REPORT ON THE BOSTON-MIDWEST PROJECT

Prepared

by

National Research Council
Committee on Selection and
Training of Aircraft Pilots

November 1945

CIVIL AERONAUTICS ADMINISTRATION
Division of Research
Report No. 52
Washington, D. C.

National Research Council

Committee on Selection and Training of Aircraft Pilots

Executive Subcommittee

M. S. Viteles, Chairman

E. C. Andrus

C. W. Bray

D. R. Brimhall

J. W. Dunlap

J. C. Flanagan

W. R. Miles

P. J. Rulon

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

November 27, 1945

Dr. Dean R. Brimhall
Asst. to Administrator for Research
Civil Aeronautics Administration
Room 5835, Commerce Building
Washington 25, D. C.

Dear Dr. Brimhall:

In 1941-1942, the Committee on Selection and Training of Aircraft Pilots conducted, in and around Boston, Massachusetts, and Columbus, Ohio, an integrated program of research involving the evaluation of predictors and criteria found to be promising in earlier research. The attached report, entitled Report on the Boston-Midwest Project, provides an analysis of the data accumulated in the course of this investigation. The report was prepared and is submitted by the Committee on Selection and Training of Aircraft Pilots with the recommendation that it be included in the series of Technical Reports issued by the Division of Research, Civil Aeronautics Administration.

The study failed to produce material of striking significance with respect to the predictors employed in the investigation. On the other hand, advances are indicated in the area of criterion measures, particularly with respect to the possibility of using a composite criterion as a means of more fully describing flight performance. This study has additional interest as another in the series sponsored by the Committee on Selection and Training of Aircraft Pilots involving a coordinated program of civilian research in aviation psychology, conducted in scattered centers, with the cooperation of local institutions and psychologists, financed by a government agency, the Civil Aeronautics Administration, through the National Research Council.

Cordially yours,

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

MSV:rm

EDITORIAL FOREWORD

Early studies conducted under grants from the Committee on Selection and Training of Aircraft Pilots, particularly those carried on during 1940 and the early part of 1941, were largely exploratory in character. These included preliminary surveys of a large number of predictors[1] and initial experiments on the development of improved criteria[2] including such instruments as the Ohio State Flight Inventory,[3] the Purdue Rating Scale,[4] and graphic and photographic techniques for the evaluation of flight performance.[5]

---

[1]McFarland, R. A., and Franzen, R. The Pensacola study of naval aviators. Final summary report. Washington, D. C.: CAA Division of Research, Report No. 38, November 1944.

McFarland, R. A., and Franzen, R. A revised serial reaction time apparatus for use in appraising flying aptitude. Washington, D. C.: CAA Airman Development Division, Report No. 34, September 1944.

McFarland, R. A., and Franzen, R. A revised two-hand coordination test. Washington, D. C.: CAA Airman Development Division, Report No. 36, October 1944.

Johnson, H. M., in cooperation with Boots, M.-L., and Wherry, R. J.; with the assistance of Rotaling, C. C., Martin, L. G., and Cassens, F. P., Jr. On the actual and potential value of biographical information as a means of predicting success in aeronautical training. Washington, D. C.: CAA Airman Development Division, Report No. 32, August 1944.

Foley, J. P., Jr., Hunt, T., Kelly, E. L., and Lepley, W. M. Studies of predictors of achievement in learning to fly. Washington, D. C.: CAA Division of Research, Report No. 27, March 1944.

[2]Johnson, H. M., and Boots, M. L. Analysis of ratings in the preliminary phase of the C.A.A. training program. Washington, D. C.: CAA Division of Research, Report No. 21, October 1943.

[3]Edgerton, Harold A., and Walker, Robert Y. History and development of the Ohio State Flight Inventory. Part I: Early versions and basic research. Washington, D. C.: CAA Division of Research, Report No. 47, July 1945.

National Research Council Committee on Selection and Training of Aircraft Pilots. History and development of the Ohio State Flight Inventory, Part II: Recent versions and current applications. Washington, D. C.: CAA Division of Research, Report No. 51, November 1945.

[4]Kelly, E. L. The development of "A Scale for Rating Pilot Competency." Washington, D. C.: CAA Division of Research, Report No. 18, July 1943.

[5]Viteles, M. S., and Backstrom, O., Jr. An analysis of graphic records of pilot performance obtained by means of the R-S Ride Recorder. Part I. Washington, D. C.: CAA Division of Research, Report No. 23, November 1943.

Viteles, M. S., and Backstrom, O., Jr. An analysis of graphic records of pilot performance obtained by means of the R-S Ride Recorder. Part II. (A final report in preparation for publication in the CAA Technical Series.)

Viteles, M. S., and Thompson, A. S. An analysis of photographic records of aircraft pilot performance. Washington, D. C.: CAA Division of Research, Report No. 31, July 1944.

In the fall of 1941, it was decided to undertake an integrated program of research involving field tryout, with as large a number of cases as were available, of the more promising selection and criterion instruments developed in the course of early research. As part of this program, centers were established at Boston, Massachusetts, and at Columbus, Ohio, to conduct the program of coordinated research which has come to be known as the Boston-Midwest Project.

The work at Boston was under the direction of Ross A. McFarland, Graduate School of Business Administration, Harvard University, and involved students undergoing pilot training drawn from Harvard University; the Massachusetts Institute of Technology; Northeastern University; Boston College; Boston University; and Tufts College, Medford, Massachusetts.

Research administered from Columbus, Ohio, was under the direction of Robert Y. Walker, Ohio State University, and involved students taking flight training at Ohio State University, Columbus, Ohio; University of Michigan, Ann Arbor, Michigan; the University of Toledo; the University of Dayton; Purdue University, Lafayette, Indiana; Tri-State College, Angola, Indiana; and Indiana State Teachers College, Muncie, Indiana. E. L. Kelly, Project Director, and the staff conducting Committee research at Purdue University, cooperated in the administration of the program at that university.

All students acting as subjects were undergoing training under the Civilian Pilot Training Program, financed by the Civil Aeronautics Administration. Dean R. Brimhall, Director of Research, and John P. Morris, Director, Civil Pilot Training Program, as well as other members of the staff of the Civil Aeronautics Administration were largely responsible for facilitating the program and for coordinating the research activity with the operating program of that agency.

The general plan for research was formulated by the Executive Subcommittee of the Committee on Selection and Training of Aircraft Pilots. Jack W. Dunlap, Director of Research, was responsible for preparing the details of the research program and for field supervision of the Boston-Midwest Project. However, in a large sense, this project was the outcome of cooperative effort of the many psychologists who had participated in the program of the Committee on Selection and Training of Aircraft Pilots. The predictors, for example, selected for study were drawn from investigations earlier conducted by E. Lowell Kelly, Purdue University; Ross A. McFarland, Harvard University; Raymond Franzen, New York City; Jack W. Dunlap, University of Rochester; G. Richard Wendt, Wesleyan University; Edward B. Greene and Clarke W. Crannell, University of Michigan, and others. The criterion instruments employed grew out of earlier research by E. Lowell Kelly, Purdue University; Harold A. Edgerton and Robert Y. Walker, Ohio State University; Morris S. Viteles and Albert S. Thompson, University of Pennsylvania.

It was originally planned that Boston would be the major center of the study in that the Contact Link Trainer and the photographic techniques were

to be employed there and not at Columbus. When, with our entry into World War II it became evident that flying restrictions would be imposed on the East Coast, certain phases of the investigation originally assigned to Boston were transferred to Columbus. Later, when a ban was put on flying on the Atlantic Coast, Boston was eliminated as a center of research, the number of supplementary centers in the Midwest area expanded, and the research entirely centered in the Midwest.

Much of the data on predictors presented in this study is only of historical interest because of more recent work by the Committee on Selection and Training of Aircraft Pilots and, more particularly, by reason of the vast amount of research on certain of the instruments described in this report subsequently conducted by the military services. However, there is much of basic interest in the data relating to criteria which include a number of criterion measures unique to this and other studies conducted by the Committee on Selection and Training of Aircraft Pilots. In addition, the report of this investigation is of interest as an example of a coordinated program of civilian research in aviation psychology, conducted in scattered centers, with the cooperation of local institutions and psychologists, financed by a government agency, the Civil Aeronautics Administration, through the National Research Council.

Acknowledgment is due to Edwin S. Ewart, Editorial Staff, Committee on Selection and Training of Aircraft Pilots, for collating the data and writing this report on the Boston-Midwest Project.

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

CONTENTS

CONTENTS (Continued)

SUMMARY

The Boston-Midwest Project, conducted under the suspices of the National Research Council Committee on Selection and Training of Aircraft Pilots, was designed to provide a field trial for a number of promising techniques for the selection of pilots, and to provide further data on certain criterion instruments for the evaluation of flight performance. The project was conducted at Boston, Massachusetts, and Columbus, Ohio. The research was continued over two Civilian Pilot Training Programs at both schools, the Fall 1941 and Spring 1942 programs in the Boston area, and the 1942 Spring and Summer programs in the Midwest area, centering around Columbus, Ohio.

All subjects were enrolled for primary flight training in the Civilian Pilot Training Program, and all were college students with the exception of 30 subjects in the Boston Spring program who were of college age, but were working in various industries in the Boston area. None of the subjects had had previous flight experience. The number of subjects used in the various analyses varied from 82 to 105, and from 49 to 81 in the Boston Fall and Spring samples, respectively, and from 20 to 103 and from 9 to 96 in the Midwest Spring and Summer samples, respectively. In certain instances the Midwest group data from Spring and Summer samples were pooled, yielding a group of 185 subjects.

The following paper-and-pencil tests were administered:

1. The Personal History (A-H) Inventory
2. The Otis Self Administering Test of Mental Ability (Form B)
3. The Bennett Test of Mechanical Comprehension (Form B)
4. The Biographical Inventory (The Inventory of Personal Data for Prospective Pilots)
5. The Strong Vocational Interest Blank for Men (Form E)

The Aviation Interview, developed under Committee auspices, was administered to a part of the subjects in the Boston area and to a part of the subjects in the Midwest area.

The psychomotor tests employed in the investigation were:

1. The Two-Hand Coordination Test
2. The Eye-Hand Coordination Test
3. The Mashburn Serial Reaction Test

The physiological tests were concerned primarily with spirometric and vascular variables. Measures on the vascular variables were taken in an upright and supine position by means of a "Tilt Table." The physiological measures used in this investigation were as follows: Body Surface, Vital Capacity, Vital Capacity/Body Surface, Tidal Air, Tidal Air/Body Surface, Systolic Blood Pressure (lying), Systolic Blood Pressure Initial Change (lying), Systolic Blood Pressure Maximum Change, Systolic Blood Pressure Time to Maximum, Diastolic Blood Pressure Initial Change (up), Diastolic Blood Pressure (lying), Diastolic Blood Pressure Maximum Change, Diastolic Blood Pressure Time to Maximum, Pulse Rate (lying), Pulse Rate Initial Change, Pulse Rate Maximum Rate, Pulse Rate Maximum Change, Pulse Rate Time to Maximum, Pulse Pressure Maximum Change, Smallest Pulse Pressure, and Pulse Pressure Time.

Flight proficiency was evaluated in terms of the following criteria:

1. Pass-Fail, a dichotomous criterion the score on which was a function, as the name implies, of whether or not the subject passed the flight training course.

2. Time measures, i.e., Time for Stage A, Time for Stage D, and Total Time for the flight training course.

3. Purdue Rating Scale, an instrument which called for rating of the subject by the Flight Instructor on 14 items associated with flight proficiency.

4. Photographic criteria, criteria represented by measures yielded by photographic records of instruments which indicated the attitude and performance of the plane, taken during the execution of certain maneuvers in the standard flight.

5. Ohio State Flight Inventory, a check sheet on which the flight examiner or check pilot noted the various critical elements of the pilot's performance during a standard check flight.

Analysis of the data indicated that the three physiological variables Body Surface, Vital Capacity, and Vital Capacity/Body Surface yielded the highest reliabilities of the several predictor variables, the reliability coefficients ranging from .76 to .94. The reliabilities of the psychomotor tests were somewhat lower, although scores from the Two-Hand Coordination Test and the Mashburn Serial Reaction Test yielded reliability coefficients of .74 or greater on two of the three independent samples for which data were available. Of the paper-and-pencil tests, only the reliability of the Biographical Inventory was determined on the basis of data gathered in this investigation. No unequivocal conclusions can be drawn since the reliability coefficients varied from .60, in terms of a sample in which test and retest were separated by an interval of three months, to .94 in terms of a sample in which test and retest were separated by an interval of four hours.

None of the selection tests predicted any of the individual criteria consistently over all independent samples of subjects, and in general the correlation coefficients were not high. Measures on the Two-Hand Coordination Test, the Mashburn Serial Reaction Test, and the Test of Mechanical Comprehension appeared to show relatively the most promise.

As measures of flight proficiency the various criteria were in general positively related, although the correlation between criteria was not high, except when two variations of the same criterion measure were considered. In the interest of obtaining a better measure of flight proficiency, composite criteria were set up, individual criterion measures being combined to yield single criterion scores.

In general, the psychomotor tests predicted the composite criteria most efficiently, the Mashburn Test yielding consistently the highest correlations. Following the psychomotor tests, the Test of Mechanical Comprehension was probably the next in rank of efficiency as a predictor, with the physiological measures and the Interview being less efficient. With the exception of the Test of Mechanical Comprehension, the paper-and-pencil tests showed little or no correlation with the composite criteria.

On the basis of their reliabilities, their predictor-criteria correlations, and their interrelationships a number of variables were selected for inclusion in several batteries of predictors. Multiple correlations between batteries and the composite criteria ranged in general between .50 and .60. Further analysis indicated that in most cases the multiple correlation was maximized after the best two measures had been combined. In many cases the predictive value of the battery was little better than the predictive value of the best single test in the battery.

It was concluded that while interpretation of the results must be guarded in view of the small number of cases involved, the value of a more detailed determination of the most satisfactory batteries on a larger and more stable population was indicated.

