RESTRICTED

Not for publication

"ABILITY-TO-TAKE-IT" TESTS: EXAMINER DIFFERENCES AND VALIDATION

by

Henry S. Odbert
Leon Festinger
Seymour Wapner

A report on research conducted, in cooperation with the Bureau of Aeronautics of the U. S. Navy, at the University of Rochester, Rochester, New York, and at the Massachusetts General Hospital, Boston, Massachusetts, on cadets at the Naval Flight Preparatory School, Williams College, Williamstown, Massachusetts, by means of a grant-in-aid from the National Research Council Committee on Selection and Training of Aircraft Pilots from funds provided by the Civil Aeronautics Administration.

October 1945

CIVIL AERONAUTICS ADMINISTRATION
Division of Research
Report No. 49
Washington, D. C.

National Research Council

Committee on Selection and Training of Aircraft Pilots

Executive Subcommittee

M. S. Viteles, Chairman

E. C. Andrus	J. C. Flanagan
C. W. Bray	W. E. Kellum
D. R. Brimhall	P. J. Rulon
J. W. Dunlap	W. R. Miles

National Research Council

1945

LETTER OF TRANSMITTAL

NATIONAL RESEARCH COUNCIL

2101 Constitution Avenue, Washington, D. C.
Division of Anthropology and Psychology
Committee on Selection and Training of Aircraft Pilots

October 17, 1945

Dr. Dean R. Brimhall
Asst. to Administrator for Research
Civil Aeronautics Administration
Room 5335, Commerce Building
Washington 25, D. C.

Dear Dr. Brimhall:

Attached is a report entitled "Ability-to-Take-It" Tests: Examiner Differences in Validation by Henry S. Odbert, Leon Festinger, and Seymour Wapner. This report has been approved by the Committee on Selection and Training of Aircraft Pilots for transmittal to the Civil Aeronautics Administration with the recommendation that it be published in the CAA Division of Research Technical Series.

The investigation described in this report grows out of the concern with the performance of pilots under stress and the effort to identify those whose flight performance would be handicapped by reason of failure to adapt to stress situations. It is interesting to note that the findings are generally negative with respect to the predictive value of so-called "Ability-to-Take-It" Tests in the ordinary training situation. There still remains unanswered the question of whether such tests may have significance in relation to combat situations. It is hoped that an opportunity may still be furnished for the follow-up of subjects who became involved in combat situations following the completion of flight training.

The investigation represents another in the series involving close cooperation between the Committee on Selection and Training of Aircraft Pilots and the military services.

Cordially yours,

Morris S. Viteles, Chairman
Committee on Selection and
Training of Aircraft Pilots
National Research Council

MSV:rm

EDITORIAL FOREWORD

Research conducted in 1941 and 1942 by Dr. J. W. Dunlap and his staff at the University of Rochester on tests of the "ability-to-take-it" indicated that two instruments showed sufficient reliability for use in further research (see CAA Division of Research Report No. 11, February 1943). Steps were taken for the validation of these tests in connection with an experiment conducted at Williams College with the cooperation of the U. S. Navy, under the direction of Dr. Stanley Cobb and Dr. Jacob E. Finesinger, Massachusetts General Hospital.

An analysis of the results of this study revealed marked differences between examiners in the scores obtained in administering the tests. Such findings, as well as the data on validation obtained through comparison of test results with various criteria are embodied in the present report.

Acknowledgment is made to the Project Directors, Dr. Stanley Cobb and Dr. Jacob E. Finesinger, for undertaking the administration of the tests and to the U. S. Navy for furnishing the subjects for this investigation.

CONTENTS

	Page
EDITORIAL FOREWORD	v
SUMMARY	ix
INTRODUCTION	1
APPARATUS AND PROCEDURES	1
The Hand Dynamometer	1
Procedure in Administering the Hand Dynamometer Test	2
Measures Analyzed in the Hand Dynamometer Test	2
The Constant Stimulus Shocker	3
Procedure in Administering the Constant Stimulus Shocker	3
Measures Analyzed on Constant Stimulus Shocker	3
Sequence used in Administering Tests	3
Subjects	4
Examiners	4
EXAMINER DIFFERENCES	5
Analysis of Variance	5
Application of t-tests	5
Cumulative Tables	5
Further Examination of Hand Dynamometer Test	5
Intercorrelations Among Measures on the Hand Dynamometer Test	11
Intercorrelations Among Measures on the Constant Stimulus Shocker	11
Correlations Between the Two Tests	14
DISCUSSION OF EXAMINER DIFFERENCES	14
Factors Applicable to Both Tests	14
Factors Specific to the Hand Dynamometer Test	17
Factors Specific to the Constant Stimulus Shocker Test	19
VALIDATION	19
Nature of Criterion Data	19
Correlations with Success in Flight Training Through WTS	21
Correlations with Success in Flight Training to June, 1945	23

Page

Special Analysis of Flight Failures, using all Available Data 26
Discussion 26

SUMMARY 30

APPENDIX A: INSTRUCTIONS FOR ADMINISTRATION OF THE HAND DYNAMOMETER TEST 31

APPENDIX B: INSTRUCTIONS FOR ADMINISTRATION OF CONSTANT STIMULUS SHOCKER 35

SUMMARY

This report is concerned with differences in test results obtained by five examiners in administering two tests designed to measure resistance to pain and fatigue. The tests, tentatively designated as measures of "ability-to-take-it," had been developed by J. W. Dunlap and others at the University of Rochester. They were included in a battery of tests administered to cadets at the Naval Flight Preparatory School at Williams College in 1943.

The appearance of apparent differences in the mean scores obtained by the five examiners made it necessary to analyze these differences in some detail. The results of the analysis are presented because they emphasize some of the difficulties in standardization of this type of test.

One test employed a Smedley hand dynamometer which was modified to give, in addition to the usual measure of strength of grip, a signal indicating when strength of grip fell below an arbitrary setting. The subject was required to exert his maximum grip twice with each hand. The instructions then required the examiner to set the signaling apparatus at 60% of the stronger of the two grips for a given hand. The length of time was recorded during which the subject was able to maintain his grip above this setting. This time measure and the measure of strength of grip have both been subjected to analysis.

The second apparatus employed was a constant stimulus shocker. The current the subject received was regulated by the examiner. He recorded dial readings for four stages of experience which the subject was required to report: (1) when he first felt anything at all (threshold); (2) when he first felt pain (painful); (3) when the shock was "very painful" but he could stand more (very painful); and (4) the terminal point when he called for the current to be turned off (terminal). These four measures have been analyzed for each hand.

Analyses of variance were performed for 12 measures to determine whether there were significant differences among the mean values obtained between five examiners. For 11 measures the differences were such as would be expected by chance less than one time in 100. Differences as large as those obtained on the twelfth measure would be expected by chance only five times in 100. Application of t-tests revealed no significant differences between Examiners A and B who had been trained in the office where the tests were developed. These examiners were, however, in general, significantly different from the other examiners. The other examiners also showed a large proportion of significant differences among themselves. Differences were also observed among examiners in the relationships between first and second measures of strength of grip. Intercorrelations among the measures on each of the tests showed no appreciable examiner differences.

The observed differences arose even though the various examiners were all well aware of the need for standardized conditions. A number of circumstances are reviewed which may account for the striking differences which

were observed, including variations in examining procedures, changes in the sequence of tests, apparatus fluctuations, and changes due to discussion among the cadets, etc. Standardization of testing conditions and the preliminary training of examiners are probably especially necessary in tests of this sort which depend in a great measure on maintaining a uniform attitude in the subjects.

On the basis of the differences among the five examiners, it was inadvisable to combine the data for the different examiners in correlating the results with success in flight training. Validation was done separately for Examiners D and E by correlating measures on the tests with success in flight training through WTS, and for all data available to June, 1945. Further, a special analysis was done of flight failers since the test findings were not considered likely to predict all causes of elimination from flight training.

The biserial coefficients of correlation for the hand dynamometer measures (including all eliminations to June, 1945) were low and in a positive direction. The biserial coefficients for the constant stimulus shocker were uniformly low and in some instances, negative. Analyses of variance comparing means of passers, means of flight failers, and means of "other failers" for the two tests showed few significant P-values and revealed that the flight failers were not uniformly lower than other failers or passers.

The evidence based on the magnitude and direction of the coefficients for both examiners for the summed grip measure and the summed time measure, as well as the evidence from the analyses of variance, may suggest that the hand dynamometer test is tapping some aspect of behavior related to success in flight training, although the present data do not indicate that the measures in their present form would increase the effectiveness of established test batteries.

