult. Had the test been too easy, there would have been skewness in the opposite direction.

Table 15. **Pilot Composite Score Distributions for Appropriate and Inappropriate Groups[a]**

Percentile	Percent of Air Force Academy Candidates at Each Percentile	Percent of Advanced AFROTC Candidates at Each Percentile	Percent of Basic Airmen at Each Percentile
95	4.6	4.2	0.9
90	4.5	3.7	0.7
85	5.1	4.1	0.8
80	5.3	4.5	0.8
75	4.9	4.4	0.9
70	5.8	5.0	0.8
65	4.8	4.8	1.2
60	5.3	4.8	1.4
55	4.6	4.1	1.9
50	4.0	4.2	2.3
45	4.6	4.6	2.0
40	5.3	5.4	2.0
35	5.5	5.4	2.3
30	4.9	5.4	3.7
25	4.8	4.2	3.6
20	5.7	5.9	4.1
15	4.9	5.6	6.2
10	5.0	5.4	7.3
05	4.9	5.8	14.5
01	5.4	8.5	42.6

[a]Sample sizes are 5,105 for Academy candidates, 15,600 for AFROTC candidates, and 2,489 for basic airmen.

One observation to be made on the score distribution of a too easy or too difficult test is that the normal model does not apply. Another is that the test distinguishes the various aptitude levels within the examinee group very poorly. It is certain that there is a fairly wide range of aptitude within the large group of airmen lumped together in the bottom percentile of Table 15, but the test is insensitive to this.

It has been shown that the ideal difficulty level of a test in relation to the group for which it is intended is such that the item of median difficulty is answered correctly by 50 percent of the group, while at the same time there is a wide range of difficulty among the other items. The range of difficulty and median difficulty of items in each AFOQT composite are shown in Table 16. Entries in the table are proportions of a group of student officers who answered the items correctly. Biographical items are not included because the concept of difficulty has a somewhat special meaning for them.

Table 16. **Difficulty Level of AFOQT Composites[a]**

Composite	Range	Median
Pilot	.20 - .85	.55
Navigator-Technical	.19 - .92	.54
Officer Quality	.18 - .85	.54
Verbal	.18 - .85	.53
Quantitative	.19 - .84	.54

[a]Based on samples of 400 or more student officers.

XV. STANDARDIZATION

In some testing situations it is desirable to construct new percentile scales based on various raw score distributions as they become available. However, uniformity of meaning of AFOQT scores regardless of time or place of collection requires that a single reference group, defined in advance, be the basis of all AFOQT percentiles. Before release for operational use, each new form of the AFOQT is standardized with respect to this group. The process of standardization consists essentially of the development of norm or conversion tables by which raw scores are converted to percentiles for the reference group. This group must be representative of groups on which the test will be used in practice.

A group composed of candidates for admission to the Air Force Academy was used for standardization through almost the whole history of the AFOQT. Following the standardization of Form G, however, this group ceased to be available for the purpose. In anticipation of this development, a method was devised to permit indirect establishment of relationships between new forms of the AFOQT and a prior group of Air Force Academy candidates.

The method involved administering AFOQT Form G to a large sample of basic airmen stratified by AFQT decile in the range of the 21st through the 100th percentile. Also administered to the same group at approximately the same time was the entire battery of Project TALENT tests. These tests had been used for a national survey of aptitudes and abilities in a sample of over 400,000 youth of high school age. By multiple linear regression methods it was possible to define groups of TALENT tests which gave the best available prediction of each AFOQT composite. Thus a TALENT composite corresponding to each AFOQT composite was defined.

The next step consisted of making conversions from the AFOQT Form G percentiles to the appropriate TALENT composite score distributions. The score on the TALENT composite which cut off the same proportion of the sample as a given Form G percentile was treated as representing that percentile. In this way percentiles were established in the TALENT composite distributions with the same meaning as the Form G percentiles. Utilizing these relationships, the process of standardizing a new form of the AFOQT is accomplished as follows:

1. Each new AFOQT composite is administered along with the tests of the corresponding TALENT composite to approximately 1,000 basic airmen stratified by AFQT decile in the range of the 21st through the 100th percentile. Only high school graduates are included in this sample.