REPORT ON THE BOSTON-MIDWEST PROJECT

INTRODUCTION

The purpose of the Boston-Midwest project, conducted under the auspices of the National Research Council Committee on Selection and Training of Aircraft Pilots, was to provide a field trial for a number of promising techniques for the selection of pilots, and to provide further data on certain criterion instruments for the evaluation of flight performance. A number of selection techniques and criterion instruments employed in this study represent earlier developments growing out of Committee research. Further, the Boston-Midwest project was planned to provide evidence bearing on certain conclusions reached during the course of earlier studies, e.g., the Pensacola Study.[1]

The project was conducted at two centers: Boston, Massachusetts, and Columbus, Ohio. Research at Boston was directed by R. A. McFarland. Research at Columbus was directed by R. Y. Walker. The research was continued over two Civilian Pilot Training Programs at both schools, the Fall 1941 and Spring 1942 programs at institutions in the Boston area, and the Spring and Summer programs in 1942 at schools in the Columbus area. Data from the Boston area were gathered at Harvard University, the Massachusetts Institute of Technology, Northeastern University, Boston College, Boston University, and Tufts College. Data from the Columbus (Midwest) area were gathered at Ohio State University, the University of Michigan, the University of Toledo, the University of Dayton, Purdue University, Tri-State College (Angola, Indiana), and Indiana State Teachers College (Muncie, Indiana).

SUBJECTS

All subjects were enrolled for primary flight training in the Civilian Pilot Training Program, and all were college students with the exception of 30 subjects in the Boston Spring program who had completed high school and were of comparable age to the college students, but who were not actually enrolled in any educational institution at the time. These subjects were working in various industries in the Boston area. None of the subjects had previous flight experience,[2] or had taken any of the selection tests in the experimental batteries.

---

[1] Descriptions of the procedures and findings of this project are discussed in: McFarland, R. A., and Franzen, R. The Pensacola study of naval aviators. Final summary report. Washington, D. C.: CAA Division of Research, Report No. 38, November 1944. Also see: Franzen, R., and McFarland, R. A. Detailed statistical analysis of data obtained in the Pensacola study of naval aviators. Washington, D. C.: CAA Division of Research, Report No. 41, January 1945.

[2] Editor's Note. It should be pointed out that the subjects in the Pensacola Study were naval cadets who had had 10 hours of previous dual flight instruction and had soloed before entering Pensacola Naval Air Station for further training. See: McFarland, R. A., and Franzen, R. Ibid.

The number of subjects on whom data are available for use in the analysis discussed in this report varies with the particular predictor or criterion variables under consideration. For the Boston Fall sample the N's vary from 82 to 105, in the Boston Spring sample, from 49 to 81. In certain analyses of the Midwest group, data from Spring and Summer samples are pooled; in which case an N of 185 is available. For each sample separately, the N's vary from 20 to 113 in the Spring sample, and from 9 to 96 in the Summer sample.[3]

PREDICTION TESTS

The selection instruments administered in this project were of four general types: paper-and-pencil tests, the interview, psychomotor tests, and physiological tests. The tests administered are listed below. It should be noted that a number of different scores were derived from certain of the individual tests. These various scoring procedures will not be outlined here, but will be described in later sections of the report when their nature is not self-evident.

Paper-and-Pencil Tests

The following paper-and-pencil tests were administered:[4]

1. P-H (Personal History) Inventory[5]
2. Otis Self Administering Test of Mental Ability (Form D)[6]
3. Bennett Test of Mechanical Comprehension (Form B)[7]

---

[3]In the analysis of specific variables in a related series the ranges of N for the various comparisons is not as great as indicated by this statement.

[4]For the procedures employed in administering the paper-and-pencil tests, the psychomotor tests, and the physiological tests, see: Dunlap, J. W. The Midwest Project, a description of experimental procedures. August 1942. (This manual is on file with the NRC Committee on Selection and Training of Aircraft Pilots.)

[5]Kogan, L. S., Wantman, M. J., and Dunlap, J. W. Analysis of the Personal History Inventory. Washington, D. C.: CAA Division of Research, Report No. 42, February 1945.

[6]Published by the World Book Company, Yonkers, N. Y.

[7]Published by the Psychological Corporation, New York City.

4. Biographical Inventory[8]

5. Strong Vocational Interest Blank for Men (Form M)[9]

Interview

A controlled "Aviation Interview"[10] was administered to the Spring group of cases in the Midwest sample obtained from Ohio State University, and the University of Michigan, and to the Boston project cases at Harvard University which comprised the Spring group at that institution. A sample rating sheet used in the study of the Aviation Interview is presented in Figure 1. The subjects were rated by the members of the interview board on the following scales:

A. Academic Background as Related to Flying

B. Family and Socio-Economic Background as Related to Flying

C. General Social Adjustment as Related to Flying

D. Desire to Fly

E. Hobbies and Diversions, and Outside Activities as Related to Flying

F. Athletic Activities and Coordination

G. Personality as Related to Flying

H. Appearance, Mannerisms, and Physique as Related to Flying

I. Fitness for Flight Training

---

[8]The "Biographical Inventory" is the Navy designation of this test, and will be used for convenience in this report. Its official title, in terms of the Committee's designation, is "Inventory of Personal Data for Prospective Pilots." For discussions of findings obtained with this test, see: Wantman, M. J. Report on the reliability of the Inventory of Personal Data for prospective pilots. December 1942. Also, Odbert, H. S. Sample analysis of responses to the Inventory of Personal Data for prospective pilots. January 1944. (Progress reports in the files of the NRC Committee on Selection and Training of Aircraft Pilots.)

[9]Published by the Stanford University Press, Stanford University, California.

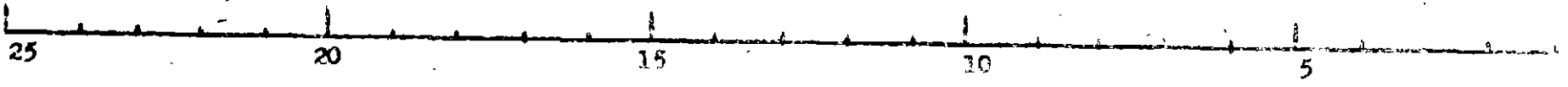
[10]The details of the interview procedure are given in: Dunlap, J. W., and Wantman, M. J. An investigation of the interview as a technique for selecting aircraft pilots. Washington, D. C.: CAA Airman Development Division, Report No. 33, August 1944.

Name of Interviewer _____

INTERVIEW CHART

Name of Subject _____

D. DESIRE TO FLY



25	20	15	10	5
AN ARDENT, IRREPRESSIBLE DRIVE TO BECOME A FLYER	A STRONG INTEREST IN FLYING	A MILD INTEREST IN FLYING	ONLY A SLIGHT INTEREST IN BECOMING A FLIER	NO FUNDAMENTAL INTEREST IN FLYING
___has a very strong and sincere interest in becoming a flier; ___would rather fly than eat; ___bubbling over with enthusiasm for flying; ___knows a lot about flying already; ___can't wait until he gets in the air; ___heart is set on flying; ___has always wanted to fly	___has a genuine desire to fly; ___wants very much to fly; ___has paid for flying instruction; ___has read widely about aviation	___would like to be a flier; ___is an average candidate; ___has a moderate desire to become a flier	___is nonchalant about flying as a career; ___prefers other things to flying; ___is prompted primarily by money	___has pseudo-interest in becoming a flier; ___is a "draft-dodger"; ___drifted into flying; ___flying is a means to some other end; ___is only curious about flying

EXPLANATION OF RATING:

_____

_____

_____

_____

FIGURE 1

Psychomotor Tests

The following psychomotor tests were administered:

Two-Hand Coordination Test[11]

Eye-Hand Coordination Test

Mashburn Serial Reaction Test[12]

Physiological Tests

The physiological tests were concerned primarily with spirometric and vascular variables. Measures on the vascular variables were taken in an upright and in a supine position by means of the "Tilt Table."[13] The physiological measures are listed below.[14]

Body Surface	Systolic Blood Pressure Maximum Change
Vital Capacity	Systolic Blood Pressure Time to Maximum
Vital Capacity/Body Surface	Diastolic Blood Pressure Initial Change (lying)
Tidal Air	Diastolic Blood Pressure (lying)
Tidal Air/Body Surface	Diastolic Blood Pressure Maximum Change
Systolic Blood Pressure (lying)	Diastolic Blood Pressure Time to Maximum
Systolic Blood Pressure Initial Change (lying)	Pulse Rate (lying)

---

[11]For a description of the two-hand coordination Test and an analysis of test results obtained on the original and revised forms of the test employing naval cadets and civilian students, see: McFarland, R. A., and Channell, R. C. A revised two-hand coordination test. Washington, D. C.: CAA Airman Development Division, Report No. 26, October 1944.

[12]A revised model of the Mashburn Serial Action Test and an analysis of test data obtained on naval cadets and civilian students is presented in: McFarland, R. A., and Channell, R. C. A revised serial reaction time apparatus for use in appraising flying aptitude. Washington, D. C.: CAA Airman Development Division, Report No. 31, September 1944.

[13]An analysis of the tilt table test obtained on naval cadets at Pensacola is described in: Preston, J. B. Statistical analysis of the tilt table test of cardiovascular adjustments in the normal subject of naval aviators. (Progress report in the files of the NDRC Committee on Selection and Training of Aircraft Pilots.)

[14]Specific procedures [illegible] and for administering the Tilt Table [illegible] Preston, J. B. et al. (Referred to in footnote 13.)

Pulse Rate Maximum Change — Pulse Pressure Change

Pulse Rate Maximum Change — Smallest Pulse Pressure

Pulse Rate Maximum Change — Pulse Pressure Time

Pulse Rate Time to Maximum

CRITERIA

The following criteria were used in this research:

Pass-Fail

A subject's score on this criterion, which is dichotomous, is dependent, as the name implies, upon whether or not he passes or fails the flight training course. A student may fail for a number of reasons. The primary reasons are failure to pass the flight test at the end of the course, administered by a CAA inspector, and being "washed out" by the instructor, or flight operations check-pilot during the course, for general ineptness. However, students may also fail the flight course because of failure in ground school, for health reasons, because of proneness to airsickness, for disciplinary reasons, and others.[15]

Time Measures

These measures are based on operational records regarding the length of instructional time taken by given students to complete various "stages" of the flight course, or to complete the entire course. Three measures are available for most cases:

Time for Stage A — usually equivalent to instructional time before solo.

Time for Stage D — the instructional time necessary to complete the last stage or section of the course,[16] usually about the last 10 hours of instruction time.

---

[15]Reasons for failure in primary flight training and distribution of failers for each of these reasons are presented in a report prepared by the NRC Committee on Selection and Training of Aircraft Pilots. The CAA-National Testing Service. Summary of test results and comparisons with success in flight training. Washington, D. C.: CAA Division of Research, Report No. 39, November 1944, p. 38.

[16]The present primary course approved by the CAA consists of only two stages, A and B, although the total number of hours (35) in the curriculum has not been changed. At the time of this research, however, the course was divided into four stages, A, B, C, and D.

Total Time -- the total amount of instructional time necessary to complete the given course.

Purdue Rating Scale

This scale calls for ratings by the instructor on 14 items associated with Flight Proficiency.[17] The Purdue Rating Scale is presented in Figure 2. However, in most of the subsequent analyses, only the rating on Item 14 ("In your opinion, considering skill, emotional stability, judgment, etc., how good an all-around pilot is he likely to become?") was used as criterion datum. In general, two ratings on each subject were obtained one at the end of Stage A, and the other at the end of the course.

Photographic Criteria[18]

The photographic criteria are represented by measures yielded by photographic records of instruments indicating the attitude and performance of the plane, taken during a standard flight.[19] The type of instrument panel from which the photographic records of the instruments were made is presented in Figure 3. The measures are of two types, (1) over-all, or "global" ratings, and (2) "flight scores." The over-all or "global" ratings were made in terms of a three-point scale (A denoting good, B denoting average, and C denoting poor) and also in terms of a two-point scale, (U denoting upper half of the group in terms of flight proficiency, and L denoting lower half). These ratings were made first by two observers independently, the cases on which the two independent judgments did not agree being subsequently re-rated by the two observers working jointly. This joint re-rating of discrepant cases resulted in a composite rating. Then in order to obtain a measure of the reliability of the composite ratings, the entire procedure was repeated, resulting in a second composite rating. These were termed respectively "Composite Rating 1," and

---

[17] For a more detailed description of this scale see: Kelly, E. L. The development of "A Scale for Rating Pilot Competency," Washington, D. C.: CAA Division of Research, Report No. 18, July 1943.

[18] A detailed description of the photographic criteria is given in: Viteles, M. S., and Thompson, A. P. An analysis of photographic records of aircraft pilot performance. Washington, D. C.: CAA Division of Research, Report No. 3", July 1943.

[19] A "Standard Flight" is a period of specified maneuvers, which are executed in a specified sequence and under specified flight conditions. By this means a standard work sample is obtained from all subjects. The principles underlying the construction and use of standard flights are described in: Viteles, M. S., and Thompson, A. P. The use of standard flights and motion photography in the analysis of aircraft pilot performance. Washington, D. C.: CAA Division of Research, Report No. 15, May 1943. Standard flights used in this project are described in: Walker, R. Y., Lipman, E., and Wantman, M. J. Manual for the administration of the Ohio State Flight Inventory. Progress report, December 1941. (Copies of this manual are in the files of the War Committee on Selection and Training of Aircraft Pilots.)

Composite Rating II." In a similar manner the discrepant cases between Composites I and II were rated jointly, resulting in a "Criterion Rating."

The second type of measure taken from the photographic records were the "flight scores." After observing each film, the record readers independently rated the photographed performance in terms of eight aspects of flight performance. These eight aspects were:[20]

1. Wing Control
2. Nose Control
3. Directional Control
4. Slip-Skid
5. Altitude
6. Airspeed
7. Control Coordination
8. Execution of Maneuver

The aspects were rated in terms of the observers' judgment as to the deviation of the observed performance from the "ideal," using a five-point scale. On this scale "1" represented the high end of the scale and "5" the low end. The eight ratings of each of the two observers were summated, and the summated ratings of the observers were then added together to obtain the "Criterion Flight Score."

