

55-17

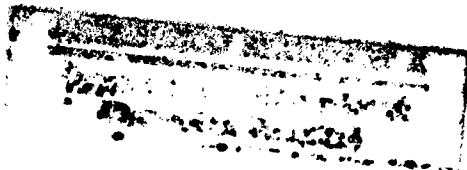
Research

REPORT

AFPTRC - TN - 55 - 17, July 1955

The Development and Validation Of MDAP English Proficiency Examination Form A

By John A. Cox, Jr.



19980923 123

AIR FORCE PERSONNEL & TRAINING RESEARCH CENTER

LACKLAND AIR FORCE BASE • SAN ANTONIO • TEXAS

HEADQUARTERS
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER

Air Research and Development Command
Lackland Air Force Base, Texas

Col Herbert N. Cowles
Commander

Dr. Arthur W. Melton
Technical Director

Dr. Charles W. Bray
Deputy Chief of Staff
For Operations

Dr. Henry S. Odbert
Chief
Technical Information Division

Initial distribution of this report has been made by the Air Force Personnel and Training Research Center. No further copies are available at the Center for distribution. Copies may be obtained in the manner specified below. All such requests for publications should specify author(s), title, date, and serial number (such as AFPTRC-TN-55-1).

Department of Defense agencies and their contractors should address requests for additional copies to: The Armed Services Technical Information Agency, Document Service Center, Knott Building, Dayton 2, Ohio. Documents should be requested from ASTIA by identification number AD, ATI, or TIP when known.

Other organizations and individuals may obtain information on cost of microfilm or photostatic copies from: Office of Technical Services, U.S. Department of Commerce, Washington 25, D.C.

Copies may also be obtained on loan through local libraries from any of the depository libraries of the Center. A list of these libraries may be obtained from: Commander, Air Force Personnel and Training Research Center, ATTN: PTDT, Post Office Box 1557, Lackland Air Force Base, San Antonio, Texas.

THE DEVELOPMENT AND VALIDATION
OF MDAP ENGLISH PROFICIENCY EXAMINATION FORM A

By John A. Cox, Jr.

Personnel Research Laboratory
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
Air Research and Development Command
Lackland Air Force Base, Texas

Project No. 7701
Task No. 77027

Approved by:
Lloyd G. Humphreys, Director
Personnel Research Laboratory

IMPLICATIONS

Foreign nationals attend United States Air Force pilot training schools under the Mutual Defense Assistance Program (MDAP). The MDAP English Proficiency Examination is currently being used overseas to select personnel who understand English well enough to profit from this instruction. This Research Report describes the development and the validation of the first form of this test.

An experimental version of the test was given to foreign students already in training in this country. Analysis of these data led to the selection of 90 written items (Part I) and 24 oral items (Part II) for the first operational version of the test.

Form A was administered overseas in 1953 to select personnel to attend pilot training schools in the United States. Another sample was tested experimentally at Lackland Air Force Base. In addition the test was given to basic airmen. Analysis of these data gave evidence of acceptable reliability of the test's part scores as well as its discriminating powers.

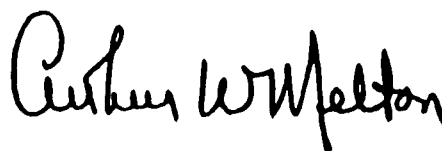
An individually-administered behavior test of English comprehension was developed as a criterion of English proficiency, and administered along with MDAP English Proficiency Examination Form A. High correlation of the English Proficiency Examination with the behavior test and with an English test used in Pre-Flight training gave assurance that it actually measures language comprehension. Predictive validity was found for Form A through correlating the test scores with success in Pre-Flight training.

Use of MDAP English Proficiency Examination Form A in selecting foreign nationals for pilot training in the United States can increase efficiency of training through entering into Pre-Flight only those candidates with enough English comprehension to have a reasonable chance to profit from the instruction. To accomplish this goal, a sufficiently large number of applicants

must be available so that a suitable qualifying level can be set in the selection test.

Since this report was prepared, Forms B and C of MDAP English Proficiency Examination have been developed.

Hq, AFPTRC
Lackland Air Force Base
San Antonio, Texas
27 June 1955



Arthur W. Melton
Technical Director



Herbert N. Cowles
Col, USAF
Commander

TABLE OF CONTENTS

	Page
List of Tables	v
List of Figures	vi
Background	1
The Experimental Test	1
Analysis of the Experimental Test	1
The Operational Test	4
Operational Administration	5
Analysis of Operational Administration	5
The United States Administration	7
Comparison of English Proficiency Scores from USAF Basic Airmen and MDAP Personnel	9
Estimates of Validity	12
The Criteria of English Comprehension	12
Behavior Test	12
English Language--Proficiency Examination No. I	16
Predictive Validity	18
Summary and Conclusions	20
References	20

LIST OF TABLES

Table	Page
1 Types of Items Used in English Proficiency Testing	2
2 Distribution Statistics for the Experimental Test by Countries	3
3 Summary of the Item Analysis of MDAP English Proficiency Examination Form A, Overseas Administration	6
4 Part-Score Distribution Statistics from the Operational Administration of MDAP English Proficiency Examination Form A	6

List of Tables (Cont.)