2. The new AFOQT composite is scored in the usual manner and the scores are distributed. The TALENT tests are scored and combined to yield the corresponding TALENT composite scores. These scores are also distributed.

3. Conversions are made between the known percentile levels in the TALENT composite distribution and the new AFOQT composite distribution. This step yields percentile norms for the new AFOQT composite.

The inappropriateness of the AFOQT for basic airmen is not an obstacle to this standardization process because the standardization is not actually based on the airman sample. The small frequencies at the upper ranges of the percentile scale for this sample can lead to some instability in the placement of the

upper percentiles. However, these are not the levels where critical decisions are made in practice. Currently, the highest minimum qualifying score in any program is the 60th percentile, and most minimum qualifying scores are much lower.

The tests in each TALENT composite, together with the integral score weights used in computing the composite scores, are shown in Table 17. The titles of the tests are fairly descriptive of their content and help to provide further insights into what is involved in aptitudes measured by the AFOQT. The tests listed as constituting the Academic composite are used in standardizing the AFOQT Officer Quality composite.

Table 17. **Composition of TALENT Composites Corresponding to AFOQT Composites[a]**

TALENT Composite	TALENT Test	Weight
Pilot	Aeronautics and Space (Information)	3
	Mechanical Reasoning	3
	Mechanics (Information)	3
	Advanced Mathematics	2
	Visualization in Three Dimensions	2
	Electricity and Electronics (Information)	1
	Visualization in Two Dimensions	1
Navigator-Technical	Introductory Mathematics	3
	Mathematics (Information)	3
	Mechanical Reasoning	3
	Visualization in Three Dimensions	3
	Electricity and Electronics (Information)	2
Academic	Advanced Mathematics	3
	Aeronautics and Space (Information)	2
	Introductory Mathematics	2
	Mathematics (Information)	2
	Reading Comprehension	1
Verbal	Aeronautics and Space (Information)	3
	Literature (Information)	2
	Mathematics (Information)	2
	Vocabulary (Information)	2
	Reading Comprehension	1
Quantitative	Advanced Mathematics	3
	Introductory Mathematics	2
	Mathematics (Information)	2

[a]Data extracted from Dailey et al., 1962, and unpublished supplement thereto.

The effectiveness of this indirect standardization procedure depends on the existence of high correlations between the AFOQT composites and the corresponding TALENT composites. These correlations are presented in Table 18, based on the sample of basic airmen on which the TALENT composites were originally developed.

Since each AFOQT form is standardized by referring back to the original TALENT composite distributions, an unchanging normative base is achieved which permits direct comparisons of scores on successive AFOQT forms. The stratification of the standardization groups permits comparison of any AFOQT composite with any other. The normative base continues in an indirect manner to be the Air Force Academy candidate group. Moreover, AFOQT scores can be related to the 12th grade Project TALENT sample from the national survey if desired.

Table 18. **Correlation between AFOQT Composites and TALENT Composites[a]**

AFOQT Composite	Correlation with Corresponding TALENT Composite
Pilot	.80
Navigator-Technical	.88
Officer Quality	.86
Verbal	.83
Quantitative	.82

[a]Based on 2,489 basic airmen on which TALENT composites were developed.

AFOQT scores of 12th grade males in a subsample of the Project TALENT national sample are shown in Table 19. The performance of this group is expressed as the percentage of cases attaining or exceeding given AFOQT percentile scores on each composite. The table has manpower implications. It can be seen, for example, that 19 percent of this group could qualify for admission to a program leading to a commission if the minimum qualifying score on Officer Quality is set at the 25th percentile. In practice, the minimum would probably be set much higher for examinees who do not meet current educational requirements.