Ohio State Flight Inventory[21]

The Ohio State Flight Inventory is effectively a check-sheet, on which the flight examiner notes down the various critical elements in the pilot's performance during a standard check-flight. A sample page from the Ohio State Flight Inventory is presented in Figure 4. The various items in the check-sheet, for the several maneuvers incorporated in the standard flight, have authoritatively been given plus or minus weights. Items which are descriptive of "desirable" performance are given plus weights, items descriptive of "undesirable" performance are given minus weights.

On the basis of these item weights, a pilot's performance on any given maneuver can be assigned a "maneuver score," merely by summing

---

[20] For a detailed definition of these variables, see: Viteles, M.S. and Thompson, A.S. Op. cit. (Referred to in Footnote 18.)

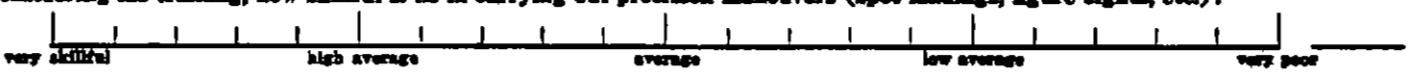
[21] For a detailed description of the Ohio State Flight Inventory, see: NRC Committee on Selection and Training of Aircraft Pilots. History and development of the Ohio State Flight Inventory, Part II: Recent versions and current applications. Washington, D. C.: CAA Division of Research, Report No. 51, November 1945.

Student Pilot's Name_____

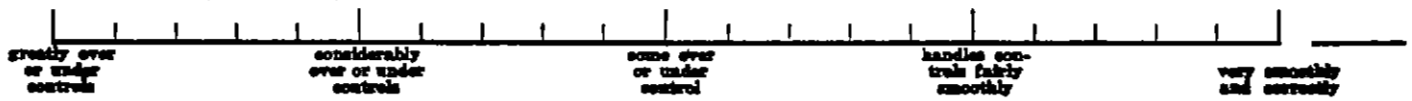
A SCALE FOR RATING PILOT COMPETENCY

No opportunity to observe

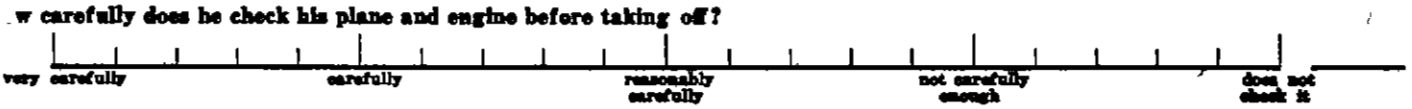
Considering his training, how skillful is he in carrying out precision maneuvers (spot landings, figure eights, etc.)?



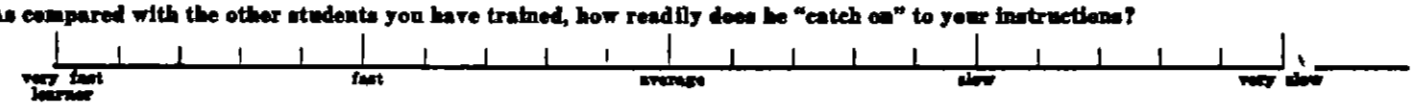
... does he handle the controls?



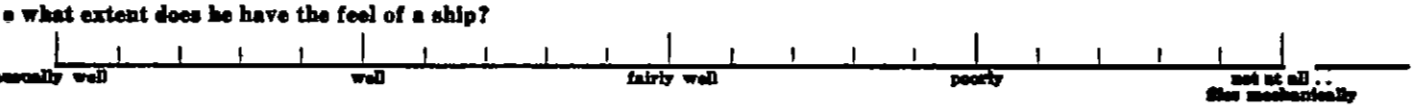
...w carefully does he check his plane and engine before taking off?



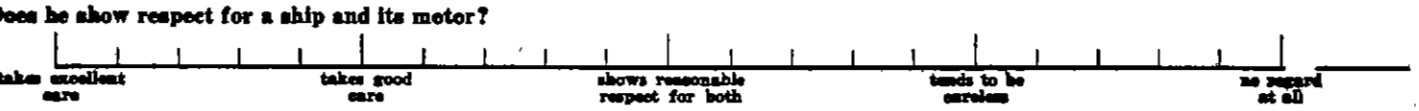
As compared with the other students you have trained, how readily does he "catch on" to your instructions?



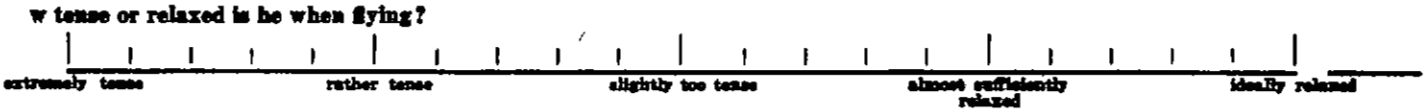
...e what extent does he have the feel of a ship?



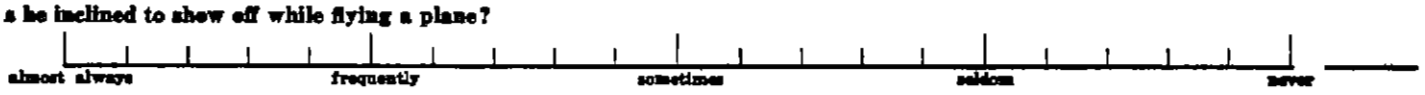
Does he show respect for a ship and its motor?



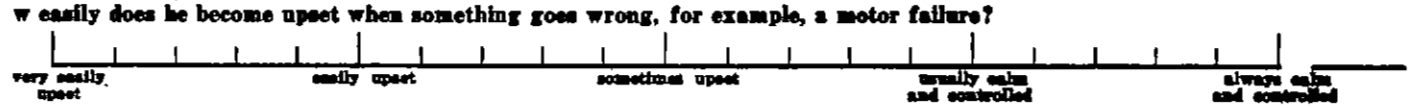
...w tense or relaxed is he when flying?



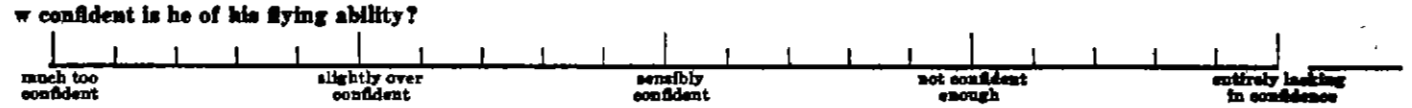
...s he inclined to show off while flying a plane?



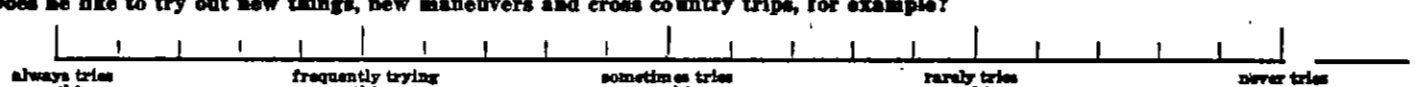
...w easily does he become upset when something goes wrong, for example, a motor failure?



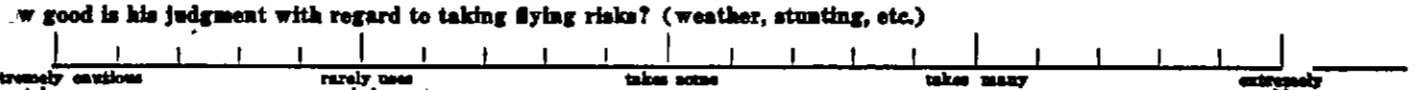
...w confident is he of his flying ability?



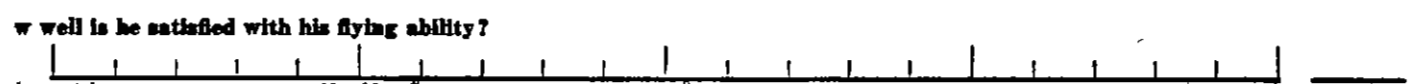
Does he like to try out new things, new maneuvers and cross country trips, for example?



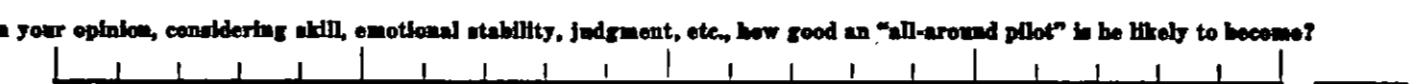
...w good is his judgment with regard to taking flying risks? (weather, stunting, etc.)



...w well is he satisfied with his flying ability?



In your opinion, considering skill, emotional stability, judgment, etc., how good an "all-around pilot" is he likely to become?



Rated by_____ Instructor

FTGI F

A SCALE FOR RATING PILOT COMPETENCY

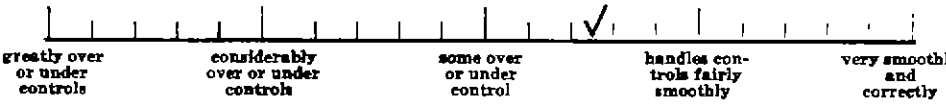
Directions for Use:

Being a good all-around pilot is not simply a matter of how well one can maneuver a ship, but also depends on the use of good judgment, on keeping one's head in emergencies, and on other traits which are difficult to measure. We have devised the scale on the back of this sheet especially to aid you in giving a many-sided description of a pilot and his flying habits.

The scale consists of 14 questions about the pilot being rated but a flat "Yes" or "No" answer is not asked for . . . rather you are given the opportunity of answering each question by checking at that point on the descriptive scale which best fits the pilot being rated. The descriptive phrases below each line should be thought of as "landmarks." Feel free to check any student as falling somewhere between these phrases if he belongs there.

EXAMPLE: If, for example, a pilot usually handles the controls fairly smoothly but sometimes slightly over or under controls, you might rate him as indicated by the check mark on the sample scale below.

How does he handle the controls?



It should be remembered that just because a pilot deserves to be rated high or low on one characteristic does not mean that he deserves to be rated equally high or low on others.

PURDUE RESEARCH FOUNDATION

CAMERA FIELD -- PIPER CUB INSTRUMENT PANEL

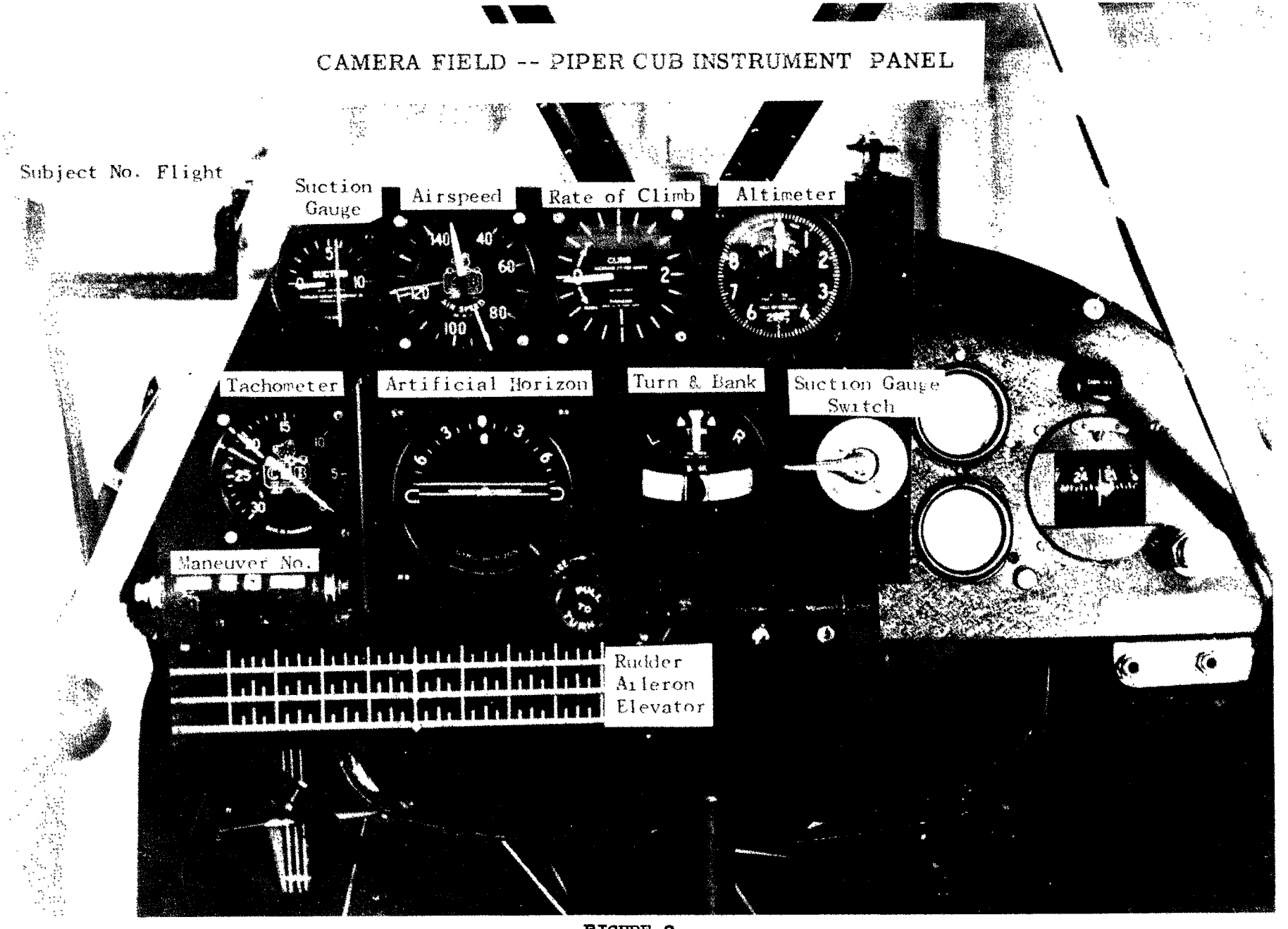


FIGURE 3

the weights on the items checked by the flight examiner as descriptive of that performance. In terms of the flight as a whole, two "over-all" scores are available, (a) Summation Score, derived by summating the individual maneuver scores and dividing by the number of maneuvers, and (b) Profile Score, derived on the basis of subjective evaluation of the "profile" of the several maneuver scores.[22]

The Flight Inventory was administered to all cases in the Midwest sample at the completion of the course, and in so far as possible also at the completion of Stages A, B, and C. In this report the criterion variables taken from the Flight Inventory are as follows:

1. Summation Score, Stage A

2. Summation Score, Stage C

Ground School Average Grade

The average Ground School Grade was determined from CAA records.