Table	Page
5 Summary of Item Analysis of MDAP English Proficiency Examination Form A, Lackland Administration	8
6 Test-Retest Distribution Statistics and Correlation Coefficients for MDAP English Proficiency Examination Form A.	8
7 Distribution Statistics for the Behavior Test Sample by Country	15
8 Summary of Covariance Analysis on Mean Behavior Test Scores of Groups Tested by Different Administrators	16
9 Intercorrelations and Distribution Statistics for MDAP English Proficiency Examination Part-Scores and Behavior Test Scores	17
10 Distribution Statistics and Correlation Coefficients for the MDAP English Proficiency Examination Form A and a Pre-Flight Placement Test	17
11 Number of Cases in the Disposition Categories by Country. . .	18
12 Mean Written and Oral Scores for Two Groups of MDAP Pre-Flight Students by Disposition Groups	19
13 Distribution Statistics and Correlation Coefficients Between Written and Oral English Proficiency Scores and Pass/Fail and Holdover in Pre-Flight Training Among MDAP Students . .	19

LIST OF FIGURES

Figure	Page
1 Distributions of Part I (written) scores, English Proficiency Examination, for basic airmen, MDAP students tested at Lackland Air Force Base, and foreign personnel tested overseas	10
2 Distributions of Part II (oral) scores, English Proficiency Examination, for basic airmen, MDAP students tested at Lackland Air Force Base, and foreign personnel tested overseas. . .	11
3 Apparatus for the last ten items of the behavior test. . . .	13

THE DEVELOPMENT AND VALIDATION
OF MDAP ENGLISH PROFICIENCY EXAMINATION FORM A

BACKGROUND¹

This report concerns the development and validation of a test of English comprehension. The test is intended for use with foreign nationals who are scheduled to attend United States Air Force pilot training schools.

In August 1952, the Air Force Personnel and Training Research Center was directed by the Human Factors Division, Directorate of Research and Development, Headquarters United States Air Force, to develop an English proficiency examination for selecting Mutual Defense Assistance Program (MDAP) personnel to attend pilot training schools in the United States. Information concerning the problems involved was obtained from conferences with operating personnel at Goodfellow Air Force Base.

THE EXPERIMENTAL TEST

Based upon facts learned in the above-mentioned conferences, the decision was made to use both written and oral materials. Items were constructed in areas of the Aircrew Classification Battery. An attempt was made to keep the subject content of items simple enough to insure that proper answering of the question would be a function of English proficiency.

Items were constructed in the areas of vocabulary, mechanical principles, word order, interpretation of data, arithmetic reasoning, aviation information, reading comprehension, and background for current affairs. The number of items included and their descriptions are presented in Table 1.

The experimental form of the test, entitled MDAP Pilot Selection Test, was a mimeographed booklet containing 97 written and 24 oral items. This form was administered to foreign student pilots at USAF bases during August and September 1952 by a traveling team of test administrators.

Analysis of the Experimental Test

The analysis of the first experimental administration was restricted to data obtained from students of all countries in Class 53-F and French students in Class 53-C. The French students were chosen because they were the largest group of students from any one country and provided large samples first. Furthermore, the instructors at Goodfellow Air Force Base reported that the French students varied more in their English proficiency than did the students from any other country. The written (Part I) and oral (Part II) portions of the test were analyzed separately.

¹ Manuscript received 19 April 1954.

Table 1

Types of Items Used in English Proficiency Testing

Item type	No. of items		Description
	Experi- mental form	Opera- tional form	
Part I (written)			
Vocabulary	23	22	A sentence with one word underlined. Choose the answer whose meaning is nearest that of the underlined word. Five choices.
Mechanical principles	10	8	Situational mechanical problems designed to measure understanding of mechanical forces, movements, or principles. Two to five choices.
Word order	15	15	A sentence is presented in three different word arrangements. Choose the one which states the sentence most simply
Interpretation of data	12	12	Simple bar graphs are presented. Each question asks for a conclusion based on the bar graph. Five choices.
Arithmetic reasoning	10	9	These are stated arithmetic problems. Five choices.
Aviation information	3	3	The stem asks for a bit of information such as the function of an airplane part or about a maneuver. Five choices
Reading comprehension	9	9	A paragraph is presented and followed by questions concerning the facts in the paragraph. Two to five choices.
Background for current affairs	15	12	These questions are of historical or geographical fact. Five choices.
Total Part I (written)	97	90	
Part II (oral)	24	24	Three paragraphs covering flying safety topics are read aloud. After each paragraph is read, the examinee has two minutes to answer questions about the paragraph. These questions are printed in the test booklet. The same paragraph is then read aloud again and one more minute allowed for answering questions before going on to the next paragraph. When possible a recording is used for the oral part. Three to five choices.

Analysis of Test Scores

Distribution statistics for Class 53-F were computed for each country. Using the French cases, correlations between the two part-scores were computed separately for Classes 53-F and 53-C. These data are presented in Table 2.

Table 2

Distribution Statistics for the Experimental
Test by Countries

<u>Country</u>	<u>Class</u>	<u>N</u>	<u>Part I</u>		<u>Part II</u>		<u>r(I, II)</u>
			<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	
France	53-C	93	75.4	6.7	17.9	2.3	.34
France	53-F	96	58.6	13.1	11.3	4.2	.51
Belgium	53-F	11	76.9	5.1	17.2	1.6	
Italy	53-F	26	66.9	7.8	14.6	4.4	
Holland	53-F	20	75.7	7.9	17.9	2.6	
Denmark	53-F	29	77.5	8.6	17.8	2.9	

Note.--Data on Part I are based on all 97 items of the experimental version. Part I is written; Part II is oral.

The report by instructors at Goodfellow Air Force Base that French students varied more in their ability to understand English than did students from other countries is confirmed by the distribution statistics in Table 2. The French of Class 53-F had the largest standard deviation on Part I (written) and only the Italians had a larger one on Part II (oral). The moderate relationship between the two part-scores should be noted. The difference in the interpart correlations of .51 and .34 is apparently due to the restriction in range of scores among the members of Class 53-C. Also to be noted is the higher mean score for the longer-trained group.

Comparison of the scores of Class 53-F students from various countries shows that the French students generally were lower on both parts of the test than were students from the other countries. These differences are significant beyond the .01 level. The students from Belgium, Holland, and Denmark performed equally well on the test. The French students of Class 53-C had mean scores on both parts equal to the mean scores of the students from Belgium, Holland, and Denmark in Class 53-F.

Three reasons for the differences between the Frenchmen of Classes 53-C and 53-F present themselves: (a) the French students learn some English after they arrive in the United States; (b) students weak in English are eliminated from training; (c) the French students of Class 53-C may have

been more proficient upon beginning flying training than were those of Class 53-F.