"ABILITY-TO-TAKE-IT" TESTS: EXAMINER DIFFERENCES AND VALIDATION

INTRODUCTION

An important consideration in introducing tests into a large-scale testing program is the assurance that different examiners will obtain uniform results. Even when special efforts are made to standardize the instructions and methods of administration, a field trial frequently reveals unanticipated differences among examiners. The present study, which was undertaken to validate two tests as predictors of success in flight training, revealed such large examiner differences that satisfactory validation was not possible. This report therefore places major emphasis on the analysis of differences in scores obtained by five examiners in administering two tests of resistance to pain and fatigue. The report concludes with validation data for the cases tested by two examiners.

The two tests, tentatively designated as measures of "ability-to-take-it," were developed by J. W. Dunlap and others at the University of Rochester. A study of the reliability of several such tests[1] led to the recommendation that two of them be validated.[2] The data analyzed in the present report were derived from a validation study undertaken by the Committee on Selection and Training of Aircraft Pilots, with the cooperation of the United States Navy and of investigators from the Massachusetts General Hospital.[3]

APPARATUS AND PROCEDURES[4]

The Hand Dynamometer. A Smedley hand dynamometer, which yields a measure of strength of grip, has been modified to give, in addition, a signal indicating when strength of grip falls below an arbitrary setting. This arbitrary setting can be adjusted for each individual.

---

[1] Dunlap, J. W., et al. Tests of the "Ability to Take It." Washington, D. C.: CAA Division of Research, Report No. 11. February 1943.

[2] A preliminary validation study of the hand dynamometer test and the constant stimulus shocker test was made early in 1942, when the two tests were administered to 195 Civilian Pilot Training students in the Standard Testing Program. Results of the investigation have been presented to the Committee on Selection and Training of Aircraft Pilots by M. J. Wantman in two progress reports: Preliminary report on the validity of the "ability to take it" tests, May, 1943, and Second report on the validity of the "ability-to-take-it" tests, July, 1943. (Copies in Committee files.)

[3] Cobb, S., Finesinger, J. E., & Chapple, E. C. Manual for the Cobb-Finesinger study on the classification of pilots. October, 1943. (Copy in Committee files.)

[4] The descriptions of apparatus and procedures are quoted in part from the report by J. W. Dunlap, et al., ibid., with some technical modifications suggested by J. D. Coakley, one of the co-authors of that study.

Two small flash-light cells are attached to the back of the dynamometer. These batteries supply the current for a buzzer, a red light, and an amber light attached to the dial face of the instrument. A small attachment on the face of the dial makes it possible to vary the point on the strength-of-grip scale at which the buzzer, the red light, and the amber light are activated. The relations of the buzzer and the two lights never vary among themselves, but the three, as a unit, can be shifted to any position on the dial. A complete circuit is formed through a rider which moves forward with the indicator hand. Thus, if the attachment is set at thirty units on the dial, the amber light will show when the indicator hand shows a grip of 35-37 kilograms, the red light will come on at 32-34 kilograms, and the buzzer will sound at thirty kilograms.

Procedure in Administering the Hand Dynamometer Test. Each subject was asked to exert his maximum grip, first with his preferred hand and then with the other. Two readings were taken for each hand. These readings were used as a base in determining a setting which the subject was required to maintain as long as possible. In this determination, two examiners (A & B) used the standard procedure specified in the instructions (see Appendix A). The other three examiners (C, D, & E) used a modified procedure.[5]

a. Standard Procedure. The examiner set the buzzer to sound at 60% of the stronger of the two grips the subject exerted with the hand being tested. The subject's task was to keep the indicator hand of the dynamometer above this point so that the amber light remained on. He was not to allow the instrument to touch his body at any time. The red light flashed on as a warning to the subject that he was relaxing his grip. The buzzer was the signal that the test was over for that hand. The examiner recorded the total time for each hand from the starting point until the sound of the buzzer. At least ten minutes intervened between hands.

b. Modified Procedure. The examiner set the buzzer to sound at 60% of the average of the two grips the subject exerted with the hand being tested. The procedure was otherwise the same.

Measures Analyzed in the Hand Dynamometer Test. The following measures obtained in the hand dynamometer test have been analyzed:

a. First measure of strength of grip, preferred hand (in kilograms)
b. Second measure of strength of grip, preferred hand (in kilograms)
c. Length of time grip was maintained, preferred hand (in seconds)
d. Length of time grip was maintained, non-preferred hand (in seconds)

Two measures of strength of grip were recorded for the non-preferred hand, but these have not been analyzed in the present study.

---

[5]The change in procedures occurred when the two examiners provided by the Committee were forced, by unforeseen circumstances, to leave the project before replacements could be thoroughly trained in the administration of the tests.

The Constant Stimulus Shocker. The constant stimulus shocker used in this study furnishes a peak current varying from 0.0 to 11.0 milliamperes when connected with a 110-volt source.[6] This current is secured from a 500-volt transformer and is rectified in a current-control tube. The output of the stimulator is connected to two electrodes, each of which is immersed in a separate beaker of saturated salt solution. The subject completes the circuit by inserting two fingers of the same hand in the two beakers. A scale on the front of the instrument indicates the amount of current used. A false rider is carried by the current-regulating knob of the instrument and indicates the maximum current used.

Procedure in Administering the Constant Stimulus Shocker. The examiner told the cadet the general purpose of the test and examined his index and middle fingers for cuts and hangnails. If defects were found on these fingers, others were selected. Collodion was used if necessary to protect minor cuts. After being blindfolded the cadet was given further instructions. He was asked to report four stages of experience as the examiner turned the knob at a slow constant speed:[7]

a. When he first felt anything at all (Threshold)
b. When he first felt pain (Painful)
c. When the shock was "very painful" but he could stand more (Very Painful)
d. The terminal point when he called for the current to be cut off (Terminal)

Measures Analyzed on Constant Stimulus Shocker. Milliampere readings were taken at the following points:

a. Threshold, preferred hand
b. Painful, preferred hand
c. Very painful, preferred hand
d. Terminal, preferred hand
e. Threshold, non-preferred hand
f. Painful, non-preferred hand
g. Very painful, non-preferred hand
h. Terminal, non-preferred hand

Sequence used in Administering Tests. The tests were administered in the following sequence:

---

[6]The shocker used in the earlier study furnished a maximum peak current of 9.6 milliamperes. It was otherwise similar to that described here. Specifications for its construction are found in the report by J. W. Dunlap et al., op. cit.

[7]The examiner was to have practiced turning the current-regulating knob at the rate of one complete revolution every 60 seconds. The instructions indicated that he should have practiced this with a stop watch until he had an error of not more than 4 seconds. (Instructions for administration of the test are presented in Appendix B.)

a. Strength of grip, preferred hand
b. Strength of grip, non-preferred hand
c. Strength of grip, preferred hand
d. Strength of grip, non-preferred hand
e. Time of sustained grip, preferred hand
f. Constant stimulus shocker, preferred hand

(10-minute interval)

g. Constant stimulus shocker, non-preferred hand
h. Time of sustained grip, non-preferred hand

Steps g and h were reversed for the first 114 subjects, each of these steps being given ten minutes after the corresponding test with the preferred hand.[8]

Subjects. The subjects were 592 cadets in the Naval Flight Preparatory School at Williams College. Forty-nine records were rejected for the reasons indicated below:

Small number of cases per examiner	20
Incomplete data	15
Sprained wrist or sore finger	10
No effort (hand dynamometer)	2
Directions misunderstood	2

Examiners. Eight examiners were used in the administration of the "Ability-to-Take-It" Tests at Williams College. Results for three examiners who tested a total of 20 cadets have not been treated in the present analysis. The five examiners whose results are here analyzed were all young men. Examiners A and B had had previous experience in the work on the "Ability-to-Take-It" Tests at the University of Rochester. These examiners tested the first 65 cadets for whom complete data are available. Examiners A and B left the testing program at the same time. Examiner C, a permanent member of the staff of the Massachusetts General Hospital, tested 29 subjects until Examiners D and E took over the administration of the "Ability-to-Take-It" Tests. Examiners D and E alternated in giving the tests until the end of the program. The 543 cases used in the final analysis were divided as follows among the five examiners:

Examiner A	44
Examiner B	19
Examiner C	28
Examiner D	221
Examiner E	231

---

[8] The change in the sequence of tests was recommended by the Executive Subcommittee of the Committee on Selection and Training of Aircraft Pilots to correct an oversight of the Technical Aide, who had proposed a sequence which departed from that used in the original investigation.