PROCEDURE

The selection tests were administered to the subjects before the beginning of their flight training, and as a condition of their admittance to flight training. A second administration of the instruments, i.e., a retest, was made for the purpose of determining their reliability, about three months after the original administration. It should be noted that all selection tests, and all criteria, could not be administered to all subjects. The subjects from the Midwest sample located at Tri-State College, and at Indiana State Teachers College did not take the Desire to Fly or the Personal History Inventories. Only the cases at Ohio State University, the University of Michigan, and Harvard University were interviewed and only the Ohio State subjects took the Link tests. Furthermore, photographic records are available on only those cases trained at Ohio State and on a part of those trained at the University of Michigan and at the University of Dayton.

DESIGN OF THE ANALYSIS AND LIMITATIONS OF THE RESEARCH

The analysis of the data from this project involved primarily correlational techniques, and can be subdivided into the following general categories, which will be treated in the results section of this report:

1. Reliability of predictors

2. Relationship between individual predictors and individual criteria

---

[22]The correlation between Summation Scores and Profile Scores was extremely high (.93). Since in this study Profile Scores are available on relatively few subjects they will not be considered further in this report.

STEEP TURNS

	RIGHT			LEFT		
CONTROL USE	Entry	Turn	Recovery	Entry	Turn	Recovery
Simultaneous.....	☐	☐	☐	☐	☐	☐
[illegible]	☐	☐	☐	☐	☐	☐
[illegible]	☐	☐	☐	☐	☐	☐
[illegible]	☐	☐	☐	☐	☐	☐
Neither.....	☐	☐	☐	☐	☐	☐
Rudder Pressure:						
Correct.....	☐	☐	☐	☐	☐	☐
Incorrect	☐	☐	☐	☐	☐	☐

PRECISION		RIGHT	LEFT
Bank.....	Constant	☐	☐
	Varies	____°	____°
Degree of Bank.....		____°	____°
Speed.....		____MPH	____MPH
Speed is.....	Constant	☐	☐
	Varies	____MPH	____MPH
Altitude is:...	Constant	☐	☐
	Varies	____ft.	____ft.
Recovers:..	On heading	☐	☐
	Off heading	____°	____°

FIGURE 4

3. Relationships between individual criteria and development of composite criteria

4. Correlations of individual predictors with composite criteria

5. Intercorrelations among predictors and development of prediction batteries

6. Prediction of composite criteria by batteries of predictors

The limitations of this research, it may be pointed out, are primarily those imposed by:

1. The relatively small number of cases available for certain comparisons

2. Differences in administrational procedures which rendered certain data from the Boston, and the Midwest sample, respectively, not comparable

3. Incomplete analysis of certain of the data which were collected

In view of these facts, many of the conclusions drawn from this research can be considered only tentative and suggestive.

RELIABILITY OF PREDICTORS

The reliability of prediction tests in this study was determined, in general, in terms of test-retest correlations. While the number of subjects in individual samples is not extremely large, the reliabilities of the prediction instruments can be evaluated in terms of the coefficients obtained on a number of independent samples.

Physiological Measures

Data on the reliability of physiological measures are available from the Boston samples only.[23] The test-retest correlations on these measures are given in Table 1. Examination of this table indicates that only three measures, Body Surface, Vital Capacity, and Vital Capacity/Body Surface, approach what might be considered a satisfactory degree of reliability. The reliability coefficients for Body Surface on the Fall and Spring samples, respectively, are .94 and .77. For Vital Capacity the coefficients on the two respective samples are .81 and .85, and for Vital Capacity/Body Surface, .83 and .76. Tidal Air, Pulse Rate, and Systolic Blood Pressure are next in order in terms of reliability, with coefficients ranging between .50 and .70. The reliability of the remaining physiological measures is considerably

---

[23] Reliability coefficients for physiological measures were not computed for the Midwest data because retests on these variables were not administered to the Midwest [illegible].

lower vital figures from both samples were taken into consideration.[24]

TABLE 1

PHYSIOLOGICAL INDICES (CIRCULATORY AND SPIROMETRIC)

BOSTON PROJECT

Test-Retest Correlations

Variable	Fall Group r_{tt}	Fall Group N	Spring Group r_{tt}	Spring Group N
Body Surface	.94	97	.77	81
Vital Capacity	.81	97	.85	81
Tidal Air	.64	97	.69	81
Tidal Air/Body Surface	.60	97	.68	81
Vital Capacity/Body Surface	.83	97	.76	81
Pulse Rate (lying)	.65	92	.65	81
Systolic B.P. (lying)	.55	92	.52	81
Diastolic B.P. (lying)	.38	92	.32	81
Smallest Pulse Pressure	.39	92	.30	81
Pulse Pressure Time	.03	92	-.07	81
Pulse Pressure Max. Change	.11	92	.15	81
Systolic B.P. Init. Change (lying)	.21	92	.18	80
Systolic B.P. Max. Change	.34	92	.35	80
Systolic B.P. Time to Max.	-.02	92	.07	80
Diastolic B.P. Init. Change (up)	.17	92	.36	80
Diastolic B.P. Max. Change	.19	92	.30	80
Diastolic B.P. Time to Max.	.26	92	.11	80
Pulse Rate Initial Change	.47	92	.29	80
Pulse Rate Maximum Change	.44	92	.37	80
Pulse Rate Time to Max.	.003	92	.23	80
Pulse Rate Max. Rate	.55	92	.48	80

The means and standard deviations for the various measures on test and retest are given in Table 2. It will be noted that there is little difference between means and between standard deviations from test to retest, or from sample to sample.

Psychomotor Measures

Data on the reliability of the psychomotor tests are available, in general, for the Boston Spring and Fall samples separately, and for the Midwest Spring and Summer samples combined. However, no data on the Eye-Hand Coordination Test are available for the Boston Spring sample,

---

[24] For a discussion of the reliability of respiratory measures obtained on naval cadets in the Pensacola research project, see: Franzen, R., and Blaine, Louise. Evaluation of respiratory measures for use in pilot selection. Washington, D. C.: CAA Division of Research, Report No. 25, January 1944.

TABLE 2

PHYSIOLOGICAL INDICES (CIRCULATORY AND SPIROMETRIC)

BOSTON PROJECT

Means and Standard Deviations of Measures on Test and Retest

Variable	Fall Group				Spring Group			
	Original Test		Retest		Original Test		Retest	
	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Body Surface	1.87	0.11	1.88	0.10	1.86	0.14	1.87	0.10
Vital Capacity	5011.80	632.60	5010.30	647.60	4904.07	547.80	4870.49	520.03
Vital Air	699.30	182.70	685.30	173.40	671.85	183.40	640.37	132.00
Vital Air/Body Surface	375.10	97.70	363.80	89.60	360.00	92.40	345.93	78.50
Vital Capacity/Body Surface	2698.40	320.20	2659.70	318.70	2626.30	256.70	2609.01	222.30
Pulse Rate (lying)	69.01	9.38	70.37	9.02	69.65	10.03	66.88	8.24
Systolic B.P. (lying)	120.91	10.31	117.85	10.74	116.84	11.32	109.75	9.31
Diastolic B.P. (lying)	72.25	7.99	60.84	9.55	67.65	12.03	63.06	7.49
Systolic Pulse Pressure	18.96	8.53	29.82	12.77	25.35	10.32	21.98	7.75
Pulse Pressure Time	9.36	6.58	8.10	6.68	9.64	6.94	10.22	6.16
Pulse Pressure Max. Change	29.89	9.81	27.41	8.30	23.75	9.47	24.80	9.42
Systolic B.P. Init. Change (lying)	3.66	8.76	3.94	7.09	3.30	8.45	3.75	9.06
Systolic B.P. Max. Change	15.36	7.74	10.16	6.82	9.81	7.95	15.10	10.58
Systolic B.P. Time to Max.	9.22	6.53	6.76	6.13	7.88	6.15	7.86	6.15
Diastolic B.P. Init. Change (up)	11.00	7.35	15.08	6.77	10.75	8.59	6.10	7.28
Diastolic B.P. Max. Change	18.65	6.72	21.60	7.03	17.76	9.19	12.75	7.24
Diastolic B.P. Time to Max.	8.30	6.45	8.77	6.75	9.19	6.72	9.60	6.14
Pulse Rate Initial Change	20.33	9.16	20.64	10.64	22.00	8.54	19.48	7.99
Pulse Rate Max. Change	25.77	8.88	27.47	9.50	32.98	9.50	32.78	8.44
Pulse Rate Time to Max.	6.14	5.46	6.42	5.66	8.56	5.57	9.14	6.39
Pulse Rate Max. Rate	94.67	11.55	97.88	12.17	101.91	11.48	98.98	10.58

and some on reliability of certain of the derived data. The various tests are, in a few instances, available from only one of the samples. For example, the reliability of the score indicating learning between trials 1 and 6 on the Two-Hand Coordination Test" is available only for the Midwest group.

The test-retest correlations for the psychomotor measures are summarized in Table 3.

Two-Hand Coordination Test

The mean score from 6 trials is the most reliable measure taken from this test, the test-retest correlations being .75 and .80 for the Boston Fall group, and the Midwest combined groups, respectively. The coefficient for the Boston Spring sample drops to .50. The variable "highest score" shows relatively high reliability (.63) for the Boston Fall sample, but this measure was not computed for other groups. The measures involving "learning" and trial number of highest or lowest score have approximately zero reliability on the basis of coefficients computed from single samples.[25]

It will be noted that the reliability coefficients for the various Two-Hand measures on the Boston Spring group are consistently and markedly lower than are the comparable coefficients for the Boston Fall group, or from the Midwest sample. Examination of Table 4, in which the means and standard deviations of the psychomotor measures on test and retest are given, indicates that the mean scores on retest are higher than the mean score on the original test. However, it may be significant that the Boston Spring sample which showed the lowest reliability coefficients on the Two-Hand measures, also showed the least increase in mean score from test to retest. This might be accounted for by activities engaged in by members of the group during this interval, by some irregularity in the administration of the test to this group, or by some fundamental difference in the samples. This latter possibility is militated against by the fact that mean scores on the original test and the standard deviations on test and retest are not markedly out of line when the Boston Spring sample is compared with the Boston Fall and Midwest samples.

The Mashburn Serial Reaction Test

Turning again to Table 3, it is evident that the most reliable measures taken from the Mashburn test are the scores based on settings 39 and 40 (i.e., total time required by the subject to reach these settings). The coefficients for these measures on the Boston Spring group and on the Midwest Spring and Summer groups combined are between .74

---

[25]Coefficients of reliability of the test as a whole, based on odd versus even trials, ranged from .79 to .88 in a study described in: Spence, K. W., Buxton, C. E., and Melton, A. W. The effects of massing and distribution of practice on two-hand coordination test scores. Washington, D. C.: CAA Division of Research, Report No. 45, April 1945.

TABLE 3

RELIABILITY OF PSYCHOMOTOR TESTS

	Boston				Midwest	
	Fall Group		Spring Group		Sp. & Sum. Gr. Combined	
Two-Hand Coordination Test	r_{tt}	N	r_{tt}	N	r_{tt}	N
Trial 1	.43	98	.24	88	.51	185
Trial 2	.47	98	.35	88	.59	185
Trial 3	.43	98	.33	88	.65	185
Trial 4	.55	98	.28	88	.62	185
Trial 5	.58	98	.39	88	.55	185
Trial 6	.53	98	.37	88	.61	185
Mean of 6 trials	.75	98	.50	88	.80	185
Learning, Trials 1 to 6					.03	185
Lowest score	.54	98				
Highest score	.63	98				
Trial number of low score	.04	98				
Trial number highest score	04	98				
Mashburn Serial Reaction Test						
13 setting	.39	98	.60	88	.61	185
26 setting	.52	98	.69	88	.68	185
39 setting	.51	98	.75	88	.76	185
40 setting	.53	98	.74	88	.74	185
26 minus 13 setting	.53	98	.45	80	.57	185
39 minus 26 setting	.53	98	.57	80	.68	185
40 minus 39 setting	.19	98	.08	80	.15	185
Time 13th Trial minus Time 39th Trial					.03	185
Eye-Hand Coordination Test						
Pattern A, Trial 1			.62	89	.53	185
Trial 2			.66	89	.58	185
Trial 3			.68	89	.50	185
Mean of 3 trials			.75	89	.65	185
Learning tr. 1 to 3					.10	185
Pattern B, Trial 1			.56	89	.49	185
Trial 2			.58	89	.64	185
Trial 3			.51	89	.56	185
Mean of 3 trials			.71	89	.63	185
Learning tr. 1 to 3					.15	185
Pattern D, Trial 1			.30	89	.25	185
Trial 2			.55	89	.36	185
Trial 3			.71	89	.34	185
Trial 4			.60	89	.44	185
Trial 5			.62	89	.34	185
Trial 6			.66	89	.46	185
Trial 7			.64	89	.37	185
Trial 8			.62	89	.37	185
Mean of 8 trials			.71	89	.44	185
Learning tr. 1 to 8					.15	185