Item Analysis

Answer sheets were available for two classes of French students; Class 53-C had been in training four and one-half months, and Class 53-F had been in training two weeks. The answer sheets were scored "right answers only." Using answer sheets from Class 53-C as the upper group and from Class 53-F as the lower group, phi coefficients were computed for each item in the test. This procedure gives an indication of item validity if the hypothesis is accepted that English proficiency will improve during training.

A second item analysis was made, using data from Class 53-F only. In this case the answer sheets were split equally into upper and lower groups on the basis of the scores for Parts I (written) and II (oral). Phi coefficients and difficulty levels (percentage passing the item) were computed for each test item. As a result of these analyses, 7 of the 97 written items were omitted from the operational test.

Reliability Estimate

To provide an estimate of the reliability of the part-scores of the experimental test, Kuder-Richardson Formula 21 (3) was applied to the data from French students of Class 53-F. For Part I (written) this estimate was .89 and for Part II (oral) .71. These estimates were based on the 90 written items and 24 oral items selected for the operational form of the test.

The analysis described above completed the study of the experimental form of the test, the MDAP Pilot Selection Test. The urgent need for an operational test did not allow a validation study to be performed. However, the characteristics of the distribution statistics, part-score correlation coefficients, and reliability estimates indicated that the measuring instrument held potential as a selection device.

THE OPERATIONAL TEST

The operational form of the test, MDAP English Proficiency Examination (EPE) Form A, is composed of items from the experimental form except for seven of the written items which were omitted. Table 1 presents the number of items of each subject-matter category used in the operational test. Copy for the test booklet and manual for administration, along with scoring keys for Parts I (written) and II (oral), were submitted to higher headquarters for publication. Sound tracks were prepared on 16 mm. film for Part II (oral) of the test by the Audio-Visual Research Division of the Human Factors Operations Research

Laboratories. These sound tracks insure standard administration of the oral part of the test.

Operational Administration

During the spring of 1953, MDAP EPE Form A was administered operationally in a number of European countries under the supervision of the Military Assistance Advisory Groups. Because the 16 mm. sound tracks were not available for use in this administration, the oral paragraphs were read aloud from the manual by test administrators. The results of this testing were used to select military personnel from various participating countries to attend pilot training schools in the United States.

Analysis of the Operational Administration

Item Analysis

Upon receipt of answer sheets from the operational administration of MDAP EPE Form A, the items of each part of the test were analyzed. The sample contained 523 cases from six countries. The analysis of each part was based on the high and low 27% of the answer sheets for that part. Phi coefficients and difficulty levels were computed for each test. A summary of these analyses is presented in Table 3.

The figures in Table 3 show that certain types of items are easier than others for the population tested. Word-order items and interpretation-of-data items were among the easiest, while the aviation information items were the most difficult for this sample of potential foreign student pilots.

The mean phi coefficients in Table 3 indicate the relationship of each group of items to its appropriate score, i.e., written or oral. They are indexes of internal consistency. The reading comprehension items have the highest index of internal consistency; that is, they are the most representative of the written part as a whole. Mechanical principles, aviation information, and background for current affairs, as groups of items, are less related to the total written scores.

Analysis of Test Scores

Means and standard deviations of the score distributions from each country were computed. These distribution statistics are presented in Table 4.

The sample used for the computations in Table 3 was larger than the sample for the data in Table 4 because additional cases had been received from overseas. The figures in Table 4 show the variation in English proficiency scores among the potential student pilots by country. The French

Table 3

Summary of the Item Analysis of MDAP English Proficiency Examination
Form A, Overseas Administration

(Sample: 523 foreign student pilots; testing dates: March-April 1953)

<u>Items</u>	<u>Item type</u>	<u>Mean difficulty level</u>	<u>Mean phi coefficient</u>
	Part I (written)		
1-22	Vocabulary	.55	.39
23-30	Mechanical principles	.68	.23
31-45	Word order	.84	.34
46-57	Interpretation of data	.81	.36
58-66	Arithmetic reasoning	.76	.33
67-69	Aviation information	.40	.27
70-78	Reading comprehension	.61	.52
79-90	Background for current affairs	.63	.23
91-114	Part II (oral)	.57	.44

Table 4

Part-Score Distribution Statistics
From the Operational Administration
Of MDAP English Proficiency Examination Form A

<u>Country</u>	<u>N</u>	<u>Part I (written)</u>		<u>Part II (oral)</u>	
		<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Belgium	68	64.1	10.9	16.0	3.9
Denmark	32	79.6	6.8	19.7	2.4
Holland	18	66.4	10.7	16.8	3.8
France	320	61.2	10.8	13.4	4.5
Portugal	12	64.9	3.5	10.8	3.8
Turkey	44	66.2	4.8	13.6	1.9
Total	494	63.4	11.2	14.2	4.5

have a wider range of talent than some of the other countries. They also represent the largest group tested in any of the countries.

The correlation between the two part-scores of the MDAP EPE Form A was computed for 523 cases collected under operational conditions overseas. The written vs. oral correlations was .56 as compared with that of .51 obtained from the experimental form on the French students of Class 53-F.

The United States Administration

From February 1953 through December 1954, MDAP EPE Form A was administered experimentally to all foreign students in Pre-Flight training at Lackland Air Force Base. This testing was done by highly trained test administrators, and the 16 mm. sound track was used to present the oral part of the test.

Item Analysis

An item analysis was made on a sample of 503 cases tested at Lackland Air Force Base in Pre-Flight Classes 53-I through 53-O. Some of the later cases had already been selected to attend pilot training school on the basis of their operational test scores. Part I (written) and Part II (oral) were analyzed separately. The answer sheets of the upper and lower 27% groups were selected on the basis of the appropriate part-score, and phi coefficients and difficulty levels were computed for each test item. A summary of this analysis is presented in Table 5.