Table 19. **Performance of 12th Grade Males on the AFOQT[a]**

	Percent of Cases at or above Percentile				
Percentile	Pilot	Navi-Tech	Officer Quality	Verbal	Quant.
95	2.4	2.1	1.1	2.8	1.2
90	3.6	2.5	2.1	4.0	1.7
85	4.4	2.9	2.7	6.0	2.0
80	5.7	3.8	3.2	6.7	2.7
75	6.6	4.4	4.2	7.5	3.2
70	7.4	5.0	4.7	10.0	3.6
65	8.7	5.8	5.5	11.3	4.0
60	10.0	6.8	6.5	12.0	5.0
55	12.7	8.0	7.3	13.0	6.0
50	15.0	8.7	8.3	14.0	7.0
45	18.0	10.0	10.0	16.0	8.0
40	20.5	13.0	11.0	18.0	10.0
35	23.5	15.5	12.5	21.0	13.0
30	27.0	18.0	14.7	24.0	15.0
25	31.0	21.0	19.0	27.0	19.0
20	35.0	27.0	24.0	31.0	22.0
15	43.0	32.0	30.0	36.0	27.0
10	51.0	41.0	41.0	45.0	35.0
05	66.0	56.0	55.0	59.0	55.0
01	100.0	100.0	100.0	100.0	100.0

[a]Based on a 4 percent subsample of 12th grade males in the Project TALENT study. Subsample size is 2,403.

Because of the continuing role of the Academy candidate group in the standardization of the AFOQT, the meaning of AFOQT scores is enhanced by an understanding of the characteristics of this group. The specific sample used in standardizing Form G and subsequent forms consists of 5,105 candidates for the class of 1964. Of this group, 773 were ultimately selected for admission. The group proved to be highly self-selected, however, particularly with respect to quantitative aptitude. This is evidenced by the distribution statistics of the group on the two CEEB aptitude tests. These are shown in Table 20. Means and standard deviations of these tests usually approximate 500 and 100, respectively.

Table 20. **CEEB Cumulative Distributions and Distribution Statistics for the AFOQT Standardization Group[a]**

CEEB Verbal Aptitude Score	Percent of Cases at or above Score	CEEB Mathematics Aptitude Score	Percent of Cases at or above Score
800	0.0	800	0.1
750	0.2	750	2.5
700	1.8	700	10.4
650	8.2	650	24.8
600	20.6	600	47.2
550	36.9	550	66.6
500	55.9	500	82.8
450	74.9	450	91.4
400	87.6	400	96.8
350	94.8	350	98.8
300	98.5	300	99.7
250	99.8	250	100.0
200	100.0	200	100.0
Mean	514.2	Mean	585.5
SD	96.1	SD	93.4

[a]Based on 5,105 candidates for the Air Force Academy class of 1964.

It seemed at least possible that an AFOQT form based on a standardization sample having very high quantitative aptitude would prove excessively difficult when used outside the Academy setting. Corrections were therefore applied to all composites by equating them with CEEB scores in an earlier and less highly self-selected candidate group. The corrections, however, tended to make some of the composites too easy for most groups to which the test was applied. The corrections were therefore removed, beginning with AFOQT Form 64, and the rectangular percentile distributions of AFOQT composites were restored.

XVI. ADJUSTMENT FOR EDUCATIONAL EFFECTS

It has long been known that the effects of formal education on AFOQT scores are to raise them appreciably. Moreover, these effects for the most part do not appear to be spurious. Since the AFOQT is administered to examinees with widely different educational levels in different programs, it follows that a given percentile can not have the same meaning in all programs.

Evaluation of the extent of these educational effects proved to be very difficult in practice. Lacking this evaluation, educational effects were dealt with by imposing lower minimum qualifying scores in programs where testing is done early in college than in programs where testing is done near graduation. This solution made for roughly equivalent minimum aptitude levels in the various programs, but it also produced depressed score distributions for some commissioning sources and tended to confound research data when studies were attempted across sources.

Recently it became possible to perform two independent studies in which the extent of educational effects could be determined initially. The two were of quite different design but yielded similar results. In one, the AFOQT was administered to AFROTC cadets as freshmen and as seniors. In the other, the Department of Defense Officer Record Examination and flying deficiency elimination rates were used as controls to permit a comparison of scores of AFROTC freshmen and OTS candidates tested near graduation from college.