TABLE 4

PSYCHOMOTOR TESTS

Means and Standard Deviations of Measures on Test and Retest

	Boston Fall Group				Boston Spring Group				Midwest Spring & Summer Groups			
	Test		Retest		Test		Retest		Test		Retest	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Two-Hand Coordination Test												
Trial 1	41.72	13.4	61.66	13.9	44.89	16.32	51.95	15.21	42.57	15.25	61.57	14.24
Trial 2	49.81	14.6	68.81	12.6	53.98	12.08	60.09	12.58	53.61	14.90	67.57	13.33
Trial 3	55.80	13.4	73.09	13.0	57.11	14.96	61.51	14.32	56.53	15.50	69.94	13.90
Trial 4	57.15	13.5	75.13	13.7	58.56	15.25	64.67	11.47	58.88	15.46	71.69	12.91
Trial 5	58.43	13.8	74.34	12.5	59.25	14.50	65.51	12.21	61.09	13.74	72.08	14.52
Trial 6	60.85	14.3	75.24	11.9	60.84	15.57	66.68	12.23	61.65	14.27	71.66	11.75
Mean of 6 trials	54.03	10.9	71.49	11.0	55.78	11.87	61.83	10.20	55.67	12.39	68.87	11.96
Learning trials 1 to 6									18.73	14.16	10.74	12.89
Lowest score	37.31	12.0	58.65	13.3								
Highest score	66.92	11.7	81.43	10.4								
Trial number of low score	1.67	1.3	1.92	1.4								
Trial number highest score	4.54	1.4	4.27	1.3								
Washburn Serial Reaction Test												
13 setting	1.81	.44	1.35	.25	1.78	.39	1.35	.28	1.72	.34	1.32	.23
26 setting	3.52	.79	2.64	.42	3.44	.67	2.64	.50	3.30	.61	2.62	.45
39 setting	5.13	1.05	3.92	.64	5.00	.90	3.90	.67	4.82	.86	3.86	.66
40 setting	5.27	1.06	4.02	.64	5.12	.91	4.00	.69	4.97	.88	3.98	.67
26 minus 13 setting	1.72	.40	1.31	.21	1.64	.38	1.28	.22	1.58	.31	1.30	.26
39 minus 26 setting	1.62	.32	1.28	.24	1.57	.29	1.29	.22	1.52	.30	1.25	.25
40 minus 39 setting	.14	.06	.09	.04	.12	.04	.10	.04	.15	.06	.12	.05
Time 13th Trial minus Time 39th Trial									.20	.26	.09	.03

TABLE 4 (continued)

PSYCHOMOTOR TESTS

Means and Standard Deviations of Measures on Test and Retest

	Boston Fall Group				Boston Spring Group				Midwest Spring & Summer Groups			
	Test		Retest		Test		Retest		Test		Retest	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Eye-Hand Coordination Test												
Pattern A, Trial 1					83.75	13.19	74.43	13.36	67.78	15.79	49.02	12.95
Trial 2					72.02	13.32	63.17	10.36	55.22	15.80	38.69	11.76
Trial 3					69.63	13.33	61.84	10.22	51.91	15.33	35.73	10.87
Mean of 3 trials					75.07	12.29	65.64	10.79	58.33	14.00	41.02	10.72
Learning, trials 1-3									16.10	15.32	13.95	9.51
Pattern B, Trial 1					65.56	14.01	54.45	11.76	43.52	10.31	33.38	9.16
Trial 2					57.85	13.09	50.37	10.34	38.03	12.33	27.59	9.42
Trial 3					55.78	12.11	49.56	13.55	36.62	11.41	28.95	8.95
Mean of 3 trials					59.62	11.67	51.48	10.73	39.02	10.34	28.51	8.56
Learning, trials 1-3									7.76	7.31	5.15	5.49
Pattern D, Trial 1					96.39	13.87	65.38	11.96	65.76	12.05	34.39	8.81
Trial 2					75.24	12.51	53.51	8.54	47.79	11.39	25.10	8.76
Trial 3					67.36	10.30	51.48	8.08	40.39	10.12	23.51	7.56
Trial 4					64.46	8.95	51.25	7.71	37.31	10.62	22.82	6.74
Trial 5					62.10	9.98	47.83	6.86	38.63	9.65	22.05	6.43
Trial 6					60.48	9.03	49.20	7.18	35.03	9.29	21.55	7.28
Trial 7					58.51	9.34	48.42	7.18	33.51	8.82	20.86	6.56
Trial 8					57.45	8.86	47.98	5.94	32.00	8.65	20.31	6.30
Mean of 8 trials					67.89	8.91	52.79	7.85	40.67	8.83	23.57	6.66
Learning, trials 1-3									33.90	10.50	14.77	6.31

and .70. For the Boston Fall group the coefficients are lower, being .51 and .53 for the 39 and 40 settings, respectively. There is a general increase in reliability from the 13 setting to the 40 setting which is probably accounted for by the progressively increased length of the test between these settings. The score derived from the difference between settings (e.g., 26 minus 13 setting) are of approximately the same reliability as the more reliable of the two measures. "Learning score" (i.e., learning between settings 13 and 39) was determined by subtracting the time required to complete the third presentation of the series of 13 patterns (settings 27 to 39) from the time required to complete the first presentation of this standard series (settings 1 to 13). This learning score shows extremely low reliability on the basis of the Midwest data.[26] The Boston Fall group yields in general, lower reliability coefficients than do the other samples.

The means and standard deviations for the Mashburn scores on test and retest are given in Table 4. It should be noted that the scores on retest are slightly, but consistently lower than are the scores on the original test, and that the standard deviations are markedly lower on retest than on the original test. The Boston Fall group, which yielded generally lower reliability coefficients, shows consistently larger standard deviations in terms of Mashburn scores on the original test than do the other samples.

---

[26]Further information on the reliability of the Mashburn is given in: Nance, R. D., Buxton, C. E., and Spence, K. W. The effect of distraction lights upon performance on the Mashburn Serial Coordination Test. Washington, D. C.: CAA Division of Research, Report No. 29, April 1944. Using scores expressed in terms of the number of matchings made on the Mashburn apparatus during certain periods of work, these authors report split-half reliability coefficients of .97 and .96 for the "long form" of the test (24 minutes work) under conditions of distraction and no-distraction, respectively; and split-half reliability coefficients of .83 and .82 for the "short form" of the test (8 minutes work) under conditions of distraction and no-distraction, respectively. The "long form" and the "short form" coefficients were determined, respectively, on samples of 50 college men. In comparing the reliability coefficients obtained by Nance, Buxton, and Spence with those reported in this study it should be noted that their scoring procedures (in which the number of matchings during a given period of work was determined) were not comparable to the scoring procedures used in the Midwest study (expressed in terms of the time taken to complete the fortieth setting on the Mashburn). Furthermore, split-half reliability coefficients can be expected to be higher than test-retest coefficients. However, it should be noted that in the Midwest project, the mean duration of the work sample on the Mashburn varied from approximately 4 to 5 minutes for the various groups (see Table 4), a considerably shorter work sample than that used by Nance, Buxton, and Spence. The high reliability coefficients reported by these authors for their "short test" and particularly for their "long test" would seem to argue for extended work samples on this instrument if reliability is the prime consideration.

Eye-Hand Coordination Test

The reliability coefficients of the various scores derived from the Eye-Hand Coordination Test are summarized in Table 3. It is to be noted that no data are available from the Boston Fall group, since this test was not administered to this sample. Inspection of this table indicates that when data from both samples are considered the scores based on the mean of a number of trials are the most reliable. There seems to be little to choose between Patterns A, B, and D as far as reliability is concerned, although for the Midwest sample the reliability coefficient for mean score on Pattern D is considerably lower than are the coefficients for mean score on Pattern A or on Pattern B.

Inspection of the means and standard deviations of the Eye-Hand Coordination Test scores on test and retest, as given in Table 4, indicates that the scores on retest are in general lower than on the original test. In general, this difference is greater for the Midwest data than for the Boston data. The reliability coefficients, however, are greater on the Boston sample than on the Midwest sample.

Link Trainer

The Link Trainer (Contact) as a predictor was used on the Midwest sample only. In general, scores on the Link Trainer were available in terms of two categories, i.e., "Turns" and "Banks." Under each of these categories three scores were available, i.e., for "Aileron," "Rudder," and "Elevator," respectively. Measures on these variables were taken from a series of clocks, and were expressed, for the "Turn" category, in terms of duration of (a) deviation in aileron control from a predetermined "ideal" performance, (b) deviation in elevator control from longitudinal balance, and (c) deviation in rudder control from the ideal. Measures for the "Banking" category on "Aileron," "Rudder," and "Elevator" were similar except that the measure on "Rudder" was expressed in terms of duration of deviation from heading. For the maneuver Straight and Level scores were obtained in terms of only two measures, namely, duration of deviations from (a) correct balance and (b) correct heading.[27]

Since the Link test was used only once (prior to flight training) it was impossible to obtain test and retest measures in determining reliability. Therefore, recourse was taken to the split-half technique. However, the traditional odd-even split could not be made, since for turn maneuvers all odd trials were to the right, and even trials were to the left, and marked differences in performances between "Turn" scores to the right and to the left were evident.

Therefore, two other measures of reliability were obtained: (a) the correlation of score on the first four trials with score

---

[27] The definitions of the scores are described in a progress report in the files of the [illegible] Selection and Training of Aircraft Pilots, prepared by [illegible], entitled Analysis of the Link predictor test. [illegible]

on the second four trials, (b) the correlation of "extreme trials" with "mean trials," i.e., trials 1, 2, 7, 8 against trials 3, 4, 5, 6. The correlations between scores on the first half of the test versus the second half for Spring and Summer groups are presented in Table 5. The correlations between scores for the extreme trials versus the mean trials, for Spring and Summer groups, are shown in Table 6.

TABLE 5

LINK TRAINER RELIABILITY
Spring Group (N = 49)
1st half versus 2nd half

	Means*		Sigmas*		
	First Half	Second Half	First Half	Second Half	r
**Aileron	18.24	19.53	15.42	16.35	.69
Turns Elevator	250.00	255.65	40.30	60.03	.00
**Rudder	9.84	11.55	5.48	5.67	.46
**Aileron	3.02	2.55	1.85	1.71	.24
Banks Elevator	68.29	71.49	11.58	12.17	.21
Rudder	82.55	87.98	13.27	12.22	-.36

Summer Group (N = 48)

	Means*		Sigmas*		
	First Half	Second Half	First Half	Second Half	r
**Aileron	75.69	73.75	16.55	19.96	.55
Turns Elevator	293.33	297.77	43.35	38.05	.09
**Rudder	5.52	7.60	3.94	5.82	-.12
**Aileron	2.32	2.06	1.90	1.61	.26
Banks Elevator	83.58	85.31	15.11	15.05	.38
Rudder	91.40	89.90	8.15	12.53	-.12

*Means and sigmas are in seconds.
**Statistics represent amount of deviation from a predetermined ideal performance.

Examination of Table 5 indicates that the "Aileron" score for the "Turn" measures yields the highest reliability coefficients in terms of either method of estimating reliability. For the Spring and Summer samples, respectively, the correlation coefficients between first half and second half scores are .69 and .55. Between mean trials and extreme trials, the correlation coefficients are .76 and .61 for Spring and Summer

samples, respectively. All of the other coefficients ("Rudder" and "Elevator" scores) are low. In general, the reliability coefficients for the various measures taken from the Link test are of such low magnitude that, on the basis of these data, the use of the Link Trainer in a battery for the prediction of flight proficiency seemed inadvisable.[28]

TABLE 6

LINK TRAINER RELIABILITY
Spring Group (N = 49)

1st half versus 2nd half

		Means*		Sigmas*		
		Mean	Extreme	Mean	Extreme	r
Turns	**Aileron	18.27	19.49	14.15	15.53	.76
	Elevator	244.71	260.92	36.90	65.03	-.08
	**Rudder	9.98	10.88	6.21	5.56	.37
Banks	**Aileron	2.99	2.73	2.32	1.82	-.07
	Elevator	70.55	14.89	69.43	12.78	-.07
	Rudder	83.82	86.71	10.67	9.53	.02

Summer Group (N = 43)

		Means*		Sigmas*		
		Mean	Extreme	Mean	Extreme	r
Turns	**Aileron	68.42	67.00	19.23	16.62	.61
	Elevator	294.85	297.65	34.86	38.86	.26
	**Rudder	5.75	7.13	4.58	4.79	.02
Banks	**Aileron	2.02	2.20	1.62	1.76	.42
	Elevator	87.04	82.35	13.99	16.90	.31
	Rudder	90.69	90.53	11.67	9.58	-.11

*Means and sigmas are in seconds

**Statistics represent amount of deviation from a predetermined ideal performance

[28]It seems possible that the reliability coefficients presented above may represent the reliability of the particular scoring method used, rather than the reliability of an individual's performance on the Link Trainer. Other methods of determining Link proficiency, e.g., the amount of practice required by a subject to reach a specified standard of proficiency on the Link Trainer, would higher coefficients. Such alternative methods of scoring have not been investigated, however.

Paper-and-Pencil Tests

Of the paper-and-pencil tests, experimental data on the Biographical Inventory alone are available, and are available only on the Boston samples, with the exception of a supplementary research which was carried out at the University of Rochester.

Biographical Inventory

The reliability coefficients for the various scores derived from the Biographical Inventory are presented, for Boston Fall and Boston Spring groups, respectively, in Table 7. In Table 8 the means and standard deviations of these scores on test and retest are summarized.

TABLE 7

BIOGRAPHICAL INVENTORY

BOSTON PROJECT

Test-Retest Correlations

	Fall Group		Spring Group	
	r_{tt}	N	r_{tt}	N
+ 1% Items, Part A	.60	86	.63	73
± 1% Items, Part A	.57	86	.68	73
+ 1% Items, Part B	.66	86	.64	73
± 1% Items, Part B	.64	86	.62	73
+ 1% Items, Parts A + B	.54	86	.66	73
± 1% Items, Parts A + B	.53	86	.65	73

TABLE 8

BIOGRAPHICAL INVENTORY

BOSTON PROJECT

Means and Standard Deviations on Test and Retest

	Fall Group				Spring Group			
	Test		Retest		Test		Retest	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
+ 1% Items, Part A	9.76	2.35	10.98	2.68	9.16	3.03	11.03	2.84
± 1% Items, Part A	4.24	3.97	5.66	4.04	2.96	5.07	5.89	4.53
+ 1% Items, Part B	4.24	1.08	4.49	1.03	4.00	1.21	4.14	1.16
± 1% Items, Part B	2.56	2.04	2.99	2.03	1.99	2.42	2.36	2.35
+ 1% Items, Parts A + B	14.03	2.37	15.47	2.60	13.16	3.34	15.08	3.16
± 1% Items, Parts A + B	6.80	4.13	8.63	4.10	3.96	5.91	7.62	5.45

It is to be noted that the reliability coefficients presented in Table 7 are not high. Further, there is in general a slight, but in some cases a marked increase in mean score between test and retest. Both of these facts might be explained by the relatively long interval which elapsed between test and retest during which time most of the subjects underwent flight training. Activities occurring in this interval, particularly flight training, conceivably could so alter a subject's interests and his replies to personal history questions that the reliability of the test, as determined by test-retest would be decreased.