EXAMINER DIFFERENCES

Analysis of Variance. A casual inspection of the means of the various examiners revealed rather large differences. Accordingly analyses of variance were performed for each measure on each test to see whether these differences were significant. Tables 1 and 2 present the means for the examiners, the quantities obtained in the analysis of variance, and the P-values for the analysis of variance.

All the analyses of variance show F-values significant beyond the 1% level, except that for the first measure of grip with the preferred hand on the hand dynamometer test, which was significant at the 5% level. It may be concluded that there are real differences among examiners for each of the measures on each of the tests.

Application of t-tests. Tables 1 and 2 also present the t-values for the differences between pairs of examiners based on the error term from the analysis of variance. The t-tests reveal no significant differences between Examiners A and B. Examiners A and B are in general significantly different from the other examiners. The other examiners also show a large proportion of significant differences among themselves.

Cumulative Tables. Tables 3 and 4 are presented to indicate in a different way the magnitude of the differences (for the preferred hand) in the two scores originally recommended as measures of the "ability-to-take-it." Table 3 shows, for each examiner, the per cent of cadets scoring at and below given levels on the measure of time of grip on the hand dynamometer. It may be seen that all 23 subjects tested by Examiner C held their grip longer than 40 seconds, whereas approximately 60% of the subjects tested by Examiners A and B held their grips 40 seconds or less. The corresponding percentages for Examiners D and E are 23% and 36%. Similar comparisons may be made at other levels.

Table 4 gives similar information for the terminal score on the constant stimulus shocker. Practically all subjects tested by Examiners A and B scored 8.0 or lower. Approximately half the subjects tested by the other three examiners scored above this point.

Further Examination of Hand Dynamometer Test. On strength of grip the second measure could be expected to be lower than the first measure because of fatigue effects. Inspection of the mean values on the hand dynamometer test (Table 1), however, shows that the differences between the first and second measures of strength of grip are not uniformly in the same direction for all examiners. Table 5 presents t-tests for the differences between first and second measures. All the examiners, except C, show the expected decrease to a significant degree. Examiner C shows an increase from first to second testing which approaches significance. It seems plausible to suppose that Examiner C introduced some variation into the time sequence or motivation of the subject to have this effect.

TABLE 1

EXAMINER DIFFERENCES ON HAND DYNAMOMETER MEASURES***

	Examiner	Means	Variance Within Examiners	Variance Between Examiners	F-Value	P-Value		t-Values for Differences between Pairs of Examiners A	B	C	D	E
Grip, First Measure, Preferred Hand	A	53.09	50.17	121.93	2.43	.05	A	-	.12	.22	.44	1.85
	B	53.32					B		-	.29	.17	1.14
	C	52.71					C			-	.63	1.79
	D	53.61					D				-	2.45*
	E	55.25					E					-
Grip, Second Measure, Preferred Hand	A	52.11	46.57	173.64	3.73	<.01	A	-	.51	1.42	.19	2.05*
	B	51.16					B		-	1.63	.71	1.99*
	C	54.46					C			-	1.56	.04
	D	52.32					D				-	3.27**
	E	54.41					E					-
Time, Preferred Hand	A	37.34	257.43	4930.03	19.15	<.01	A	-	.16	7.76**	4.91**	3.55**
	B	36.63					B		-	6.46**	3.57**	2.63**
	C	67.43					C			-	5.30**	6.46**
	D	50.35					D				-	2.42*
	E	46.70					E					-
Time, Non-preferred Hand	A	33.64	276.58	8627.97	31.20	<.01	A	-	.88	9.01**	7.31**	5.30**
	B	29.63					B		-	8.14**	6.06**	4.67**
	C	69.86					C			-	4.83**	6.51**
	D	53.74					D				-	3.55**
	E	48.17					E					-

*P-value between .05 and .01
**P-value less than .01
***Number of cases for Examiner A is 44; B, 19; C, 28; D, 221; E, 231

TABLE 2

EXAMINER DIFFERENCES ON CONSTANT STIMULUS
SHOCKER MEASURES (PREFERRED HAND)***

	Examiner	Means	Variance Within Examiners	Variance Between Examiners	F-Value	P-Value		t-Values for Differences between Pairs of Examiners A	B	C	D	E
Threshold, Preferred Hand	A	.84	.08	22.32	295.61	<.01	A	-	.80	11.21**	17.09**	28.18**
	B	.58					B		-	9.76**	12.56**	20.12**
	C	1.38					C			-	.53	9.60**
	D	1.41					D				-	20.79**
	E	1.91					E					-
Painful, Preferred Hand	A	2.11	1.97	100.75	51.14	<.01	A	-	.73	4.24**	9.95**	12.40**
	B	2.39					B		-	2.77**	6.03**	7.71**
	C	3.55					C			-	3.09**	5.07**
	D	4.42					D				-	4.20**
	E	4.97					E					-
Very Painful, Preferred Hand	A	3.61	2.59	114.96	44.31	<.01	A	-	.01	5.96**	9.59**	11.45**
	B	3.61					B		-	4.83**	6.61**	7.89**
	C	5.93					C			-	.72	2.22*
	D	6.16					D				-	3.19**
	E	6.64					E					-
Terminal, Preferred Hand	A	5.15	3.16	134.24	42.46	<.01	A	-	.49	5.70**	10.76**	9.73**
	B	4.91					B		-	5.09**	8.00**	7.28**
	C	7.60					C			-	1.99*	1.12
	D	8.31					D				-	1.88
	E	8.00					E					-

*P-value between .05 and .01
**P-value less than .01
***Number of cases for Examiner A is 44; B, 19; C, 28; D, 221; E, 231

TABLE 2 (Continued)

EXAMINER DIFFERENCES ON CONSTANT STIMULUS SHOCKER MEASURES (NON-PREFERRED HAND)***

	Examiner	Means	Variance Within Examiners	Variance Between Examiners	F-Value	P-Value		t-Values for Differences between Pairs of Examiners A	B	C	D	E
Threshold, Non-preferred Hand	A	.67	.09	23.72	267.07	<.01	A	-	1.05	10.14**	16.06**	26.57**
	B	.58					B		-	9.17**	12.30**	19.55**
	C	1.40					C			-	.95	9.53**
	D	1.46					D				-	18.39**
	E	1.97					E					-
Painful, Non-preferred Hand	A	2.15	1.54	81.66	52.95	<.01	A	-	1.09	6.43**	9.49**	18.03**
	B	2.52					B		-	4.24**	5.30**	7.[illegible]**
	C	4.08					C			-	.24	2.92*
	D	4.09					D				-	6.[illegible]**
	E	4.80					E					-
Very Painful, Non-Preferred Hand	A	3.50	2.16	105.00	50.04	<.01	A	-	.02	7.05**	8.95**	12.07**
	B	3.51					B		-	6.37**	6.[illegible]**	9.30**
	C	6.29					C			-	2.09*	.45
	D	5.68					D				-	5.42**
	E	6.42					E					-
Terminal, Non-preferred Hand	A	4.81	2.90	138.74	47.80	<.01	A	-	.67	7.48**	10.90**	10.90**
	B	4.50					B		-	6.71**	8.30**	8.27**
	C	7.90					C			-	.06	.09
	D	7.58					D				-	.07
	E	7.87					E					-

*P-value between .05 and .01
**P-value less than .01
***Number of cases for Examiner A is 44; B, 19; C, 28; D, 221; E, 231

TABLE 3

PER CENT OF SUBJECTS TESTED BY EACH EXAMINER WHO FELL AT AND BELOW EACH LEVEL INDICATED ON THE HAND DYNAMOMETER TIME MEASURE (PREFERRED HAND)

Time	Examiner A	B	C	D	E
120 secs.			100.0%		
115			96 4		100.0%
110			96.4		99.6
105			96.4		99.6
100			96.4	100.0%	99.6
95			89.3	99.5	99.1
90			89.3	99.5	99.1
85			85.7	97.7	99.1
80			82.1	96.8	97.8
75			75 0	94.6	96.5
70		100.0%	71 4	92.3	94.8
65		94.7	57 1	86.9	91.8
60	100.0%	89.5	39 3	76.9	76.6
55	90.9	84.2	21 4	62.0	65.8
50	81.8	84.2	17 9	52.9	61.0
45	70.5	68.4	3 6	34.8	50.2
40	59.1	63.2		23.1	36.4
35	43.2	52.6		13.6	29.9
30	36.4	31.6		10.9	19.9
25	18.2	26 3		3.1	13.0
20	6.8	21.1		2.7	5.6
15	4.5	10 5		.5	2.2
10					1.3
5					
N	44	19	28	221	231