Results from the latter study are illustrated in Table 21. The table is an adaptation of conversion tables for AFROTC and OTS groups who have been equated on the control variables. Both groups are heterogeneous with respect to type of college and major field of study, and they represent a difference of about three years in educational level. An example of reading the table is that a Pilot raw score of 133 represents the same degree of pilot aptitude in the AFROTC program as a raw score of 177 in the OTS program, and that this degree of aptitude exceeds that of 90 percent of the examinees for whom the test is appropriate. There is evidence that educational effects on the pilot composite are greatest for those entering pilot training.

Table 21. **Effects of Education on AFOQT Scores[a]**

	Raw Scores									
	Pilot		Navigator-Technical		Officer Quality		Verbal		Quantitative	
Percentile	AFROTC	OTS	AFROTC	OTS	AFROTC	OTS	AFROTC	OTS	AFROTC	OTS
90-95	133 & above	177 & above	144 & above	160 & above	133 & above	139 & above	51 & above	56 & above	53 & above	57 & above
80-85	121-132	163-176	128-143	147-159	128-132	135-138	47-50	52-55	51-52	55-56
70-75	113-120	149-162	120-127	139-146	122-127	131-134	43-46	50-51	49-50	53-54
60-65	105-112	133-148	112-119	132-138	116-121	127-130	40-42	48-49	46-48	51-52
50-55	98-104	121-132	104-111	126-131	112-115	123-126	38-39	46-47	42-45	49-50
40-45	91-97	113-120	98-103	118-125	108-111	119-122	36-37	44-45	38-41	47-48
30-35	83-90	105-112	91-97	110-117	104-107	115-118	32-35	40-43	34-37	45-46
20-25	75-82	98-104	83-90	102-109	96-103	111-114	28-31	36-39	30-33	41-44
10-15	63-74	91-97	72-82	91-101	87-95	103-110	24-27	31-35	26-29	34-40
01-05	62 & below	90 & below	71 & below	90 & below	86 & below	102 & below	23 & below	30 & below	25 & below	33 & below

[a]Based on AFOQT Form 68 conversion tables.

In general, three years of college has the effect of increasing the percentile score by roughly 5 to 30 points, depending on the composite being considered and the level of the initial score. Pending the accumulation of additional data, it is recommended that examinees with intermediate amounts of education at the time of testing be evaluated on a third set of conversion tables which reflects half of the difference between the AFROTC and OTS tables. For example, a raw Pilot score of 155 for such an examinee should fall at the lower limit of the 90th percentile.

The AFOQT now incorporates into its scoring manual a set of multiple conversion tables based on AFROTC, OTS, and intermediate educational levels. In general, each table is for use with any examinee whose educational level at the time of testing is appropriate for that table. Some increase in disqualification rates follows from the introduction of intermediate and OTS tables, but mean aptitude levels of qualified examinees are also increased, and percentiles are given the same meaning in all programs.

XVII. MINIMUM QUALIFYING SCORES

Minimum qualifying scores are essential to a testing program if aptitude standards are to be maintained uniformly over a period of time. Minimum qualifying scores are a part of the program and not necessarily built into the test itself. In the case of Air Force tests, minimum qualifying scores are established by Headquarters, United States Air Force, and are promulgated by directive. Such scores are currently set on one or more composites in nearly all programs for which the AFOQT is used. Only the Verbal and Quantitative composites have no minimum qualifying scores for any program.

Minimum qualifying scores are not the same in all programs, and they are subject to change at any time. Changes are made in accordance with the availability of applicants for the various programs and the needs of the Air Force. Where there are many applicants to fill a small quota, minimum qualifying scores may be set very high. If the need for personnel to fill a quota is such that most applicants must be accepted, minimum qualifying scores must be set low. In this case, applicants with mediocre or borderline aptitudes are entered into the program, and it can be expected that the elimination rate will rise.

The effects of varying the minimum qualifying scores can be predicted from expectancy tables. These may be based on empirical data or worked out theoretically. In either case, the tables permit evaluation of the numbers and characteristics of selectees to be expected with any minimum qualifying score or combination of scores. If current elimination data are available, the tables can be constructed to show also the number of graduates which any qualified applicant group will yield.