Comparison of the mean difficulty level of each group of items in Tables 3 and 5 shows that the Lackland administration reflects selection on the basis of the overseas administration. The mean phi coefficients of the item types also show some variation from overseas testing to the Lackland testing. Most of this variation is associated with changes in item difficulties. On the whole, the analysis shows the groups of items to be acceptable from the point of view of internal consistency.

Reliability Estimate

Retest correlations and distribution statistics were obtained for 120 men for whom both operational and retest scores were available. These figures are presented in Table 6.

The correlation coefficients of .75 and .73 probably underestimate the true reliabilities of the part-scores. More than one month elapsed between the testing dates. The overseas administration occurred at the end of a course in English while the Lackland administration followed an ocean voyage and introduction to the United States Air Force.

Table 5

Summary of Item Analysis of MDAP English Proficiency Examination
Form A, Lackland Administration

(Samples: 503 foreign Pre-Flight students)

<u>Item</u>	<u>Item type</u>	<u>Mean difficulty</u>	<u>Mean phi</u>
Part I (written)			
1-22	Vocabulary	.68	.35
23-30	Mechanical principles	.74	.19
31-45	Word order	.90	.23
46-57	Interpretation of data	.92	.19
58-66	Arithmetic reasoning	.85	.27
67-69	Aviation information	.53	.37
70-78	Reading comprehension	.74	.38
79-90	Background for current affairs	.69	.28
90-114	Part II (oral)	.68	.50

Table 6

Test-Retest Distribution Statistics and Correlation Coefficients
For MDAP English Proficiency Examination Form A

(Sample: 120 foreign Pre-Flight students)

<u>Place of testing</u>	<u>Part I (written)</u>		<u>Part II (oral)</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Overseas	73.8	7.0	9.6	3.1
Lackland	68.9	8.6	8.7	4.1
	$r_{tt} = .75$		$r_{tt} = .73$	

The lower part-score means and larger standard deviations of data resulting from the Lackland administration are probably the result of natural statistical phenomena. It has been found that in any retest situation a given man's second score is apt to be nearer the average score of the group than was his first score. In Table 4 it is noted that the Part I

(written) mean of the total overseas group was 63.4. Men who made scores of 54 and below on Part I (written) were not shipped to America, accounting for the mean score of 73.8 for the overseas testing of men in Pre-Flight. However, when these men were retested their mean scores moved closer to the original group means.

A factor in the wider deviation of the Lackland testing is the error implicit in any measurement--a number of the men who had achieved a score of 55 overseas would make lower scores on their Lackland retesting, thereby increasing the range and standard deviation of the Lackland testing.

The results are complicated by unpredictable changes in motivation. While overseas, the men knew that failure on the test would preclude their coming to America. When tested here they had no information as to the use of the test results.

Comparison of English Proficiency Scores From USAF Basic Airmen and MDAP Personnel

The MDAP EPE Form A was administered to 307 basic airmen. The test was administered to the airmen exactly as it had been to foreign Pre-Flight students. The graph in Figure 1 presents distributions of the Part I (written) scores for basic airmen, for MDAP students who took the test at Lackland Air Force Base, and for foreign personnel who were tested overseas.

All three distributions in Figure 1 demonstrate a statistical character called "negative skewness." This term is applied to distributions in which the more able examinees tend to pile up on high scores while the scores of less able persons are more widely spread over the lower part of the score range. This is a desirable distribution because it permits of finer and more accurate discriminations among the less able examinees. In general, the curve representing the scores of basic airmen fits rather closely the curve representing the groups of MDAP Pre-Flight students. The curve representing the scores obtained under operational conditions is more nearly "normal" than either of the other two curves. The shift to the right from the operational administration distribution to that of the Lackland administration illustrates the selection that takes place when the operational test scores are used to eliminate foreign students from participation in the training program.

The graph in Figure 2 compares the Part II (oral) score distributions of the three samples described above. The Part II (oral) score distributions also show negative skewness. However, on the oral part of the test the basic airmen were definitely superior to either of the foreign student samples. The result of using the scores as a selection device again is evident in the shift to the right from the overseas administration to the Lackland administration.