PER CENT OF SUBJECTS TESTED BY EACH EXAMINER WHO FELL AT AND BELOW EACH LEVEL INDICATED ON THE TERMINAL SCORE OF THE CONSTANT STIMULUS SHOCKER (TRANSFERRED HAND)

Score	Examiner A	B	C	D	E
11.0 ms.	100.0%		100.0%	100.0%	100.0%
10.0	97.7		89.3	79.2	89.6
9.0	97.7		78.6	68.8	81.0
8.0	97.7	100.0%	53.6	48.0	52.8
7.0	90.2	89.5	35.7	22.6	25.5
6.0	63.6	73.7	28.6	11.3	11.3
5.0	47.7	47.4	14.3	5.0	4.3
4.0	25.0	26.3	3.6	2.3	1.7
3.0	15.9	10.5		.9	.4
2.0	4.5	10.5		.5	
1.0	2.3	5.3			
0.0					
N	44	19	28	221	231

TABLE 5

DIFFERENCES BETWEEN FIRST AND SECOND MEASURES OF STRENGTH OF GRIP

Examiner	N	M_1-M_2	t	P
A	44	.98	1.960	.05
B	19	2.16	3.085	<.01
C	28	-1.75	2.011	.06
D	221	1.29	4.778	<.01
E	231	.84	3.000	<.01

Inspection of the means of the five measures on the hand dynamometer test shows both increases and decreases from the preferred to the non-preferred hand for the different examiners. Examiners A and B show decreases from preferred to non-preferred hand, which approach significance (Table 6). Examiners A and B are the ones who were previously observed to resemble each other and differ from the others. The other three examiners all show increases, one of which is clearly significant. It may be safely assumed that the procedure of Examiners A and B varied somewhat from the procedure used by Examiners C, D, and E.

TABLE 6

DIFFERENCES BETWEEN PREFERRED AND NON-PREFERRED HAND ON TIME OF SUSTAINED GRIP

Examiner	N	M_{p-n}	t	P
A	44	3.70	1.878	.06
B	19	7.00	2.229	.04
C	28	-2.43	.568	.58
D	221	-3.39	3.000	<.01
E	231	-1.47	1.324	.19

Intercorrelations Among Measures on the Hand Dynamometer Test. Table 7 presents intercorrelations among measures on the hand dynamometer test for each examiner separately. The correlations between first and second measures of strength of grip reveal no marked examiner differences, ranging from .81 to .91 for the five examiners. These figures might be taken as indicative of the degree of reliability of the measure when test and retest are separated by a ten-minute interval.

The correlations between time measures on the preferred and on the non-preferred hand range from .11 to .67.[9]

Correlations between the time measures and the measures of strength of grip hover about zero.

Intercorrelations Among Measures on the Constant Stimulus Shocker. Intercorrelations among measures on the constant stimulus shocker are shown in Table 8 for each examiner separately. It may be seen that all the correlations are positive, ranging from .15 to .93. Examination of the table reveals no very marked or consistent examiner differences in the size of the correlations. There is perhaps some indication that correlations between corresponding measures on the two hands are more uniformly high for Examiners A and B than for the other examiners.

---

[9]These figures may be compared with correlations of .51 and .49 between time measures for right and left hands, reported in the study by J. W. Dunlap et al., op. cit.

TABLE 7

INTERCORRELATIONS OF HAND DYNAMOMETER MEASURES*

	Examiner	Grip, First Measure, Preferred Hand	Grip, Second Measure, Preferred Hand	Time, Preferred Hand	Time, Non-preferred Hand
Grip, First Measure, Preferred Hand	A	-	.85	.10	.05
	B	-	.91	.06	-.05
	C	-	.89	-.12	.10
	D	-	.85	-.05	.04
	E	-	.81	.00	.03
Grip, Second Measure Preferred Hand	A	-	-	.02	.13
	B	-	-	.10	.04
	C	-	-	-.11	-.04
	D	-	-	.03	.05
	E	-	-	-.02	.07
Time, Preferred Hand	A	-	-	-	.39
	B	-	-	-	.67
	C	-	-	-	.11
	D	-	-	-	.46
	E	-	-	-	.51
Time, Non-Preferred Hand	A	-	-	-	-
	B	-	-	-	-
	C	-	-	-	-
	D	-	-	-	-
	E	-	-	-	-
Mean	A	53.09	52.11	37.34	33.64
	B	53.32	51.16	36.63	29.63
	C	52.71	54.46	67.43	69.86
	D	53.61	52.32	50.35	53.74
	E	55.25	54.41	46.70	48.17
σ	A	6.24	5.63	12.02	11.44
	B	6.99	5.81	15.96	16.75
	C	7.13	6.73	17.03	16.32
	D	7.25	7.31	15.27	16.93
	E	7.01	6.56	17.11	17.01

*Number of cases for Examiner A is 44; B, 19; C, 28; D, 221; E, 231

TABLE 8

INTERCORRELATIONS OF CONSTANT STIMULUS SHOCKER MEASURES*

			PREFERRED HAND				NON-PREFERRED HAND			
		Examiner	Threshold	Painful	Very Painful	Terminal	Threshold	Painful	Very Painful	Terminal
PREFERRED HAND	Threshold	A	-	.47	.46	.50	.73	.53	.50	.41
		B	-	.57	.57	.57	.81	.53	.51	.45
		C	-	.43	.44	.37	.76	.39	.38	.36
		D	-	.39	.33	.20	.59	.35	.28	.18
		E	-	.34	.39	.25	.56	.28	.22	.20
	Painful	A		-	.88	.71	.49	.83	.74	.56
		B		-	.92	.73	.50	.86	.78	.71
		C		-	.76	.67	.49	.69	.64	.57
		D		-	.86	.63	.25	.68	.65	.47
		E		-	.85	.64	.23	.65	.62	.53
	Very Painful	A			-	.83	.46	.80	.82	.73
		B			-	.83	.44	.87	.84	.78
		C			-	.86	.47	.69	.83	.79
		D			-	.76	.23	.68	.75	.59
		E			-	.80	.37	.72	.73	.69
	Terminal	A				-	.46	.72	.79	.83
		B				-	.25	.76	.80	.88
		C				-	.56	.67	.85	.92
		D				-	.15	.53	.65	.79
		E				-	.23	.61	.68	.80
NON-PREFERRED HAND	Threshold	A					-	.64	.68	.55
		B					-	.46	.36	.17
		C					-	.58	.51	.56
		D					-	.34	.27	.16
		E					-	.42	.35	.30
	Painful	A						-	.91	.75
		B						-	.93	.81
		C						-	.76	.70
		D						-	.85	.54
		E						-	.87	.68
	Very Painful	A							-	.90
		B							-	.91
		C							-	.90
		D							-	.68
		E							-	.81
	Terminal	A								-
		B								-
		C								-
		D								-
		E								-
	Mean	A	.64	2.11	3.61	5.15	.67	2.15	3.50	4.81
		B	.58	2.39	3.61	4.91	.58	2.52	3.51	4.50
		C	1.38	3.55	5.93	7.60	1.40	4.08	6.29	7.90
		D	1.41	4.42	6.16	8.31	1.46	4.09	5.68	7.38
		E	1.91	4.97	6.64	8.00	1.97	4.80	6.42	7.87
	σ	A	.15	1.15	1.54	1.80	.17	.93	1.33	1.60
		B	.12	1.28	1.43	1.76	.10	1.26	1.46	1.65
		C	.35	1.29	1.71	2.10	.41	1.54	1.96	2.26
		D	.28	1.55	1.82	1.89	.29	1.26	1.51	1.76
		F	.28	1.31	1.38	1.60	.32	1.22	1.37	1.57

*Number of cases for Examiner A is 44; B, 19; C, 28; D, 221; E, 231

A rough indication of trends in the data may be obtained by directing attention toward the mid-correlation in each group of five. Five of these correlations are .80 or higher:

a. Terminal (preferred) and Terminal (non-preferred)[10]
b. Terminal (preferred) and Very Painful (preferred)
c. Terminal (non-preferred) and Very Painful (non-preferred)
d. Very Painful (preferred) and Painful (preferred)
e. Very Painful (non-preferred) and Painful (non-preferred)

The tendency may be noted for the higher correlations to occur between corresponding measures on the two hands and between adjacent measures on the same hand. Correlations of threshold measures with other measures tend to be the lowest in the table.

Correlations Between the Two Tests. Correlations between the two tests are presented for each examiner separately in Table 9. They range from -.63 to .43. The general tendency is for the correlations to fluctuate about zero. There is no evidence of consistent examiner differences or of any consistent relationship between the two tests.