Tables 22 and 23 illustrate the process. These tables were developed theoretically on the basis of data from an empirical validation study. Table 22 represents the selection of undergraduate student pilots where minimum qualifying scores are set on both the Pilot and Navigator-Technical composites. Horizontal and vertical lines drawn through the table represent minimum qualifying scores, each arbitrarily set at the 30th percentile. By altering the location of the lines, the effects on inputs to the pilot training program can be observed.

Table 22. **Pilot and Navigator-Technical Score Distributions for 1,000 Unselected Candidates for Pilot Training[a]**

Pilot Percentile	Navigator-Technical Percentile										
	01-05	10-15	20-25	30-35	40-45	50-55	60-65	70-75	80-85	90-95	Total
90-95	0	0	1	2	4	6	10	14	23	39	99
80-85	0	2	3	5	8	10	13	16	20	21	98
70-75	1	3	6	8	10	12	14	16	16	13	99
60-65	2	5	8	10	12	13	14	14	13	9	100
50-55	4	8	10	12	13	13	13	12	10	6	101
40-45	6	10	12	13	13	13	12	10	8	4	101
30-35	9	13	14	14	13	12	10	8	5	2	100
20-25	13	16	16	14	12	10	8	6	3	1	99
10-15	21	20	16	13	10	8	5	3	2	0	98
01-05	39	23	14	10	6	4	2	1	0	0	99
Total	95	100	100	101	101	101	101	100	100	95	994

[a]Theoretical data based on a correlation of .69 between tests. The actual number of cases is 994 because of commulative rounding errors.

Table 23 shows the expected number of graduates from the examinees in Table 22. Neither the minimum qualifying scores nor the elimination rate in Table 23 will necessarily apply in practice. Hence the table is illustrative only. From a table of this kind, however, the number of graduates per 1,000 examinees can be determined for any combination of minimum qualifying scores on tests with known validities and intercorrelations, and for any elimination rate.

Table 23. **Pilot and Navigator-Technical Score Distributions for Graduates from 1,000 Candidates for Pilot Training[a]**

Pilot Percentile	Navigator-Technical Percentile										
	01-05	10-15	20-25	30-35	40-45	50-55	60-65	70-75	80-85	90-95	Total
90-95	0	0	1	2	4	6	9	13	21	36	92
80-85	0	2	3	4	7	9	11	14	17	18	85
70-75	1	2	5	7	8	10	11	13	13	11	81
60-65	2	4	6	8	9	10	11	11	10	7	78
50-55	3	6	7	9	10	10	10	9	7	4	75
40-45	4	7	8	9	9	9	8	7	6	3	70
30-35	6	8	9	9	8	8	6	5	3	1	63
20-25	8	10	10	8	7	6	5	4	2	1	61
10-15	11	11	8	7	5	4	3	2	1	0	52
01-05	16	9	6	4	2	2	1	0	0	0	40
Total	51	59	63	67	69	74	75	78	80	81	697

[a]Theoretical data based on a Pilot validity of .40 and an elimination rate of .21 in the qualified group.

Tables 22 and 23 can be used to extract the probability of successful completion of training with any combination of test scores. The probability, for example, is .64 at the minimum qualifying score shown for both tests, and it increases to .92 at the highest score levels. A summary of the effectiveness of this pilot selection system with minimum qualifying scores as shown is that, while 21 percent of the selectees were eliminated from training, 43 percent of the rejected group would have been eliminated had this group been allowed to enter the program.

REFERENCES

This list of references does not constitute a complete bibliography of AFOQT publications. References which deal with the AFOQT only incidentally or which deal with tests related to the AFOQT have been excluded unless they contain data used in this report. Unpublished reports and reports published before 1960 have also been excluded.

AFM 35-8, *Air Force Military Personnel Testing Manual.* Washington: Department of the Air Force, 7 May 1968.

Dailey, J.T., Shaycoft, Marion F., & Orr, D.B. *Calibration of Air Force selection tests to Project TALENT norms.* PRL-TDR-62-6, AD-285 185. Lackland Air Force Base, Texas: Personnel Research Laboratory, Aerospace Medical Division, May 1962.