In view of this fact it seemed desirable to obtain another estimate of the reliability of this instrument on the basis of test-retest scores under conditions in which a shorter time elapsed between test and retest. Such an investigation was undertaken at the University of Rochester, under the direction of M. J. Wantman, in connection with the testing of applicants for the National Testing Service program.[29]

The Biographical Inventory was administered to a group of 23 applicants at the first, and again at the last of a testing period in which a number of tests were administered. (A plausible excuse was given for the re-administration of the test so that the motivation on the two administrations was probably the same.) The time elapsing between test and retest was four hours.

The reliability coefficient, in terms of test-retest, for this group was .94, the total scores being computed on the basis of "A + 1%" items.[30] Even though the four hour time interval may have been so short that the applicants recalled some of their previous responses, it may be concluded from this small study that the Biographical Inventory is more reliable than the figures obtained from the Boston samples would imply.[31, 32]

---

[29]Test-retest correlation coefficients of .525 and .603 were obtained on two samples of 307 cases and 1334 cases respectively in this study. See Table 23, p. 35. Op. cit. (Referred to in Footnote 15.)

[30]"A + 1%" items are those items in Part A of the Inventory which carry positive unit weights, and predicted the criterion groups in the validation study at the 1% level of significance.

[31]In an early study at Purdue University, conducted by E. L. Kelly in connection with the development of the Biographical Inventory, the split-half reliabilities for two scoring keys of the instrument, in terms of preliminary weights were reported as .84 and .81. See: Kelly, E. L. The relationship of background and personality factors to pilot competency. (Progress report on file with the NRC Committee on Selection and Training of Aircraft Pilots.)

[32]In connection with this study, under the direction of M. J. Wantman, it was found that responses to certain weighted items in the instrument were changed from test to retest with greater frequency than were responses to certain other weighted items. This would suggest that certain items in the [illegible] are in need of revision.

Evaluation Interview

As reported by Dunlap and Wantman,[33] the reliabilities of the Interview scales, as estimated by the Spearman-Brown prophecy formula applied to the mean intercorrelation between raters, are as follows:

Interview Scales

	A	B	C	D	E	F	G	H	I
Rel. Coeff.	.87	.80	.84	.88	.81	.82	.81	.77	.81

It will be noted that the reliabilities for individual scales range from .77 to .88. The authors state that "While these reliabilities are not as high as one would like for making accurate predictions on individuals, they are sufficiently high to give dependable predictions for groups of individuals."

Although from this project no experimental data are available bearing upon the reliabilities of the Otis Self Administering Test of Mental Ability, the Strong Vocational Interest Blank, or the Test of Mechanical Comprehension, the reliabilities as published by the authors of these instruments are as follows: The Otis Self Administering Test of Mental Ability (Form D), .92; the Strong Vocational Interest Blank for men (Form M), .80; and the Bennett Test of Mechanical Comprehension (Form B), .80.

Summary

From examination of the experimental data presented in this section it is evident that on the basis of the two independent samples for which data were available the three physiological variables, Body Surface, Vital Capacity, and Vital Capacity/Body Surface yielded the highest reliabilities of the several predictor variables.

The reliabilities of the psychomotor measures were slightly lower, although scores from the Two-Hand Coordination Test and the Mashburn Serial Reaction Time Test yielded reliability coefficients of .74 or greater on two out of the three independent samples on which data were available. Among the psychomotor tests those scores based on the work sample of maximum length, or on mean measures, were the most reliable, as might be expected.

Of the paper-and-pencil tests only the reliability of the Biographical Inventory was determined on the basis of samples involved in this investigation. No unequivocal conclusions on the basis of data presented here can be drawn since the coefficients vary from .60 on a sample in which test and retest were separated by an interval of three months to .94 on a sample in which test and retest were separated by an interval

---

[33] Dunlap, J. W., and Wantman, M. J. Op. cit. (Referred to in Footnote 10.)

of four hours. While certain items may have been recalled after the shorter interval thus rendering the coefficient unduly high, the fact that during the longer interval of three months the subjects engaged in flight training may so have altered their interests and personal histories as to render the coefficients in the neighborhood of .60 too low.

Since the reliabilities for the Otis, Strong, and Test of Mechanical Comprehension are not taken from samples involved in this experiment, they can hardly be compared directly with the other reliability coefficients presented in this report.

RELATIONSHIPS BETWEEN INDIVIDUAL PREDICTORS AND INDIVIDUAL CRITERIA

Determination of the relationships between the various predictors and the several criteria are important in determining not only the best predictors, but also in determining the best predictable criteria. In this section, the correlations between predictors and criteria will be examined. In general, in each group of predictors (physiological measures, psychomotor measures, etc.) only those measures which are relatively reliable will be considered.

Correlations with Pass-Fail, Time for Stage A, and Total Time

In Table 9 are presented the correlation coefficients between the more reliable predictors and the above criterion measures. Data are available from the Boston Fall and Spring groups, and from the Midwest Spring and Summer groups. Examination of the correlations with the criterion Pass-Fail indicates that none of the predictor variables consistently predict this criterion in terms of all four samples. For example, the correlations involving the Washburn Test score on trials 27 to 39 vary from -.37 to .41.[34] Similarly, some of the correlations between paper-and-pencil test scores and Pass-Fail vary from sizable positive to sizable negative values. It might be noted, however, that the correlations between Tidal air and this criterion are greater than -.30 for three of the four samples. The correlations with Pass-Fail are all biserial coefficients, and the number of cases in the "failing" category is in general small. Therefore, these correlation coefficients are undoubtedly unstable.

Examination of the correlations between the test variables and Time for Stage A indicates that none of these variables predict this criterion over all four samples. The correlations between physiological measures and Time for Stage A are all low. The Two-Hand Coordination Test scores, and the Washburn Test scores show some degree of prediction on the Midwest Spring and Summer samples, the correlations being around the magnitude of .30, but show little or no degree of prediction in terms of the Boston Fall and Spring samples. The Two-Hand Coordination Test is not correlated with this criterion to any degree. The correlations between

---

[34] A time score is used on the Washburn Test, thus a high score denotes poor performance.

TABLE 9

PREDICTOR-CRITERION CORRELATIONS (SELECTED VARIABLES)

BOSTON AND MIDWEST GROUPS

Variable	PASS-FAIL*				TIME STAGE A				TOTAL TIME			
	Boston Fall grp	Boston Spr. grp	Midwest Spr. grp	Midwest Sum. grp	Boston Fall grp	Boston Spr. grp	Midwest Spr. grp	Midwest Sum. grp	Boston Fall grp	Boston Spr. grp	Midwest Spr. grp	Midwest Sum. grp
Physiological Test (Phys.)												
Body surface	.10	.02	-.24	.13	.12	-.06	-.07	-.03	.00	-.03	.03	-.02
Vital capacity	.17	-.12	-.23	-.27	-.11	-.17	-.03	.00	-.06	-.17	-.16	-.36
Tidal air	.04	-.58	-.48	-.03	.15	.07	-.09	.09	.05	.04	.04	.16
Tidal air/Body surface	.03	-.61	-.42	-.33	.14	.09	-.08	.09	.06	.05	.04	.17
Vital cap./Body surface	.13	-.17	-.09	-.34	-.19	-.08	.01	.02	-.07	-.09	-.21	-.41
Pulse rate (lying)	-.08	-.09	-.20	-.25	-.12	.05	-.06	.02	.11	.14	.20	.27
Psychomotor Tests (P-M)												
Two-Hand: trial 6			-.20	-.04			-.29	-.25			-.16	-.30
Two-Hand: Mean	-.02	-.07	-.18	-.17	-.20	.01	-.33	-.31	-.23	.01	-.02	-.16
Mashburn: through trial 39	.18	-.22	.45	.20	.21	.12	.36	.39	.50	.09	.15	.14
Mashburn: trials 14 to 26	.26	-.02	.49	.25	.10	.15	.32	.33	.47	.10	.13	.04
Mashburn: trials 27 to 39	.15	-.37	.41	.16	.17	.19	.30	.34	.41	-.02	.16	.16
Eye-Hand: pattern A		.15	.40	.10		.02	.20	.23		-.01	.01	-.06
Eye-Hand: pattern B		-.26	.25	.12		.12	.23	.19		-.02	.10	.11
Eye-Hand: pattern D		-.15	.29	.07		.05	.13	.19		-.04	.05	.12
Paper-and-Pencil Tests (P and P)												
BI ± 1% A	.23	.37			-.13	-.28			-.21	-.32		
BI + 1% A	.25	.31	-.15	-.23	-.15	-.25	-.09	-.20	-.20	-.20	.07	-.07
BI + 1% B			-.02	-.07			.00	-.08			-.14	-.03
BI ± 1% A + B	.22	.25			-.14	-.28			-.18	-.23		
BI + 1% A + B	.25	.27	-.15	-.27	-.14	-.27	-.08	-.24	-.20	-.21	.01	-.07
Mechanical Comprehension	-.03	.32	-.31	-.35	-.34	-.27	-.12	-.15	-.37	-.27	-.24	-.17
Otis	.17	.13	-.07	-.14	-.17	-.01	-.04	-.06	-.10	-.08	-.04	-.07
Aviation Information				-.24				-.01				-.14
Number of Cases (N)	98-105	82-90	Phys.:50 Others:139-140	P-M:96 Others:40-47	83-84	72-80	Phys.:49 Others:137-138	P-M:96 Others:40-47	83-84	72-80	Phys.:46 Others:121-122	P-M:90 Phys.:45 P&P:35

*Correlations with Pass-Fail are expressed in terms of biserial coefficients. Other coefficients are Pearsonian.

paper-and-pencil tests and this Time criterion approach the lower limits of statistical, and possibly practical, significance only in terms of the Boston Spring sample. In general, there are no consistent relationships evident between the predictor variables and the criterion Time for Stage A.

A similar situation is evident in regard to Total Time for Course, when data from all four samples are considered, although a few relatively high correlations on single samples are evident. The correlation between Vital Capacity and Total Time [illegible] on the basis of the Midwest Summer group, but small and negative [illegible] three samples. Similarly, the correlations between Mashburn scores and Total Time range from .41 to .50 on the Boston Fall sample, but are .16 or lower on the other three.

In considering, in summary fashion, the correlations between the selected predictors and Pass-Fail, Time for Stage A, and Total Time, it is evident that none of these variables predict any of the three criteria consistently over the four samples, and that in general, the correlations are not high. The best prediction, in terms of more than a single sample, was between Two-Hand Coordination Test scores, Mashburn Test scores, and the criterion Time for Stage A, when data from the Midwest group only (Spring and Summer samples) are considered.

In Table 10 are presented the correlations between the selected predictors and ratings on the Purdue Rating Scale of Pilot Competency[35] made at Stage A and also at Stage D. Data are available only on three samples, the Boston Spring group, and the Midwest Spring and Summer groups.

Examination of the correlations between physiological measures and ratings at Stage A indicate that, in terms of the size of the samples on which the coefficients are based, none differ significantly from zero, the highest value (between Pilot Data, Flying, and ratings at Stage A) being .36.* Furthermore, no consistent trends over the three samples, even in regard to the sign of the coefficients, are evident. Inspection of the correlations between the physiological measures and ratings at Stage D indicates that while the coefficients are in general, negative, none of the variables consistently predict the criterion to any marked degree over the three samples.

Similarly, there are no consistent relationships between the psychomotor measures and ratings at Stage A, or at Stage D, except the fact that in general, the coefficients are in the expected direction, i.e., the Two-Hand Coordination Test, on which a high score denotes good performance, correlates negatively, while the Mashburn Test, on which a [illegible] denotes poor performance), correlates positively [illegible] Purdue Scale.

---

[35] See Kelly, G. I., [illegible] (referred to in Footnote 17.)

* Since the [illegible] is as great as .29, a value of approximately [illegible] the standard error of a coefficient

TABLE 30

CORRELATIONS BETWEEN SELECTED PREDICTORS AND PURDUE RATING SCALE

MIDWEST AND BOSTON (Spring Groups)

Variable	PURDUE RATING SCALE STAGE A Midwest Spring	Midwest Summer	Boston Spring	PURDUE RATING SCALE STAGE D Midwest Spring	Midwest Summer	Boston Spring
Physiological Tests (Phys.)						
Body surface	.14	-.16	-.28	.16	-.01	-.15
Vital capacity	.21	.23	-.18	.14	-.03	-.07
Tidal air	.04	.15	-.23	-.22	-.34	-.21
Tidal air/Body surface	.00	.21	.00	-.26	-.26	-.15
Vital cap./Body surface	.17	.36	-.09	.02	.01	-.04
Pulse rate (lying)	.19	-.18	.10	-.13	-.02	.04
Psychomotor Tests (P-M)						
Two-Hand: trial 6	.33	.03		.14	.28	
Two-Hand: Mean	.44	.04	.10	.07	.22	.11
Mashburn: through trial 39	-.29	-.07	-.19	-.08	-.39	-.13
Mashburn: trials 14 to 26	-.26	-.06	-.27	-.06	-.23	-.18
Mashburn: trials 27 to 39	-.24	-.09	-.16	-.03	-.42	-.10
Eye-Hand: pattern A	-.17	-.08	-.17	-.03	-.31	-.13
Eye-Hand: pattern B	-.11	-.04	-.21	-.01	-.26	-.10
Eye-Hand: pattern D	-.36	-.16	-.22	-.24	-.29	-.14
Paper-and-Pencil Test (P and P)						
BI ± 1% A			.35			.43
BI + 1% A	.15	.18	.36	-.01	.07	.40
BI + 1% B	.25	-.12		.02	.44	
BI ± 1% A + B			.27			.33
BI + 1% A + B	.22	.12	.37	.02	.26	.39
Mechanical Comprehension	.24	-.04	.23	.19	.48	.13
Otis	.31	-.33	-.05	.10	-.10	-.16
Aviation Information		-.57			-.35	
Number of Cases (N)	25	P & P: 9 Others: 25	76-82	Phys.: 34 Others: 39-40	P & P: 12 Others: 34	76-82

(A few relatively high coefficients appear in the case of specific samples.) The mean score of the Hand-Eye Coordination Test correlates with Stage A rating on the Midwest Spring sample, .44, and trial 39, and trials 27 to 39 on the Mashburn Test correlate with Stage D rating -.39 and .42 on the Midwest Summer sample. Comparable coefficients on the other samples are markedly lower, being less than .10. A similar situation exists in regard to the Eye-Hand Coordination Test, although Eye-Hand pattern D correlates better than .20 with Stage A and Stage D ratings respectively, on two out of the three samples.