DISCUSSION OF EXAMINER DIFFERENCES

Examination of the data on the two "Ability-to-Take-It" Tests has revealed marked differences among the scores obtained by subjects tested by different examiners. Consideration of possible explanations of these differences will be divided into three sections: those applying to both tests, those specific to the hand dynamometer test, and those specific to the constant stimulus shocker.

Factors Applicable to Both Tests. The following factors may be considered applicable to both tests:

a. Discussion Among Subjects. While all subjects were told not to discuss the tests with other cadets, this instruction may not have been altogether successful. If such discussion among subjects occurred, it might have resulted in systematic changes over a period of time. Such changes might show up in differences among three groups of subjects: those tested by Examiners A and B, those tested by Examiner C, and those tested by Examiners D and E. There was practically no temporal overlap in the testing of these three groups (the only exception being that Examiner D tested his first four subjects while Examiner C was still testing). It will be noted that on most measures Examiners A and B obtained lower scores than Examiners C, D, and E. On the constant stimulus shocker the scores obtained by examiner C were all significantly higher than those ob-

---

[10] Correlations between terminal measures on preferred and non-preferred hands range from .79 to .92. In the study by J. W. Dunlap, correlations of .84 and .79 were reported between terminal measures on right and left hands.

TABLE 9

CORRELATIONS BETWEEN HAND DYNAMOMETER AND CONSTANT STIMULUS SHOCKER MEASURES*

			CONSTANT STIMULUS SHOCKER							
			PREFERRED HAND				NON-PREFERRED HAND			
		Examiner	Threshold	Painful	Very Painful	Terminal	Threshold	Painful	Very Painful	Terminal
SAND DYNAMOMETER	Grip, Time, Measure, Preferred Hand	A	.25	.21	.30	.36	.30	.36	.43	.39
		B	.12	.05	-.07	-.07	.03	-.22	-.25	-.23
		C	.06	.48	.32	.19	.02	.27	.20	.23
		D	.16	.17	.16	.19	.14	.25	.24	.22
		E	-.04	.09	.15	.19	.07	.06	.17	.19
	Grip, Second Measure, Preferred Hand	A	.26	.31	.32	.37	.27	.43	.44	.39
		B	.01	-.09	-.25	-.18	-.07	-.38	-.37	-.33
		C	.09	.38	.28	.16	.06	.21	.19	.21
		D	.18	.17	.17	.17	.14	.25	.27	.07
		E	-.02	.09	.11	15	.09	.03	.12	.18
	Time, Preferred Hand	A	.15	.09	.17	.18	.09	.22	.26	.25
		B	-.41	.05	-.04	-.24	-.63	.05	.05	.14
		C	-.18	-.05	-.18	-.19	.03	-.12	-.07	-.26
		D	.02	.04	.06	.12	-.03	.00	.00	.06
		E	-.17	.01	-.03	.04	-.10	.08	.05	-.04
	Time, Non-Preferred Hand	A	.13	-.04	.03	.09	.02	.12	.12	.11
		B	-.13	.03	.02	.42	-.52	-.11	-.05	.22
		C	-.04	.07	.17	.31	.18	.08	.17	.30
		D	10	.12	.12	17	-.08	.09	.11	.14
		E	05	.00	.00	.05	.03	.02	.07	.07

*Number of cases for Examiner A is 44; B, 19; C, 28; D, 221; E, 231

tained by Examiners A and B. The scores obtained by Examiners D and E were generally higher than those obtained by Examiner C, although not always significantly so. On the hand dynamometer test Examiners A and B obtained time scores significantly lower than those of the other examiners. Examiner C, however, obtained scores significantly higher than those of Examiners D and E.

b. Variations in Calibre of Cadets. It is possible that there are differences in the quality of cadets tested by different examiners. This possibility could be checked in part by referring to the original Navy selection data on the cadets.

Three kinds of selection data were available: (1) A Flight Aptitude Rating which consisted of a combination of scores on the Test of Mechanical Comprehension and the Biographical Inventory. (2) The Aviation Classification Test. This was a test of general intelligence which was found to predict ground school performance. (3) The Test of Mechanical Comprehension. This was a test of the ability to handle familiar physical concepts and has been found to predict ultimate success or failure in flight training.

The scores on these Navy selection items were used to evaluate whether the different examiners tested cadets of a different quality. This evaluation was made by the method of analysis of variance. Since there were no differences between Examiners A and B, and also between Examiners D and E on the "Ability-to-Take-It" tests, the cases of Examiner A were grouped with the cases of Examiner B, and the cases of Examiner D were grouped with cases of Examiner E in doing the analysis. One analysis of variance was done for each of the three kinds of Navy selection data. In the case of the Flight Aptitude Rating, the letters were assigned numerical values: A = 1; B = 2; C = 3; D = 4; and E = 5. The results for the three analyses of variance are presented in Table 10. The examiner differences obtained on the "Ability-to-Take-It" tests cannot be accounted for by differences in the quality of cadets tested by the different examiners, in so far as these three kinds of Navy selection data are measures of the quality of cadets tested.

c. Training of Examiners. Differences in the training of examiners may have affected many details of administration in both tests.[11]

---

[11] It is interesting to note that the time measures obtained by Examiners A and B (between 30 and 37 seconds) are considerably lower than those reported in the earlier study (52 seconds for both right and left hands). It may be doubted that much of this difference could be accounted for by the fact that the settings in the earlier study were determined by the strongest of three grips instead of by the stronger of two. Scores obtained by Examiners A and B on the constant stimulus shocker correspond rather closely to those reported in the earlier study. This correspondence may not be important, however, as a different shocker was used in the earlier study.

TABLE 10

DIFFERENCES IN QUALITY OF CADETS TESTED BY DIFFERENT EXAMINERS AS EVIDENCED BY NAVY SELECTION TESTS

Navy Selection Test	Examiner	N	Mean	Variance within Examiners	Variance between Examiners	F-value	P-value
Flight Aptitude Rating	A + B	63	2.62				Not Signif-
	C	28	2.79	1.26	1.96	1.56	icant
	D + E	422	2.87				
Aviation Classification Test	A + B	49	75.53				Not Signif-
	C	24	78.83	217.46	102.17	0.47	icant
	D + E	388	77.09				
Test of Mechanical Comprehension	A + B	53	52.81				Not Signif-
	C	26	53.31	30.27	26.90	0.89	icant
	D + E	423	52.57				

None of the analyses of variance show significant F-values.

Examiners A and B were both from the Statistical Office at the University of Rochester, where the tests were developed. One of them had a good deal to do with the construction of the tests and the development of the procedures. Both had been thoroughly trained in the administration of the tests. The training for the other examiners was not equally systematic, although the administration of the tests was discussed with them in detail and all tests were administered with the aid of written instructions, including verbatim instructions to the subjects.

d. Personality Differences Among the Examiners. It appears possible that in tests of the present sort personality differences among examiners might affect their rapport with the subjects or the degree of effort they would elicit, even though all examiners followed the same verbatim instructions to the subjects.

Factors Specific to the Hand Dynamometer Test.

a. Method of Determining Setting for Measure of Sustained Grip. It has already been noted that Examiners A and B required subjects to hold a grip 60% of the higher of the two grips achieved with a given hand in the earlier trials, whereas the other examiners set the requirement at the average of the two grips. The average values of strength of grip were somewhat higher for Examiners C, D, and E, however, with the result that the settings for the five examiners were not greatly different. The average settings actually used were 60% of the following values:

Examiner A	53.80
Examiner B	53.74
Examiner C	53.58
Examiner D	52.96
Examiner E	54.83

Thus it hardly seems that this difference in procedures could account for the large differences observed in the time measures.

b. Variations in other Details of Administration. It appears possible that examiners may have started the stop watch at different points in testing strength of grip. The instructions are not specific on this point.

c. Differences in Time Intervals Between Measures. The instructions for the sequence of tests required that four measures of strength of grip be obtained, alternating between preferred and non-preferred hands, and that the first measure of strength of grip then be obtained on the preferred hand. It was recognized that this procedure was not ideal, as the subject was undoubtedly fatigued at the time. The compromise appeared to be necessary in order to fit the test into the time limits of the total testing program. The examiners were aware of the difficulty, however, and it is possible that some examiners allowed a longer rest period than others. Differences in the speed with which they computed the setting and adjusted the instrument might also be factors.