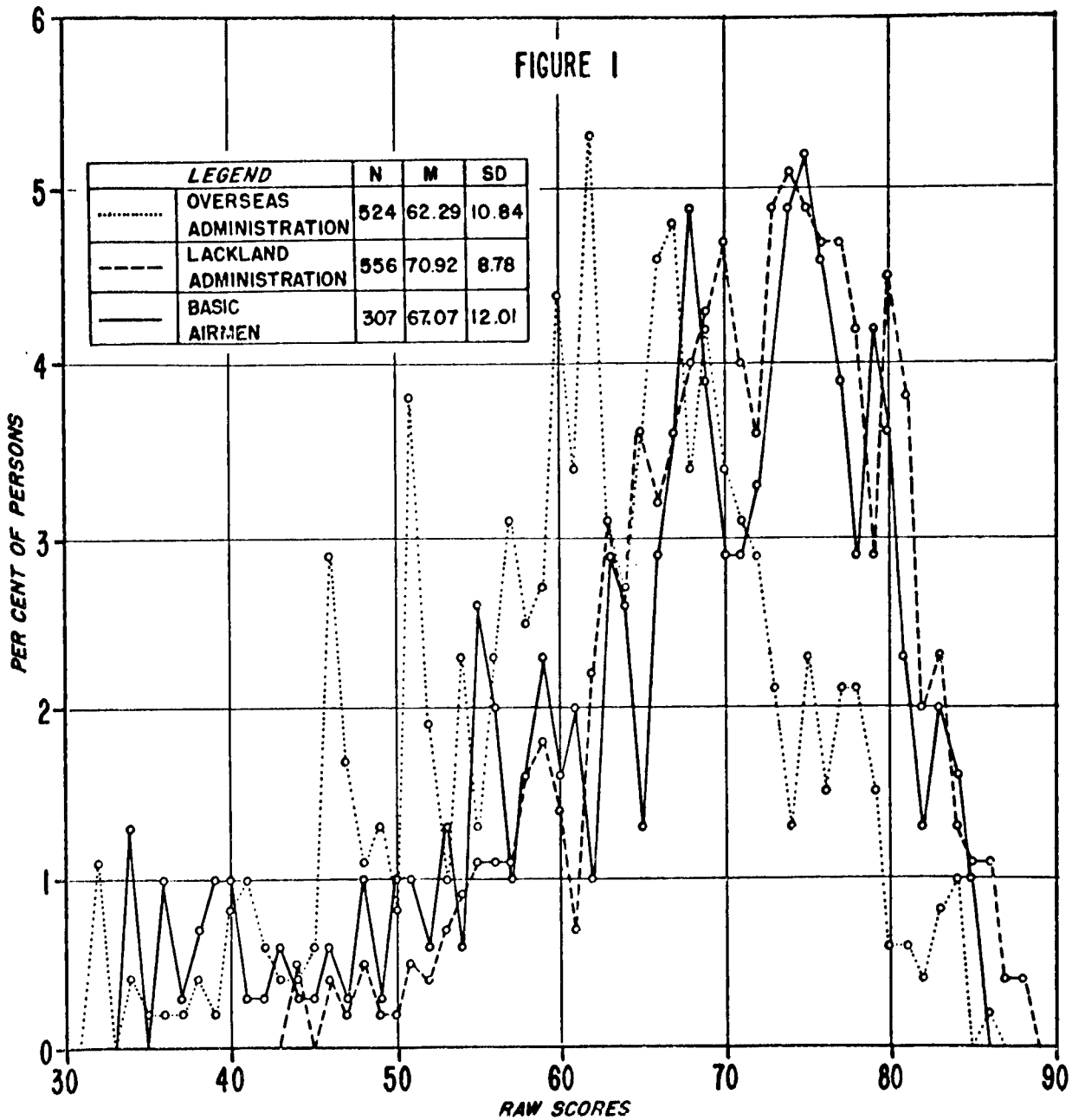


Fig. 1. Distributions of Part I (written) scores, English Proficiency Examination, for basic airmen, MDAP students tested at Lackland Air Force Base, and foreign personnel tested overseas.

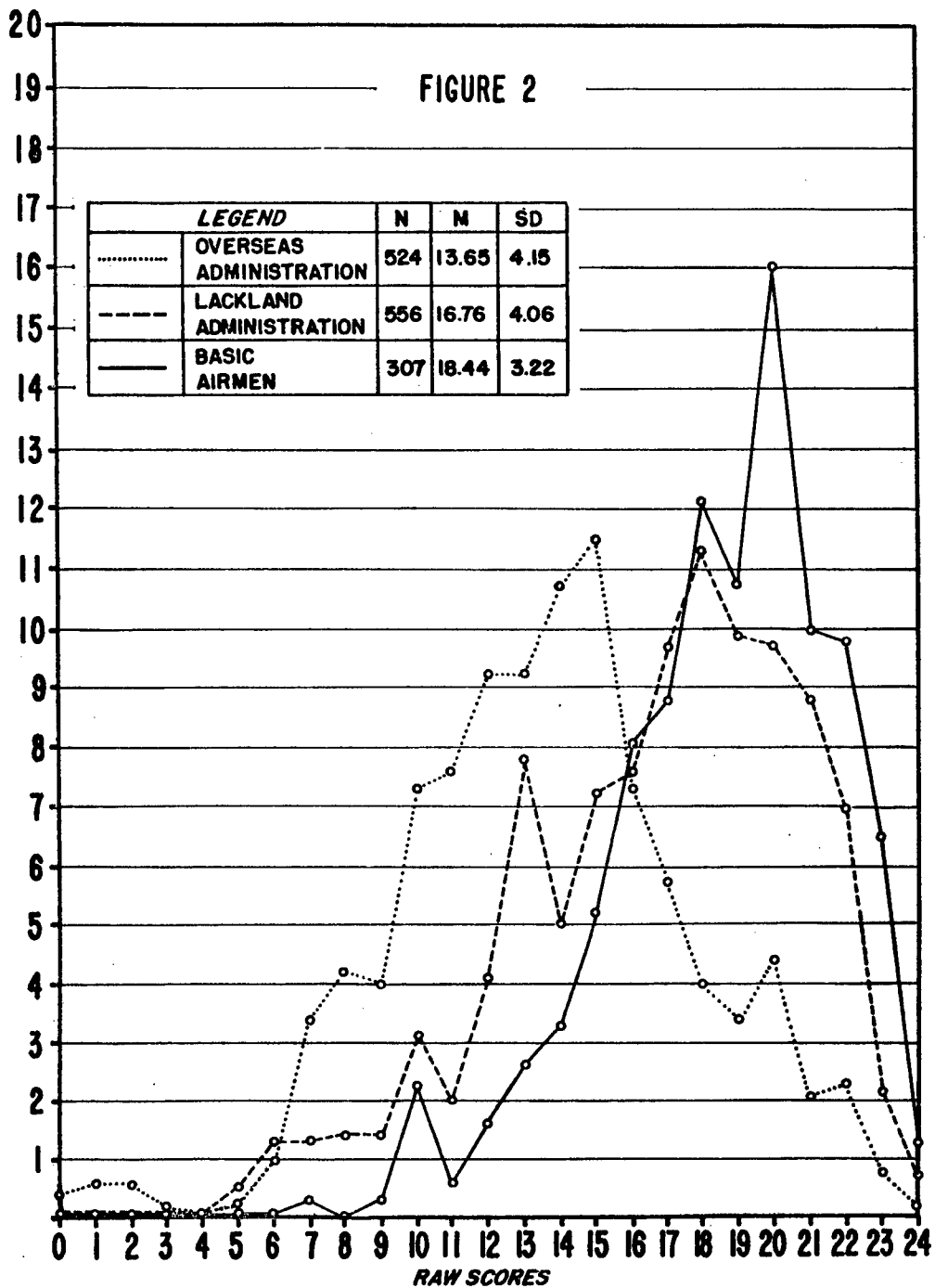


Fig. 2. Distributions of Part II (oral) scores, English Proficiency Examination, for basic airmen, MDAP students tested at Lackland Air Force Base, and foreign personnel tested overseas.