The correlations between the scores on the Biographical Inventory and ratings at Stage A and at Stage D are relatively high for the Boston Spring sample, but are considerably lower when the Midwest samples are considered.[37] However, the number of cases in terms of which the Midwest correlations were computed, particularly on the Summer group, is extremely small. The correlations of the Test of Mechanical Comprehension, and the Otis with the criterion are not consistent over the three samples. The Test of Aviation Information was administered only to the Midwest Summer sample. Examination of Table 10 indicates that scores on this test are negatively related to ratings on the Purdue Scale and rather markedly so, although the number of cases on which these correlations are based is extremely small.

In Table 11 are presented the correlations of the selected predictors with the criterion measures obtained from the photographic records (Camera Criteria), and the Ohio State Flight Inventory (OSFI) Summation Scores.[38] These measures were available from the Midwest (Spring and Summer) samples only. Camera Criterion measures were taken of performance at the end of the course. Ohio State Flight Inventory measures were taken at the end of Stage A (i.e., approximately at the time of soloing), and at the end of Stage D, the final "stage" in the flight training course as then constituted.

---

[37]It should be noted that only the scores based on summation in terms of positive item weights on Section A and on Sections A and B, of this inventory were obtained on both the Boston and the Midwest group. For the Boston group other scores were obtained in terms of the algebraic summation of positive and negative item weights on Section A, and on Sections A and B of this inventory. For the Midwest group the only other score computed was the summation in terms of positive item weights on Section B.

[38]These criteria have been defined on pages 7-13 of this report. In summary it may be noted that in terms of Camera Criterion V the given sample was divided into three categories, "Good," "Average," and "Poor," based on ratings of flight proficiency made by observers who viewed the photographic records in slow motion projection. Camera Criterion VI divided the sample into two categories (upper half and lower half) in terms of flight proficiency. The OSFI summation score represents a measure derived from a summation of the plus and minus unit weights carried by the items on the inventory which are marked during a given flight. The actual scores used in the analysis are standard scores, in terms of which a low score denotes poor performance.

TABLE 11

CORRELATIONS OF SELECTED PREDICTORS WITH CAMERA CRITERIA AND OHIO STATE FLIGHT INVENTORY

Variable	MIDWEST SPRING				MIDWEST SUMMER			
	Cam. Cr. A	Cam. Cr. VI	OSFI Sum. A	OSFI Sum. D	Cam. Cr. A	Cam. Cr. VI	OSFI Sum. A	OSFI Sum. D
Physiological Tests (Phys.)								
Body surface	-.39	-.21	-.30	-.11	-.09	-.03	.31	.00
Vital capacity	-.43	-.16	-.41	-.14	.30	.14	.08	.11
Tidal air	-.26	-.20	.09	-.03	.16	.39	.25	.14
Tidal air/Body surface	-.17	-.16	.14	-.02	.16	.38	.21	.14
Vital cap./Body surface	-.29	-.07	-.36	-.09	.36	.19	.00	.10
Pulse rate (lying)	-.04	.05	-.09	.01	.13	.04	.25	-.19
Psychomotor Tests (P-M)								
Two-Hand: trial 6	-.34	-.20	-.36	-.25	-.09	-.46	-.26	-.34
Two-Hand: Mean	-.34	-.36	-.27	-.38	-.10	-.44	-.14	-.24
Mashburn: through trial 39	.37	.34	.40	.46	.00	.15	-.03	.31
Mashburn: trials 14 to 26	.24	.19	.50	.33	-.10	.10	.02	.28
Mashburn: trials 27 to 39	.16	.18	.29	.31	.03	.05	-.31	.24
Eye-Hand: pattern A	.15	.06	.34	.21	.21	.60	.22	.41
Eye-Hand: pattern B	.21	.25	.08	.32	.28	.58	.21	.44
Eye-Hand: pattern D	.34	.20	.50	.31	.13	.44	.26	.38
Paper-and-Pencil Tests (P and P)								
BI + 1% A	.19	.33	-.13	.11	-.45	-.19	-.11	-.07
BI + 1% B	-.08	.04	.00	-.06	.12	-.09	-.35	-.48
BI + 1% A + B	.13	.28	-.13	.07	-.38	-.22	-.24	-.28
Mechanical Comprehension	-.38	-.24	-.30	-.30	.05	-.45	-.23	-.39
Otis	-.06	-.19	.39	.00	.43	.20	.10	.00
Aviation Information					.08	.13	.03	.08
Number of Cases (N)	29-34	29-34	25	Phys.: 34 Others: 47-48	P & P: 11-12 Others: 33	P & P: 11-12 Others: 33	P & P: 9 Others: 25	P & P: 11-12 Others: 34

Examination of the correlations between the physiological measures and the Camera Criteria V and VI indicate no consistent relationships when data from the Spring and Summer samples are considered. The correlations based on the Summer sample are predominantly positive, from the Spring sample predominantly negative.[39] A similar situation is evident from examination of the correlations between these physiological measures and the OSFI scores for Stage A and Stage D. Although certain of the coefficients are relatively high (Tidal Air against Camera Criterion V is -.43 for the Spring group), the general reversal in sign from Spring to Summer groups and the small N renders any generalizations regarding the predictive value of these physiological variables extremely hazardous.

Examination of the correlations between the psychomotor tests and the criteria indicates that for both samples, criterion correlations of the Two-Hand Coordination Test are negative, while the criterion correlations of the Mashburn Test and Eye-Hand Coordination Test scores are in general, positive. The signs of these coefficients are in the expected direction since in terms of the criteria, and in terms of the Mashburn Test and Eye-Hand Coordination Test low scores indicate good performance, while on the Two-Hand Coordination Test high scores denote good performance.

Other than this, little of predictive significance is indicated by these coefficients, particularly in view of the N on which they are based, and in view of the fact that the correlations with Camera Criterion VI are biserials. Certain of the coefficients between the Two-Hand scores and the criterion measures are relatively high, lying between the 5% and 1% level of significance, but considerable variation in the size of the coefficients is evident, e.g., for the Spring sample the coefficients vary from -.20 to -.43, and for the Summer sample from -.09 to -.46. No generalizations can be made regarding differences between correlation of the Two-Hand Coordination Test with various criterion measures, or between samples.

A similar situation prevails in regard to the scores on the Mashburn. In terms of the Spring sample, the Mashburn score through trial 39 predicts all of the criteria except Camera Criterion VI relatively well, the correlations being significant at better than the 5% level. However, in terms of the Summer sample, all criterion correlations involving the Mashburn are extremely low.

Again examination of the correlations between the Eye-Hand Coordination Test and the Criteria indicates that few consistent relationships from sample to sample are evident, although in single samples certain of the individual coefficients are relatively high.

In discussing the relationship between the paper-and-pencil tests and the criterion measures it should be noted that for the Summer sample,

---

[39] High scores on the Criteria and the OSFI scores indicate poor performance.

the N's on which the correlations are based are extremely small, i.e., between 9 and 12 cases. It should be noted, however, in regard to these tests that a negative correlation indicates predictive value, since the criterion scales are inverted, a low score denoting a good performance, while in the case of the paper-and-pencil tests a high score denotes good performance.

While there are a few relatively high correlations between scores on the Biographical Inventory (B.I.) and the criteria, particularly in terms of the extremely small Summer group, there are no consistent relationships between this test and the criterion measures. It may be of some interest, however, to note that in terms of the Spring sample in which a greater number of cases are represented, the highest correlations between the B. I. and the criteria are positive, indicating a negative relationship between scores on this test and flight proficiency.

The correlations between the Test of Mechanical Comprehension (M.C.) and the criteria are all negative (with one exception) indicating at least that some measure of prediction is possessed by this test. Although the correlations between this test and the criteria, particularly for the Spring sample, are higher than are the correlations involving other paper-and-pencil measures, only one correlation is significant at the 5% level.

The correlations involving the Otis test vary considerably in magnitude, and the Aviation Information Test, on the basis of the extremely small Summer Sample, does not predict the criteria. In considering the correlations between paper-and-pencil tests and the criteria it is evident that few consistent relationships exist with the possible exception of the Test of Mechanical Comprehension. In fact, the entire table of correlations between the selected predictors and the criterion measures involving the photographic records and the OSFI is characterized by few meaningful trends when both samples are considered. In terms of these data it might be said, however, that the measures on the Two-Hand Coordination Test, trial 39 on the Mashburn Test, and the Test of Mechanical Comprehension show relatively the most promise.

The correlations between the Aviation Interview scores and the various criterion measures are summarized in Table 12. The Aviation Interview scores were obtained only from the subjects in the Spring groups of the Boston and Midwest samples. The correlations from the Midwest sample involving the criteria Pass-Fail, and the Time measures are based on from 106 to 113 cases taken from three schools, who were interviewed by three different interview boards. Fewer cases were available for the criterion measures involving the Purdue Scale, and for the Camera Criteria and the OSFI which were obtained only in the Midwest sample.

Examination of the correlations from the Boston sample indicates that the coefficients are in general, low, particularly against the criteria Pass-Fail and Time Stage A. Although none of the coefficients are high (maximum correlation coefficient .30), Scales D, E,

TABLE 12

CORRELATIONS BETWEEN AVIATION INTERVIEW SCALES AND CRITERIA

	Midwest Spring	Boston Spring	Midwest Spring	Boston Spring	Midwest Spring	Boston Spring	Midwest Spring	Boston Spring	Midwest Spring	Boston Spring	Midwest Spring Camera Criteria V	Midwest Spring Camera Criteria VI	Midwest Spring OSFI Sum. Sc. St. A	Midwest Spring OSFI Sum. Sc. St. D
	Pass-Fail		Time St. A		Time Tot.		Purdue Sc. Stage A		Purdue Sc. Stage D					
Scale A: Academic Background	-.11	.07	-.18	.03	-.32	-.04	.06	.21	.23	.21	-.26	-.14	-.17	.00
Scale B: Family and Social Background	.06	-.02	-.18	-.05	-.28	-.11	.30	-.02	.12	.15	-.02	-.10	-.23	.03
Scale C: General Social Adjustment	.05	.07	-.11	-.03	-.42	-.19	.28	.10	-.01	.16	.02	-.21	.00	-.02
Scale D: Desire to Fly	.10	.05	-.20	-.13	-.37	-.21	.03	.21	.19	.28	-.34	-.44	-.14	-.04
Scale E: Hobbies, Diversions, etc.	-.09	.22	-.26	-.15	-.33	-.26	.09	.18	.18	.27	-.50	-.36	-.16	-.08
Scale F: Athletic Activities	.07	-.01	-.23	.01	-.54	-.05	.00	.01	.07	.09	-.21	-.48	-.18	-.15
Scale G: Personality	.05	-.07	-.18	.01	-.45	-.11	.16	.13	.14	.16	-.22	-.35	-.04	-.17
Scale H: Appearance and Physique	-.06	-.09	-.17	-.08	-.45	-.18	.22	.12	.07	.05	-.10	-.18	-.05	-.09
Scale I: Fitness for Flight Training	.01	-.12	-.28	.07	-.47	-.13	.18	.23	.13	.30	-.29	-.34	-.22	-.06
Mean Scale Rating	.03	-.01	-.22	-.06	-.50	-.18	.09	.15	.16	.21	-.26	-.38	-.16	-.09
N	113	60	112	54	106	53	20	49	38	52	33	33	47	73

[illegible] to fly [illegible] correlate [illegible], highest with the [illegible]

Turning to the coefficients [illegible] or the Midwest sample, it is evident that while, as in the case of the Boston sample, the correlations against Pass-fail are generally low, a number of relatively high correlations are evident between interview scores and the several criteria which were available from both Boston and Midwest samples, particularly Total Time. The correlations of Scales A to I with Total Time vary between -.45 and -.54. The correlations of interview scores with Time Stage A, and with the measures involving the Purdue Rating Scale are markedly lower, however. Correlations between certain of the interview scales and the camera criteria are relatively high, particularly Scales D, E, F, I, and the mean scale rating. Correlations against the OSFI are, however, lower.

It should be noted that the correlations from the Midwest sample are in general higher than the correlations based on the Boston sample, and that in general there is little relationship between the relative size of the coefficients in terms of the two samples, i.e., the scales which best predict given criteria in the Midwest sample in general do not show such relative prediction efficiency in terms of the Boston sample. In terms of the Midwest sample alone, however, Scales F to I correlate well with Total Time.[40]

DEVELOPMENT OF COMPOSITE CRITERIA

In considering flight criteria it should be recognized that there is no single measure of the broad and inclusive variable which is frequently referred to as general flight proficiency. Instead, it is necessary to obtain measures of a number of more specific variables or criteria which, logically, can be expected to be associated with general flight proficiency. Since each specific criterion represents a measure of only a part of the broad variable "General Flight Proficiency" it seemed desirable to combine the best of the specific criteria into a composite criterion. Before selecting criteria for inclusion in the composite, however, it is necessary to examine the various criteria individually.