Further differences were also undoubtedly present in the time interval between the time measure on the preferred hand and that on the non-preferred hand. The original manual of procedures for the study at Williams College provided that the time measure with the preferred hand was to be followed by the testing of the preferred hand on the constant stimulus shocker and that this in turn was to be followed in order by the time measure on the non-preferred hand and the shocker measure on the non-preferred hand. The instructions required a ten-minute rest between the first and second shocker measures and also between the first and second dynamometer measures. Shortly after the testing program was under way, however, it was recognized that the conditions of examination in the earlier study would be more closely approximated by the sequence outlined on pages 3-4. The investigators were requested to change to this procedure after Examiners A, B, and C had completed their testing and after Examiner D had tested 8 subjects and Examiner E had tested 14. As a consequence, the interval between the first time measure and the second time measure was several minutes shorter for the subjects tested by Examiners A, B, and C than it was for a large majority of the subjects tested by Examiners D and E.

Factors Specific to the Constant Stimulus Shocker Test.

a. Apparatus Fluctuations. Systematic apparatus fluctuations over a period of time might show up in differences among the three groups of subjects mentioned above (A & B, C, D & E). Special precautions had been taken in the construction of the instrument, however, to insure that current output would remain constant.

b. Mechanics of Administration. It will be recalled that the examiner was to practice turning the knob at the rate of one revolution per minute until he had an error of not more than four seconds. Any marked differences in rate or in the smoothness with which the knob was turned might seriously affect the results. No evidence is available on this point. It is also possible that examiners differed in their techniques of noting and recording readings of Threshold, Painful, and Very Painful.

VALIDATION

The marked differences among the five examiners who administered the "Ability-to-Take-It" Tests to the Williams College cadets made it inadvisable to combine the data for the different examiners in correlating the results with success in flight training. Since three of the examiners each tested too small a number of cases to justify validation, the present analysis is restricted to those cases tested by Examiners D and E. Results for each examiner are treated separately.

Nature of Criterion Data. The Bureau of Medicine and Surgery of the United States Navy made two examinations of the training records of the cadets. At the time of the earlier examination (December, 1944), no cadets had completed intermediate training and there were indications that many had not entered primary training. It was reasonably certain, however, that all subjects had passed through the WTS stage of training.[12] A preliminary analysis was made, therefore, in which only the records through WTS were considered, i.e., subjects who were eliminated in either Naval Flight Preparatory School or WTS were treated as "failers," while all others were treated as passers. Table 11 shows the number of passers and the frequency of elimination for each reason, for each examiner.

A second search for criterion data in June, 1945, revealed that many cadets had still not completed intermediate training. It seemed desirable, however, to use all data then available in a further check on the validity

---

[12] Naval flight training includes the following stages prior to commissioning: Naval Flight Preparatory School, WTS Flight Training, Preflight School, Primary Training, and Intermediate Training.

TABLE 11

FREQUENCY OF PASSERS AND OF ELIMINATIONS
FOR EACH REASON INDICATED

	Eliminations through WTS		All Eliminations through June, 1945	
Eliminations	Exam. D	Exam. E	Exam. D	Exam. E
Flight failure	19	13	39	31
Ground school training (academic)	10	26	14	35
Enforced withdrawal (July, 1944)	0	0	17	11
Voluntary withdrawal (July, 1944)	0	0	3	4
Dropped at own request	7	4	14	11
Disciplinary (including Marriage)	5	3	5	3
Psychologically unsuited	1	1	1	1
Not officer material	0	0	0	1
Not physically qualified on routine physical examination	1	0	2	1
Physical training	0	0	1	0
Killed in crash	0	0	0	1
Other reasons, including death not from crash	3	3	4	4
Reason unknown	0	0	0	1
Total Eliminations	46	50	100	104
Passers (including those still in training)	174	178	120	124
TOTAL	220	228	220	228

of the tests. Table 11 includes a tabulation of the frequency of elimination for each reason in this second analysis.[13]

Results will be presented first for the analysis through WTS and then for the analysis using all data available in June, 1945.

Correlations with Success in Flight Training Through WTS. Table 12 presents biserial coefficients of correlation between performance through WTS training and each of the measures on the hand dynamometer. The table also shows data for a score obtained by summing the time measures for preferred and non-preferred hands.[14]

The biserial coefficients of correlation range from .08 to .24. No completely adequate method of evaluating the statistical significance of the biserials is available.[15] It is possible, however, to use the formula

$$\sigma_{bis} = \frac{\left(\frac{\sqrt{pq}}{z} - r^2\right)}{\sqrt{N}}$$

---

[13] Three items in Table 11 require explanation: "enforced withdrawal," "voluntary withdrawal," and "dropped at own request." In July, 1944, records of all cadets in preflight school were examined by the U. S. Navy to determine which cadets should be eliminated in a restriction of the training program. Attention was given to information on most aspects of performance ordinarily leading to elimination. Some weight was also given to the flight aptitude rating derived from the Navy selection battery. Cadets withdrawn at the time of this "deselection" program were referred to as enforced withdrawals. Such cadets were given the privilege of re-entering flight training when the training program was later expanded. Those who re-entered training by June, 1945, are listed in Table 11 as "passers," or, if they have already been finally eliminated, are listed under a specific reason for elimination. The 23 cases listed as "enforced withdrawals" in Table 11 are those who had not re-entered flight training by June, 1945.

At the time of the deselection program other cadets were given an opportunity to withdraw voluntarily with the privilege of re-entering flight training when quotas were increased. Cadets who availed themselves of this privilege but who had still not re-entered flight training in June, 1945, are listed as voluntary withdrawals. These cadets are to be distinguished from those dropped at own request. This reason for elimination is not confined to the restriction of the training program in July, 1944. Cadets may ask to be dropped at any stage of flight training, but cadets who are dropped are ordinarily not allowed to re-enter flight training at a later date.

[14] It was recommended in the original report (J. W. Dunlap, et al., op. cit.) that the average of two trials with the preferred hand be used. The present procedure was adopted, however, as approximating more closely the conditions of the original study.

[15] If the significances of differences between passers and failers were of particular interest they could be more accurately determined by the t-test.

TABLE 12

VALIDATION DATA, HAND DYNAMOMETER MEASURES*
(Eliminations through WTS)

	Examiner	M_p	M_f	σ_{Total}	r_{bis}
Grip, First Measure, Preferred Hand	D	54.26	51.22	7.26	.24
	E	55.48	54.55	7.02	.08
Grip, Second Measure, Preferred Hand	D	52.95	50.02	7.32	.23
	E	54.64	53.70	6.58	.08
Time, Preferred Hand	D	50.90	48.41	15.30	.09
	E	47.36	45.02	17.13	.08
Time, Non-preferred Hand	D	54.76	49.37	16.89	.18
	E	49.19	45.98	16.81	.11
SUMMED SCORE Time (2 measures)	D	105.66	97.78	27.57	.16
	E	96.55	91.00	29.39	.11

*Number of cases

Passers: D = 174; E = 178
Failers: D = 46; E = 50
Total: D = 220; E = 228

as an approximation.[16] As this approximation is rough, it is advisable that only correlations greater than three times the standard error of a biserial of .00 be considered significant. According to this criterion the minimum value to be considered for Examiner D is .29; for Examiner E, .28. None of the obtained coefficients quite reaches this level. The two highest biserial correlations are for strength of grip on the hand dynamometer, a measure which was not designated as a test of "ability-to-take-it." These two correlations are for Examiner D. On the same measure Examiner E shows quite low correlations (.03 for both first and second measures of strength of grip). The correlations for the summed time scores are .16 for Examiner D and .11 for Examiner E.

Table 13 presents similar information for the constant stimulus shocker test. Information is included for a score obtained by adding the terminal scores for preferred and non-preferred hands.[17] The coefficients in Table 13 range from -.04 to .13. No correlations approach significance. Furthermore, the fact that all but two correlations are positive is not important because all measures are fairly highly intercorrelated. The biserial correlations for the summed terminal scores are .13 for Examiner D and .04 for Examiner E.

Multiple correlations were computed separately for each examiner, using the summed scores for each test. These multiple correlations were only .23 for Examiner D and .12 for Examiner E.

Correlations with Success in Flight Training to June, 1945. Table 14 presents biserial coefficients of correlation for the hand dynamometer test, all eliminations recorded in June, 1945, being treated as failers. This table includes three additional scores not analyzed earlier in the report, namely, two measures of strength of grip with the non-preferred hand, and a score obtained by adding four measures of strength of grip, two with each hand.[18] In general, the coefficients do not differ greatly from those presented in Table 12, although those for Examiner E tend to increase somewhat. With this division of passers and failers, coefficients of .25 are three standard errors above zero for both examiners. None of the coefficients in the table attains significance according to this test. The correlation for the summed grip score is .21 for each examiner. The correlation for the summed time score is .21 for Examiner D and .14 for Examiner E.