ESTIMATES OF VALIDITY

The preceding sections have described the techniques used in constructing MDAP EPE Form A. Various work has been reported which evaluates the test in terms of reliability estimates and distribution statistics for several samples. The present portion of the report is concerned with validity.

Validity of two types has been investigated. The first is construct validity (2). The concept involves the rationale that a test designed to measure a given dimension should correlate sufficiently with other measures of the same dimension. Thus by constructing two or more instruments to measure a specific dimension (in this case English comprehension) and correlating measurements taken with the instruments, validity can be inferred.

The second type of validity under investigation is predictive validity (2). The concept is one of practical use. If test scores can be found predictive of a practical criterion (such as school grades), then the test scores can be used to select those potential students who are most likely to succeed once they have entered the school.

The Criteria of English Comprehension

Two measures of English comprehension, both developed independently of MDAP EPE Form A, have been used to gather evidence of construct validity. The first is a behavior test developed in Personnel Research Laboratory, Air Force Personnel and Training Research Center. The second is a test developed by instructors in Pre-Flight.

Behavior Test

As an estimate of English comprehension independent of the MDAP EPE Form A, a behavior test of 28 items was developed. The test items are directions presented orally to the examinee in an interview type of situation. The directions, in simple English, are short sentences that tell the examinee to do something. If the examinee responds appropriately, he receives credit for the item. The score is the number of correct responses made.

The equipment used in the behavior test includes a pencil, a tablet, a book, a box containing ten 3" x 5" cards (one red and nine white), a line drawing of an aircraft, and a box, illustrated in Figure 3. A table and two chairs furnish the interview room. The 28 behavior test items are presented below:

- "1. Show me something you have in your pockets. Anything will do.
2. Put your left hand on the table.
3. Walk around your chair and sit down.

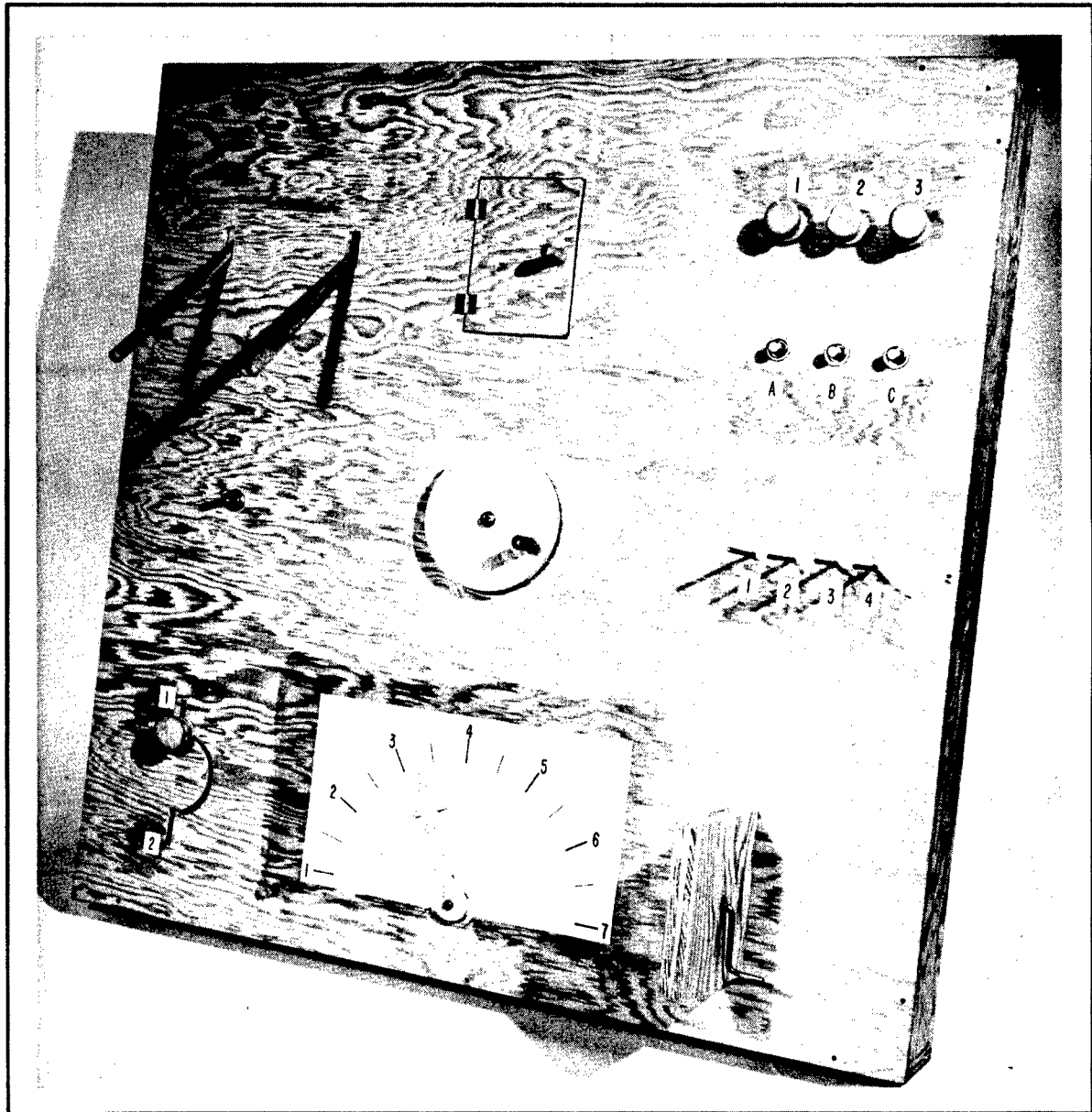


Fig. 3. Apparatus for the last ten items of the behavior test.