Germain, J. "Validity of the U.S. Aircrew Classification Battery in a sample of Spanish pilots." In Geldard, F.A., & Lee, Marilyn C. (Eds), *Proceedings, First International Symposium on Military Psychology.* Publication 894. Washington: National Academy of Sciences-National Research Council, 1961.

Gregg, G. *The effect of maturation and education experience on Air Force Officer Qualifying Test scores.* AFHRL-TR-68-107, AD-687 089. Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, Personnel Research Division, July 1968.

Miller, R.E. *Prediction of technical training criteria from AFOQT composites.* WADD-TN-60-215, AD-246 658. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, September 1960.

Miller, R.E. *Predicting first year achievement of Air Force Academy cadets, class of 1964.* PRL-TDR-64-18, AD-448 480. Lackland Air Force Base, Texas: Personnel Research Laboratory, Aerospace Medical Division, July 1964.

Miller, R.E. *Development of officer selection and classification tests-1966.* PRL-TR-66-5, AD-639 237. Lackland Air Force Base, Texas: Personnel Research Laboratory, Aerospace Medical Division, June 1966.

Miller, R.E. *Relationship of AFOQT scores to measures of success in undergraduate pilot and navigator training.* PRL-TR-66-14, AD-656 303. Lackland Air Force Base, Texas: Personnel Research Laboratory, Aerospace Medical Division, October 1966.

Miller, R.E. *Development of officer selection and classification tests-1968.* AFHRL-TR-68-104, AD-679 989. Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, Personnel Research Division, July 1968.

Miller, R.E. & Valentine, L.D., Jr. *Development and standardization of the Air Force Officer Qualifying Test-64.* PRL-TDR-64-6, AD-600 782. Lackland Air Force Base, Texas: Personnel Research Laboratory, Aerospace Medical Division, March 1964.

Valentine, L.D., Jr. & Creager, J.A. *Officer selection and classification tests: their development and use.* ASD-TN-61-145, AD-269 827. Lackland Air Force Base, Texas: Personnel Laboratory, Aeronautical Systems Division, October 1961.

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)* Personnel Research Division Lackland Air Force Base, Texas 78236	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3. REPORT TITLE

INTERPRETATION AND UTILIZATION OF SCORES ON THE AIR FORCE OFFICER QUALIFYING TEST

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

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

Robert E. Miller

6. REPORT DATE May 1969	7a. TOTAL NO. OF PAGES 29	7b. NO. OF REFS 11
8a. CONTRACT OR GRANT NO b. PROJECT NO 7717 c. Task 771706 d.	9a. ORIGINATOR'S REPORT NUMBER(S) AFHRL-TR-69-103 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

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

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Personnel Research Division Lackland AFB, Texas 78236

13. ABSTRACT

This report summarizes a large body of data relevant to the proper interpretation and use of aptitude scores on the Air Force Officer Qualifying Test. Included are descriptions of the AFOQT testing program and the general characteristics of the test itself. Technical concepts are introduced by a brief explanation to assist users of AFOQT scores who are not test specialists. Technical data include an extensive sampling of validation studies covering prediction of success in pilot training, navigator training, technical training, and academic courses. Relationships to other well known tests and the Air Force structure of career areas and utilization fields are indicated. Several types of reliability data are presented, together with intercorrelations of the aptitude composites both with and without the elevating effects of overlapping subtests. The Air Force percentile scoring system is discussed in relation to the normal probability curve and the stanine scale. Score distributions are provided for officers, candidates for programs leading to a commission, basic airmen, and 12th grade males. Procedures used in standardizing new forms of the AFOQT through the Project TALENT aptitude composites are described, including operations which maintain relationships with Air Force Academy candidates and the TALENT national sample. Effects of applying minimum qualifying scores and adjustments for level of formal education at the time of testing are explained.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Air Force Officer Qualifying Test						
aptitudes						
classification tests						
difficulty						
distributions						
educational effects						
intercorrelations						
minimum qualifying scores						
officer selection and classification						
percentile scale						
Project TALENT						
reliability						
selection tests						
standardization						
test construction						
validity						

Unclassified
Security Classification