---

[16] Soper, H. E. On the probable error of the biserial expression for the correlation coefficient. Biom., 1914, 10.

[17] The original report (J. W. Dunlap, et al., op. cit.) recommended taking the average of four terminal scores, two for each hand. Time limitations made it impossible to extend the testing to obtain additional measures.

[18] It was not possible to include these scores in the earlier treatment of examiner differences because original data on the non-preferred hand were not available for the first subjects tested.

TABLE 13

VALIDATION DATA, CONSTANT STIMULUS SHOCKER MEASURES*
(Eliminations through WTS)

	Examiner	M_p	M_f	σ_{Total}	r_{bis}
Threshold, Preferred Hand	D	1.41	1.40	.28	.02
	E	1.90	1.92	.28	-.04
Painful, Preferred Hand	D	4.48	4.14	1.55	.13
	E	5.01	4.82	1.32	.08
Very Painful, Preferred Hand	D	6.21	5.96	1.82	.08
	E	6.66	6.56	1.39	.04
Terminal, Preferred Hand	D	8.43	7.89	1.89	.16
	E	8.02	7.97	1.60	.02
Threshold, Non-preferred Hand	D	1.46	1.46	.28	.00
	E	1.97	1.97	.32	-.01
Painful, Non-preferred Hand	D	4.13	3.97	1.27	.07
	E	4.82	4.71	1.22	.05
Very Painful, Non-preferred Hand	D	5.75	5.43	1.51	.12
	E	6.46	6.29	1.37	.07
Terminal, Non-preferred Hand	D	8.00	7.43	1.76	.18
	E	7.91	7.78	1.57	.05
SUMMED SCORE Terminal (2 measures)	D	16.42	15.32	3.46	.18
	E	15.93	15.74	3.01	.04

*Number of cases

Passers:	D = 174; E = 178
Failers:	D = 46; E = 50
Total:	D = 220; E = 228

TABLE 14

VALIDATION DATA, HAND DYNAMOMETER MEASURES*
(All Eliminations to June, 1945)

	Examiner	M_p	M_f	σ_{Total}	r_{bis}
Grip, First Measure, Preferred Hand	D	54.68	52.37	7.26	.20
	E	56.07	54.31	7.02	.16
Grip, Second Measure, Preferred Hand	D	53.51	50.93	7.32	.22
	E	55.41	53.27	6.58	.20
Grip, First Measure, Non-preferred Hand	D	50.86	48.87	7.12	.18
	E	52.76	50.62	7.09	.19
Grip, Second Measure, Non-preferred Hand	D	49.20	47.40	6.84	.16
	E	51.93	49.71	6.75	.21
Time, Preferred Hand	D	52.94	47.30	15.30	.23
	E	49.25	43.98	17.13	.19
Time, Non-preferred Hand	D	55.30	51.63	16.89	.14
	E	49.14	47.70	16.81	.05
SUMMED SCORE Grip (4 measures)	D	208.24	199.57	26.31	.21
	E	216.22	207.91	24.82	.21
SUMMED SCORE Time (2 measures)	D	108.24	98.93	27.57	.21
	E	98.40	91.68	29.39	.14

*Number of cases

Passers: D = 120; E = 124
Failers: D = 100; E = 104
Total: D = 220; E = 228

Table 15 presents similar information for the constant stimulus shocker. In this table the coefficients are uniformly low. The coefficients for the summed terminal scores are .06 for Examiner D and .08 for Examiner E.

Special Analysis of Flight Failures, using all Available Data. It has been suggested that the measures of "ability-to-take-it" might not be expected to predict all the causes of elimination from flight training. Analyses of variance were performed for each measure, therefore, in which flight failers were distinguished from those eliminated for other reasons, and from passers.

Results for the hand dynamometer measures appear in Table 16. Seven of the 16 analyses are significant at the 5% level, and one at the 1% level. Inspection of the mean values, however, reveals that results are not consistent, and further that a low P-value can seldom be attributed to the inferiority of flight failers. On the summed time scores, for example, the mean value for flight failers tested by Examiner D is below that for passers, but slightly above that for other failers. The flight failers tested by Examiner E scored somewhat higher than the passers, and other failers scored considerably lower. Similar inconsistencies are also seen in the single time measures. On the summed grip scores the flight failers tested by both examiners scored slightly below other failers, but this difference was much less than that between passers and all failers. The same situation is found in seven of the eight analyses of single grip measures. In the eighth analysis the other failers score somewhat higher than flight failers, but the chief difference is still between passers and all failers.

Results of the analyses of variance for the constant stimulus shocker measures appear in Table 17. None of the 18 analyses is significant at the 5% level, and there is no consistency in the relative positions of the three groups.

The results indicate that a further analysis of flight failers as a separate group would not be profitable on either test.

Discussion. The validation study revealed no biserial coefficients of correlation more than three standard errors above a coefficient of zero. The analyses of variance comparing means of passers, means of flight failers, and means of "other failers" showed few significant P-values and revealed that the flight failers were not uniformly lower than other failers or passers.

The biserials for the constant stimulus shocker (including all eliminations to June, 1945) are uniformly very low and sometimes negative. The biserials for the hand dynamometer measures are positive and slightly higher, although not significantly so. The general agreement in magnitude and in direction of the coefficients for both examiners for the summed grip measure and the summed time measure, as well as the evidence from the analyses of variance, may suggest that the hand dynamometer test is tapping some aspects of behavior which are related to success in flight training. The

TABLE 15

VALIDATION DATA, CONSTANT STIMULUS SHOCKER MEASURES*
(All Eliminations to June, 1945)

	Examiner	M_p	M_f	σ_{Total}	r_{bis}
Threshold, Preferred Hand	D	1.41	1.41	.28	-.01
	E	1.92	1.89	.28	.08
Painful, Preferred Hand	D	4.36	4.48	1.55	-.05
	E	5.06	4.87	1.32	.09
Very Painful, Preferred Hand	D	6.01	6.34	1.82	-.11
	E	6.72	6.55	1.39	.08
Terminal, Preferred Hand	D	8.32	8.30	1.89	.01
	E	8.14	7.36	1.60	.11
Threshold, Non-preferred Hand	D	1.44	1.47	.28	-.06
	E	1.96	1.98	.32	-.02
Painful, Non-preferred Hand	D	4.08	4.11	1.27	-.02
	E	4.91	4.67	1.22	.12
Very Painful, Non-preferred Hand	E	5.72	5.64	1.51	.03
	E	6.50	6.32	1.37	.08
Terminal, Non-preferred Hand	D	8.01	7.72	1.76	.10
	E	7.93	7.82	1.57	.04
SUMMED SCORE Terminal (2 measures)	D	16.33	16.02	3.46	.06
	E	16.07	15.68	3.01	.08

*Number of cases
Passers: D = 120; E = 124
Failers: D = 100; E = 104
Total: D = 220; E = 228

TABLE 16

COMPARISON OF PASSERS, FLIGHT FAILERS, AND OTHER FAILERS: HAND DYNAMOMETER*
(All Eliminations to June, 1945)

		EXAMINER D			EXAMINER E		
		Means	F-Value	P-Value	Means	F-Value	P-Value
Grip, First Measure, Preferred Hand	Passers	54.68		Not significant	56.07		Not significant
	Flight Failers	52.26	2.76		54.07	1.84	
	Other Failers	52.44			54.41		
Grip, Second Measure, Preferred Hand	Passers	53.51		Between .05 & .01	55.41		Between .05 & .01
	Flight Failers	50.80	3.43		53.07	3.13	
	Other Failers	51.02			53.35		
Grip, First Measure, Non-preferred Hand	Passers	50.86		Not significant	52.76		Not significant
	Flight Failers	48.28	2.26		50.30	2.71	
	Other Failers	49.24			50.76		
Grip, Second Measure, Non-preferred Hand	Passers	49.20		Not significant	51.98		Between .05 & .01
	Flight Failers	47.10	1.90		50.00	3.34	
	Other Failers	47.59			49.60		
Time, Preferred Hand	Passers	52.94		Between .05 & .01	49.25		
	Flight Failers	48.67	4.06		49.37	4.92	< .01
	Other Failers	46.42			41.80		
Time, Non-preferred Hand	Passers	55.30		Not significant	49.14		Not significant
	Flight Failers	50.92	1.33		52.03	1.62	
	Other Failers	52.08			45.95		
SUMMED SCORE Grip (4 measures)	Passers	208.24		Not significant	216.22		Between .05 & .01
	Flight Failers	198.44	3.02		207.43	3.32	
	Other Failers	200.29			208.12		
SUMMED SCORE Time (2 measures)	Passers	108.24		Between .05 & .01	98.40		Between .05 & .01
	Flight Failers	99.59	3.17		101.40	3.87	
	Other Failers	98.51			87.75		