4. Pull the book toward you.
5. On the picture of the plane show me the rudder.
6. The propeller.
7. The elevator.
8. The wheels.
9. The motor.
10. The ailerons.
11. Using your hand, show me how a plane banks, dives, and lands.
12. Stand and lift your chair. (Now put it down.)
13. Write your name on the tablet.
14. Tear off that page you wrote on and put it under the book.
15. Open the box. Find the red card and drop it on the table.
16. Push the book to me.
17. Place the pencil in the box.
18. Slide your feet back and forth on the floor.
19. Open the door.
20. Pull the left lever.
21. Push the center button.
22. Turn the crank.
23. Set the dial needle on two.
24. Throw the switch.
25. Turn the wheel three turns.
26. Pull both end knobs up.
27. Put a ring on the 2d and 4th pins.
28. Slide the button to position two.⁰⁰

In administering the behavior test the examinees were told collectively that they were to be interviewed to determine their comprehension of English. The examinees were asked to follow directions given by the interviewer. All testing was done at the beginning of Pre-Flight training.

Arrangements were made to administer four behavior tests at a time. When possible, each administrator tested in a separate room. For some of the cases, two administrators operated in a large room, one at each end. Observation showed no effect on examinees as a result of two being tested in the same room at the same time. The examinees had to listen closely to the administrator to understand what they were being instructed to do. The instructions were given only once. No instruction was repeated. Assignment of examinees to administrators was random. The examinees were seated in a single room. The administrators drew the examinees from the front row and worked back to those on the back row. As each examinee was tested he was sent out of the building and to other duty. This removed the possibility of conference between men who had been tested and those who had not.

Seven persons acted as behavior test administrators. The examinees were tested the same day with the behavior test and Form A of the MDAP English Proficiency Examination. Foreign Pre-Flight students from eight countries comprised the 107-case sample. The countries and their respective

numbers of cases are presented in Table 7 along with mean scores and standard deviations on the behavior test.

Table 7

Distribution Statistics for the Behavior
Test Sample by Country

<u>Country</u>	<u>N</u>	<u>M</u>	<u>SD</u>
Belgium	19	21.2	4.3
Cuba	6	18.2	2.3
Holland	23	26.1	1.6
Italy	20	21.8	2.6
Norway	23	24.4	1.9
Iran	10	20.4	3.4
Columbia	5	19.2	4.7
Nicaragua	1	28.0	0.0
Total	107	22.8	3.8

The reliability of the behavior test has been estimated with Kuder-Richardson Formula 21. This estimate produced a coefficient of .70. While moderate in size, this coefficient is acceptable for a test of 28 items which are collected in interview fashion. The behavior test is deemed to have sufficient reliability for use as a criterion.

As a further evaluation of the behavior test procedure, an analysis of administrator differences has been made. The examinees were assigned to administrators in a random fashion. While there were seven administrators in all, two of these seven administered 27 tests each. The other five administrators each gathered fewer cases. To test the hypothesis that no differences existed between these two administrators, an analysis of covariance was computed as follows. The cases collected by administrators A and B (27 cases each) were considered as samples A and B. The means of the two samples were equated on the basis of both written and oral scores from the MDAP EPE. Then the adjusted means of the behavior test scores in the two samples were compared, using the F ratio. This procedure is described by Johnson (1). No significant difference was found between the adjusted behavior test means for the two samples. This result supports the hypothesis that no differences existed between the procedures used by the two administrators. Thus, it made no difference in an examinee's score which of these administrators gave him the behavior test. A summary of the covariance analysis is presented in Table 8.

Table 8
Summary of Covariance Analysis on Mean Behavior Test Scores
Of Groups Tested by Different Administrators

(Sample: 54 MDAP Pre-Flight students)

Variation	df	SS _y	SS _{x₁}	SS _{x₂}	SP _{yx₁}	SP _{yx₂}	SP _{x₁x₂}	Adjusted			
								df	SS _y	MS _y	F
Within	52	826.82	6657.33	890.67	1131.33	490.89	1474.67	50	532.19	10.64	
Between	1	4.17	8.17	10.67	-5.83	-6.67	9.33	1	13.13	13.13	1.23 ^a
Total	53	830.98	6665.50	901.33	1125.50	484.22	1484.00	51	545.32		

Note.--y: Behavior test scores.

x₁: Scores on Part I (written) of MDAP English Proficiency Examination.

x₂: Scores on Part II (oral) of MDAP English Proficiency Examination.

^a_p > .05

Having developed an independent measure of English comprehension ability with satisfactory reliability and collected a 107-case sample in which both behavior test scores and EPE scores were available, the relationship between the measures was determined. Correlation coefficients were computed between written and oral scores from the MDAP EPE Form A and the behavior test scores. The results are presented in Table 9.

The correlation coefficients between the MDAP EPE part-scores and the scores from the behavior test are considered as estimates of how well the proficiency examination measures ability to comprehend English. In this light, they are "validity coefficients." Thus, the coefficients of .60 for the written score and .65 for the oral score show the MDAP EPE Form A to have construct validity as a measure of English comprehension ability.

The coefficients of .60 and .65 are more impressive when considered in the light of the estimated reliability of the two instruments. If the tests are, or could be made, more reliable, the proficiency test scores might produce higher coefficients when correlated with behavior test scores. The coefficients estimating the relationships between proficiency test scores and behavior test scores are limited in size by the imperfect reliability of both instruments.

English Language--Proficiency Examination No. I

The MDAP EPE Form A was compared with another test independently developed by instructors in the Pre-Flight school, English Language--Proficiency Examination No. I. The Pre-Flight test was composed of 20 items

Table 9

Intercorrelations and Distribution Statistics
For MDAP English Proficiency Examination
Part-Scores and Behavior Test Scores

(Sample: 107 MDAP Pre-Flight students)

<u>Variable</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>M</u>	<u>SD</u>
1. Part I (written)		.68	.60	72.0	10.8
2. Part II (oral)	.68		.65	17.6	4.1
3. Behavior test	.60	.65		22.8	3.8

presented orally and 130 items presented in written form. The written items were comparable in subject matter to those in MDAP EPE Form A. The Pre-Flight test was used to place MDAP students in three classes of ability for English training. Scores on MDAP EPE Form A and the Pre-Flight placement test were collected for 203 foreign students in Pre-Flight training. These scores were then correlated as shown in Table 10.

Table 10

Distribution Statistics and Correlation Coefficients
For the MDAP English Proficiency Examination Form A
And a Pre-Flight Placement Test

(Sample: 203 foreign student pilots)

<u>Test</u>	<u>Mean</u>	<u>SD</u>	<u>r</u>	
			<u>1</u>	<u>2</u>
1. Part I (written)	73.08	9.10		
2. Part II (oral)	17.72	3.65		
3. Pre-Flight placement test	73.37	15.12	.78	.66

The fact that the two measuring instruments were developed independently to measure the same function makes the correlations important. The coefficients of .78 and .66 show the two tests to be much alike. This result lends confidence in the use of both tests.

Predictive Validity

Having determined that the MDAP EPE scores are reasonably valid estimates of ability to comprehend English, attention turned to the relationships between these scores and success in Pre-Flight school. The results from Pre-Flight school were chosen for investigation because they are the first school success information to become available. Upon completion of the training period the status of each student was determined by one of three categories: graduate, eliminate, or holdover.

A 376-case sample of MDAP Pre-Flight students was collected. Each student had been tested overseas with the MDAP EPE Form A and selected, on the basis of English proficiency, to participate in USAF flying training. Further, each case had been classified at the conclusion of Pre-Flight training as graduate, eliminate, or holdover. Table 11 shows the number of cases by country in each of the three classifications. The mean scores for the several groupings are presented in Table 12.

Table 11

Number of Cases in the Disposition
Categories by Country

<u>Country</u>	<u>Graduate</u>	<u>Eliminee</u>	<u>Holdover</u>	<u>Total</u>
France	219	3	32	254
Holland	16	0	0	16
Belgium	38	1	1	40
Denmark	21	0	0	21
Turkey	19	4	15	38
Portugal	7	0	0	7
Total	320	8	48	376
Other than French	103	5	16	122

To determine the relationships between test scores from the MDAP EPE and success in Pre-Flight training, biserial correlation coefficients were computed using graduation as the upper category of the dichotomy, and elimination plus holdover as the lower category. The coefficients and distribution statistics are presented in Table 13.

The correlation coefficients in Table 13 show that a definite positive relationship exists between both written and oral part-scores of the

Table 12

Mean Written and Oral Scores for Two Groups
Of MDAP Pre-Flight Students
By Disposition Groups

Group	Graduate			Eliminee			Holdover		
	N	Mw	Mo	N	Mw	Mo	N	Mw	Mo
French	219	62.8	13.3	3	54.0	12.0	32	52.5	11.6
Other than French	101	70.1	16.7	5	64.0	11.0	16	64.1	11.8

MDAP EPE and success in Pre-Flight training among foreign students. It should be noted that while the graduates form by far the largest percentage of the total group in the sample, all correlation coefficients are significant beyond the .01 level of confidence.

Table 13

Distribution Statistics and Correlation Coefficients Between
Written and Oral English Proficiency Scores and Pass/Fail
And Holdover in Pre-Flight Training Among MDAP Students

	<u>Nt</u>	<u>Mu</u>	<u>Ml</u>	<u>Mt</u>	<u>SD t</u>	<u>P</u>	<u>r_{bis}</u>	<u>SE_{r_{bis}}</u>
Part I (written)								
French	254	62.89	52.51	61.46	9.14	.862	.61	.075
Other	122	70.11	64.14	69.08	7.43	.828	.45	.116
Total	376	65.20	56.88	63.96	9.36	.851	.49	.067
Part II (oral)								
French	254	13.36	11.59	13.11	3.45	.862	.28	.093
Other	122	16.74	11.64	15.86	4.35	.828	.66	.095
Total	376	14.42	11.61	14.00	3.98	.851	.39	.071

The present data show the written-part scores to be more highly related to school success among French students than among foreign students from countries other than France. On the other hand, the oral scores

are more highly related to school success among foreign students from countries other than France than among French students. Information is not available in the present data to explain why these differences occurred.

SUMMARY AND CONCLUSIONS

This report has recounted the procedures used in development of MDAP English Proficiency Examination (EPE) Form A. Reliability of the test has been estimated. Distribution statistics for several samples of foreign military personnel, as well as one of basic airmen, have been reported. Validity has been considered from two points of view, construct validity and predictive validity. The major conclusions drawn from the above information are these:

1. The MDAP (EPE) Form A is a measure of ability among foreign student pilots to comprehend both written and oral English.
2. The test has predictive validity for Pre-Flight training and is acceptably reliable for use as a selective instrument.
3. Both written and oral parts have suitable reliability and validity and are only moderately related to each other. They can effectively be used separately as selection devices; a person qualifies on both parts of the test to become eligible for entry into pilot training.

Future research involving the MDAP EPE Form A should include investigation of differences between item types as to their effectiveness in measuring English comprehension. Test scores should also be related to success in Primary and Basic Pilot Training.

REFERENCES

1. JOHNSON, P.O. Statistical methods in research. New York: Prentice-Hall, 1950.
2. Joint Committee of American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education. Technical recommendations for psychological tests and diagnostic techniques. Washington, D.C.: Amer. Psychol. Ass., 1954.
3. KUDER, G.F., & RICHARDSON, M.W. The theory of estimation of test reliability. Psychometrika, 1937, 2, 151-160.