*Number of cases

Passers: D = 120: E = 124
Flight Failers: D = 39: E = 30
Other Failers: D = 61: E = 74

TABLE 17

COMPARISON OF PASSERS, FLIGHT FAILERS AND OTHER FAILERS: CONSTANT STIMULUS SHOCKER*
(All Eliminations to June, 1945)

		EXAMINER D			EXAMINER E		
		Mean	F-Value	P-Value	Mean	F-Value	P-Value
Threshold, Preferred Hand	Passers	1.41		Not significant	1.92		Not significant
	Flight Failers	1.37	0.46		1.81	2.04	
	Other Failers	1.43			1.92		
Painful, Preferred Hand	Passers	4.36		Not significant	5.06		Not significant
	Flight Failers	4.59	0.50		4.97	0.71	
	Other Failers	4.41			4.83		
Very Painful, Preferred Hand	Passers	6.00		Not significant	6.72		Not significant
	Flight Failers	6.56	1.39		6.72	0.73	
	Other Failers	6.19			6.48		
Terminal, Preferred Hand	Passers	8.32		Not significant	8.14		Not significant
	Flight Failers	8.50	0.35		8.12	1.48	
	Other Failers	8.17			7.75		
Threshold, Non-preferred Hand	Passers	1.44		Not significant	1.96		Not significant
	Flight Failers	1.50	0.67		1.92	0.85	
	Other Failers	1.45			2.00		
Painful, Non-preferred Hand	Passers	4.08		Not significant	4.91		Not significant
	Flight Failers	4.07	0.06		4.71	1.14	
	Other Failers	4.14			4.65		
Very Painful, Non-preferred Hand	Passers	5.72		Not significant	6.50		Not significant
	Flight Failers	5.63	0.09		6.46	0.71	
	Other Failers	5.64			6.27		
Terminal, Non-preferred Hand	Passers	8.01		Not significant	7.93		Not significant
	Flight Failers	7.95	1.24		8.13	0.93	
	Other Failers	7.58			7.70		
SUMMED SCORE Terminal (2 measures)	Passers	16.33		Not significant	16.07		Not significant
	Flight Failers	16.45	0.68		16.25	1.26	
	Other Failers	15.75			15.45		

*Number of cases

Passers: D = 120; E = 124
Flight Failers: D = 39; E = 30
Other Failers: D = 61; E = 71

present data, however, do not indicate that the measures in their present form would increase the effectiveness of established test batteries.

SUMMARY

The first part of this report is concerned with an analysis of differences in scores obtained by five examiners in the administration of two tests of "ability-to-take-it." The analysis was undertaken before comparing the scores with records of performance in flight training, because it was known that procedures were not completely uniform throughout the test.

The results indicate striking differences, for which a number of possible explanations are offered, including variations in examining procedures, changes in the sequence of tests, apparatus fluctuations, personality differences among the examiners, changes due to discussion among the cadets, and variations in the quality of cadets.

No effort is made in this report to estimate with any finality the relative importance of the various factors suggested, since the study was not designed to furnish an answer to such a question. It is of interest that the difficulties encountered in the present study arose even though the various individuals concerned were all well aware of the need for standardized conditions.

The findings serve to re-emphasize the well recognized importance of assuring that testing conditions and apparatus are completely standardized and examiners thoroughly trained. Such assurance is especially necessary before extensive validation studies are undertaken.

The second part of this report is concerned with the validation of the "ability-to-take-it" tests. Validation was done separately for the two examiners with the greatest number of cases by correlating measures on the tests with success in flight training through WTS, and for all data available to June, 1945. In addition, a special analysis was done of flight failures, because it was suggested that the measures of "ability-to-take-it" might not be expected to predict all the causes of elimination from flight training. There is a suggestion that the "Hand Dynamometer" test is tapping some aspects of behavior which are related to success in flight training. The general conclusion may be made that the "ability-to-take-it" tests could not predict the criteria employed.

APPENDIX A

INSTRUCTIONS FOR ADMINISTRATION OF THE HAND DYNAMOMETER TEST

APPENDIX A

INSTRUCTIONS FOR ADMINISTRATION OF THE HAND DYNAMOMETER TEST

Say: This is a hand dynamometer for the measurement of strength of grip. We are interested in seeing just how strong your grip is as compared with your fellow man. Two trials for each hand is all that is required.

Keep the instrument away from your body. Adjust the handle to the point where you feel you can give your best performance. Squeeze as hard as you can for a moment before relaxing.

(Examiner demonstrated by adjusting handle to his best position, places some magnesium carbonate on his hand so that the instrument does not slip; places dynamometer about six inches from body -- arm slightly bent with elbow also about six inches from body. Squeezes momentarily about as hard as he can.)

Say: You see how it is done. Now you try it.

(Rubs chalk on subject's hands and sees that he assumes correct position.)

Say: Measurements of the strength of grip of both hands are required. Are there any questions? Are you right handed or left handed?

(First test with preferred hand.)

Say:
1. The right hand first please.....
2. The left hand.....
3. Right hand.....
4. Left hand.....

(Measurements of maximum grip are recorded for both right and left hands as well as the reading of grip set.)

Say: We are also interested in seeing how long you can hold the dynamometer above a certain point. I am going to place the indicator at that point.

(Indicator is placed at 60 per cent of maximum grip -- for each hand, respectively.)

Say: You are to squeeze the handle until the amber lights up. Do not squeeze too hard, for you just waste your strength. Keep the amber light glowing as long as you can. When your grip weakens, the red light will go on. The red light is a warning signal; therefore, increase your grip so that the amber light is again on. When the red light goes on, fight as long as you can to keep your grip from getting weaker. The buzzer sounding after the red light is the end of the test. Keep the dynamometer away from your body. Watch the lights. Try to keep your grip as steady as possible. Any questions? When I say go, start.....

(Time is called off at five second intervals. At finish of test, total time and grip set are recorded.)

Say: After a ten minute rest, the endurance of your other hand will be tested.

(After ten minutes, proceed similarly with untested hand. The shock test is administered between the fatigue tests for each hand.)

APPENDIX B

INSTRUCTIONS FOR ADMINISTRATION OF CONSTANT STIMULUS SHOCKER

APPENDIX B

INSTRUCTIONS FOR ADMINISTRATION OF CONSTANT STIMULUS SHOCKER

Say: We wish to determine how much pain you can stand in terms of electric shock. You will place two fingers in the salt solution after which we will gradually increase the amount of electric shock. When you first place your fingers in the solution, there will be no sensation of shock or pain since the current will be cut off. As the current is increased, you will be asked to answer a series of four questions regarding your sensations. This test will gradually increase in painfulness until you cannot bear it, unless you have a lot of "guts." What we want to see is whether or not you can "take it." Remember that, although the test will become very painful, it will not hurt you permanently. Now may I see your right hand.

(Examiner carefully examines the index and middle fingers for cuts or hangnails. If none is apparent, immerse the fingers in the salt solution. If cuts or hangnails are apparent, select two fingers free from these imperfections. Have the assistant blindfold the subject.)

Say: I want you to tell me four things--first, when you feel any sensation at all; second, when the sensation becomes painful--by painful, I mean whatever you consider painful; third, I want you to tell me where the sensation becomes very painful but where you think you could stand some more if you had to; and fourth, you are to report where you can no longer take it. Then I will immediately cut off the current. Do not take your hand out of the solution until I tell you to do so. Now I am going to start the test.

(Examiner turns the rheostat control at the rate of one complete revolution every 60 seconds. He should have practiced this with a stop watch until he has an error of not more than 4 seconds.)

Say: Please tell me when you first feel sensation.

(No more comments should be made until the student reports sensation. If he reports sensation below 5 on the scale, stop the test. It means there is a minute cut on one of the fingers. Select another finger of the same hand and restart the test. When sensation is reported at a scale value of 5 or more, record this value.)

Say: Now I am going on and you report where it is painful.

(No comments until the student reports pain. Then record dial reading.)

Say: Now we are going on until you report the sensation is very painful but you think you can stand some more.

(No comments until the student reports very painful. Record dial reading.)

Say: Now we are going on until you cannot stand it any longer. Let us see if you can take it.

(No comments. As soon as subject says stop, cut the current by reversing the rheostat. After 10 minute interval or more, test the other hand.)