Evaluation of Criteria

Ohio State Flight Inventory. The principal measure derived in terms

---

[40]The data here summarized are the same as those treated in more detail by Dunlap, J. W., and Wantman, M. J. Op. cit. (Referred to in footnote 10.) These authors conclude that while the interview "did well in predicting certain of the criteria of competence in flying" it fails to have practical significance in view of the fact that it is extremely expensive when compared to group-administered paper-and-pencil tests, and adds little or nothing to the predictive power of such tests when included in a battery with them.

of this instrument was the "summation score." As described on page 13, the summation scores were determined by summing maneuver scores for the flight and dividing this sum by the number of maneuvers on which it was based, yielding in effect a mean maneuver score. Maneuver scores were obtained by converting the algebraic sum of the unit weights for the marked items on the inventory sheet for a given maneuver into standard scores based on the distribution of algebraic sums on approximately 175 Stage D flights. These standard scores ranged from 1.0, best performance, to 5.0, poorest performance.

In addition to the summation scores, "profile" scores were also determined for 66 of the cases. These scores were based on the rankings by six raters of "profiles" of maneuver scores for the check flight. The rank orders were converted to scale scores, and the scale scores for the six raters summated. Although the profile scores will not be considered further in this report, it may be noted in passing that the correlations between the profile scores and the summation scores were .94 and .91 for the Spring and Summer samples, respectively.[41]

Camera Criteria. Two types of scores were available from the analysis of the photographic records: (1) Criterion Ratings and (2) Criterion Flight Scores.[42] The reliability coefficients of the flight scores (determined by correlating two measures, taken independently on the same data) were .87 and .97, respectivaly, for the two Midwest samples. The reliability of the Criterion Ratings cannot be compared directly with the reliability of the Flight Scores, since in the former case the reliability was expressed in terms of corrected coefficients of contingency. The Criterion Ratings, however, apparently were comparable in terms of reliability, the corrected contingency coefficients being .75, .86, and .95, respectively, for three groups of students rated.

These measures seemed sufficiently reliable for use as criterion data, and although no experimental data on the validity of these measures are available, they possess a "logical" or "face" validity since the ratings were based on objective records of flight performance, and since the observational procedures provided for (1) a detailed comparison of how the maneuvers were performed against how they should have been performed (i.e., ideal performance), and (2) a careful cross-check between independent ratings of two observers.[43]

Other Criteria. Except for determination of their relationships with other criteria, evaluation of the remaining criteria, on the basis

---

[41]Detailed description of the methods of scoring the inventory and research data from its use is found in: Op. cit. (Referred to in Footnote 21.)

[42]As noted previously, the Criterion Flight Scores represent the summation of ratings given by two observers on eight "aspects" of flight performance.

[43]See Viteles, H. G., and Thompson, L. G. Op. cit. (Referred to in Footnote 18.)

of experimental results, as impossible. It should be noted, however, that "time measures," since they represent an index of the amount of training time necessary before a student pilot is considered ready to advance from one "stage" of flight training to the next, or is considered to show sufficient proficiency to be graduated from the training course, apparently have considerable face validity. On the other hand, this measure is dependent upon the validity and reliability of individual instructors' judgment, and the situation is further complicated by the restrictions as to minimum time set by CAA regulations and by the fact that procedures for granting extra time may vary from airport to airport.[44]

No experimental data on the reliability of the Purdue Rating Scale are available, and in this study only item 14 of that instrument was used. However, it should be noted that this scale defines a clear-cut continuum in terms of which over-all flight proficiency can be rated, and that experimental evidence of the validity of this instrument is available.[45]

While the Pass-Fail criterion might be considered of great practical importance, it is suspect because of its non-critical nature and because of its probable unreliability. Subjects who are "washed out" or drop out because of failure in ground school, because of sickness, or for disciplinary or other reasons are classed as "failures." Furthermore, the flight inspector's grade at the end of the course, on the basis of which a large proportion of the "washouts" are failed, is based on a single check-flight. This short sample of the subject's performance may or may not be a representative and reliable work sample indicative of the individual's ability.[46]

---

[44] In a study by R. C. Rogers, Training time as a criterion of flight proficiency (unpublished report in the files of the NRC Committee on Selection and Training of Aircraft Pilots), it was concluded that, except in the case of extremely poor flight trainees, subjects who took longer to solo, or to go from one stage to the next, or to complete the course, did not necessarily represent poorer end products than did those who proceeded more rapidly.

[45] An analysis of the Purdue Rating Scale is given in: Kelly, E. L., Op. cit. (Referred to in Footnote 17.) In this report, data on the validity of the various items in the instrument are presented in terms of the degree to which the items differentiated previously selected criterion groups.

[46] For an evaluation of inspectors' ratings, see: Johnson, H. M., and Boots, M. L. Analysis of ratings in the preliminary phase of the CAA training program. Washington, D. C.: CAA Division of Research, Report No. 21, October 1943. Also, Festinger, L., Kogan, L. S., Odbert, H. S., and Wapner, S. An analysis of inspectors' ratings on Form ACA 342Z. (Final report to be published in the CAA Technical Series of the NRC Committee on Selection and Training of Aircraft Pilots.)

The reliability of a single flight, as measured by photographic records of student pilot performance on two successive flights, is described in: Wapner, S., Festinger, L., and Odbert, H. S. Consistency of student pilot performance as observed in photographic records. Progress report, January 1945. (Report in the files of the NRC Committee on Selection and Training of Aircraft Pilots.)

Interrelationships Between Criteria. In developing a composite criterion, consideration of the interrelationships between specific criteria is important. In Table 13 are presented the intercorrelations between Time Measures and the ratings on the Purdue Scale. With the exception of Time for Stage D all of these measures are available from both the Midwest Spring and Summer, and the Boston Spring samples. It is to be noted that intercorrelations between criteria in terms of the Boston Fall sample are not available.

Examination of Table 13 indicates that the only marked relationship between criteria available on all three samples is that between the ratings on the Purdue Scale at Stages A and D, respectively, the coefficients varying between .54 and .74.[47] It should be noted that these measures are not completely independent, both being made by the same individual, i.e., the student's flight instructor. The correlations between Time for Stage D and Total Time are high, in terms of the Midwest samples, but this relationship is rendered spurious by the fact that Time for Stage D is included in Total Time, and further by the fact that the majority of the extra time allotted to the various students was probably granted in Stage D.

The correlation between Total Time and Time for Stage A is high for the Boston sample, but low for the Midwest samples. It should be noted, of course, that Time for Stage A is included in Total Time. The relationship between the ratings on the Purdue Scale and the Time Measures is relatively high for the Midwest Summer sample, but markedly lower for the other samples. With the exception of the correlations between Time for Stage A and Time for Stage D, which are low but negative, it can be stated that as measures of proficiency the remaining measures are positively related, since although the correlations between Time and the Purdue Scale are negative, time spent in the course is negatively related to proficiency.

In Table 14 are presented the intercorrelations of the Camera Criterion measures, the OSPI measures, and Ground School grades, and in addition, the correlations of the Time, and Purdue Rating Scale measures with the above criteria. Correlations in this table are from the Midwest Spring and Summer samples only, since the photographic records, and the OSPI measures were not taken on the Boston sample.

Examination of the intercorrelations between Camera Criterion measures and OSPI measures reveals that the only marked relationship in terms of both samples is that between Camera Criteria V and VI. This, however, is to be expected since these two measures

---

[47] In another study it was found that instructors' ratings on Item 14 of the Purdue Scale at the end of Stage A correlated .76 with instructors' ratings on this item at the end of the course, based on a sample of 30 cases. See: Kelly, E. L., and [illegible], W. A. A preliminary study of certain predictors of success in civilian pilot training. Washington, D. C.: CAA Division of Research, Report No. 7, December 1942.

TABLE 1.

INTERRELATIONSHIPS BETWEEN TIME MEASURES AND PURDUE RATING SCALE CRITERION MEASURES

	Midwest Spring					Midwest Summer					Boston Spring			
	Time Stage A	Time Stage D	Time Total	PRS Item 14 A	PRS Item 14 D	Time Stage A	Time Stage D	Time Total	PRS Item 14 A	PRS Item 14 D	Time Stage A	Time Total	PRS Item 14 A	PRS Item 14 D
N =	164	164	164	25	42	90	90	90	25	34	54	53	49	32
Time Stage A	---					---					---			
Time Stage D	-.06	---				-.18	---				---			
Time Total	.19	.90	---			.15	.76	---			.73			
PRS Item 14 A	-.42	-.01	-.21	---		-.23	-.49	-.40	---		.00	-.13	---	
PRS Item 14 D	-.24	-.07	-.27	.54	---	-.60	-.43	-.55	.68	---	.26	-.29	.74	---

TABLE 14

INTERCORRELATIONS BETWEEN CRITERIA, MIDWEST SAMPLE

	Midwest Sample					Midwest Suspect			
	Cam. Crit. V	Cam. Crit. VI	OSFI Stage A	OSFI Stage B	Grad. Sch. Grds.	Cam. Crit. V	Cam. Crit. VI*	OSFI Stage A	OSFI Stage B
Cam. Crit. V (C.C.V)	---					1			
Cam. Crit. VI (C.C.VI)	.90	---				.97	1		
OSFI Stage A (OSFI-A)	.60	.65	---			.16	.46	1	
OSFI Stage B (OSFI-B)	.61	.64	.47	---		.22	.34	.62	1
Grad. Sch. Grades	-.17	.06	.24	.07	---	-.13	-.21	.15	.08
Time Stage A	.46	.53	.30	.30	.02	.42	.52	.20	.23
Time Stage B	.13	.24	.34	.21	-.10	.21	.14	.21	.14
Time Total	.42	.48	.37	.24	-.03	.44	.45	.43	.23
PRS Item 14 A (PRS-A)	-.35	-.69	-.24	-.05	.16	-.26	-.35	-.04	.07
PRS Item 14 B (PRS-B)	-.14	-.07	-.48	-.12	.36	.04	-.07	.00	-.13
Number of Cases (N)	PRS-A & OSFI-A: 17 Others: 35-36	PRS-A & OSFI-A: 17 Others: 45-46	PRS-A & OSFI-A: 17 Others: 21-24 PRS-B & OSFI-A: 17 Others: 45-46	PRS-A & OSFI-B: 17 Others: 21-24 PRS-B & OSFI-B: 17 Others: 45-46	Time measures: 127-135 PRS-A, OSFI-A: 24 PRS-B, OSFI-A: 24	Time measures: 127-135 PRS-A, OSFI-A: 24 PRS-B, OSFI-A: 24	PRS-A, OSFI-A: 24 Others: 33	24-25	PRS-A, OSFI-A: 24 Others: 33-34

*The correlations between Cam. Crit. VI and each of the other criterion variables for both samples are in terms of biserial r's.

represent merely different types of [illegible] Criterion V placing the subjects in three groups (above average, average, and below average) and Criterion IV in two groups (upper half and lower half).

The correlation between the Camera Criteria (particularly Camera Criterion V) and the OSFI measures is relatively high for the Spring sample, but markedly lower in terms of the Summer sample. The correlation between the OSFI summation scores for check flights at Stages A and D respectively is relatively high for the Summer sample, but considerably lower for the Spring sample. Correlations between ground school grades and other criteria are in general low. However, although low consistent relationships are evident, the correlations between the Camera Criterion measures and the OSFI measures are positive.

In the last five rows of Table 1 are presented the correlations of the Purdue Rating Scale and Time measures with the Camera Criterion and the OSFI measures. It is evident that the correlation of the Camera Criteria with Time for Stage A and with Total Time are relatively high for both samples, being in the neighborhood of .45. The correlations with Time for Stage D are uniformly lower. The correlations of Time A and Total Time measures with the Camera Criteria are higher, for both samples, than are the correlations between these Time measures and the OSFI measures. Although the correlations are not high there is a greater relationship (with one exception) between the Camera Criteria and ratings on the Purdue Scale at Stage A, than between the Camera Criteria and Purdue Scale ratings at Stage D. This trend is indicated in both samples. In summary it can be noted that, as measures of flight proficiency, the above criteria are positively related, the negative correlations of the Purdue scale with the other criteria being accounted for by the fact that for these measures alone a high score denoted excellence of performance, while for the other measures a low score denoted excellence. In general, the Camera Criteria correlated higher with the other criteria than did the OSFI measures, and both of these measures, in general, correlate higher with Time for Stage A than with Time for Stage D.

No correlations have been presented between Pass-Fail and the other criteria, since data on these relationships are available for only the Midwest Spring sample. For this sample, these correlations (biserials) are presented below. (No data are available for failing cases on Purdue Rating Scale D or the Camera Criteria.)

	Time Stage A	Time Stage D	Time Total	Purdue Scale A	OSFI Stage A	OSFI Stage D	Grnd. Sch. Grade
Pass-Fail	.20	.41	.43	-.31	.49	-.24	.19
N	181	164	164	25	52	43	135

Composite Criteria. Four measures were selected for inclusion in the composite criterion, namely:

Total Time

Purdue Rating Scale

Ohio State Flight Inventory

Camera Criterion

Selection of these measures was determined by their apparent validity, because of their intercorrelations, and in the case of the Camera Criterion because it was demonstrably reliable. Conversely, the other two criteria available, Inspectors' Flight Test grade, and Ground School grades, were both unsatisfactory.

Two types of criterion distributions were set up, (a) dichotomous composite criterion, and (b) continuous composite criterion.

Dichotomous Composite, Four-fold Criterion. Four-fold criterion cut-off points, or dichotomies, were established authoritatively so that unsatisfactory performance on the four criteria was defined as follows:

1. Total Time -- 40 hours or greater

2. Purdue Rating Scale -- score of 7 or lower

3. OSFI -- summation score of 3.6 or lower

4. Camera Criteria -- rating of 0 (i.e., rated in "Poor" category)

There were thus 16 different criterion patterns of these four criteria which could be obtained. Two methods of scoring this criteria were developed, the hurdle method in which a subject must obtain "satisfactory" status on all four criteria to be considered as passing on the Composite, and the compensation method, in which a subject passes on the composite if he obtains "satisfactory" status on three of the individual criteria, or on both Criteria 3 and 4 (OSFI, and Camera Criteria, respectively).

Dichotomous Composite, Three-fold Criterion. Since photographic records were not available on all subjects, a three-fold composite was also set up in which the Camera Criterion was eliminated, and for which a larger number of cases would be available than for the four-fold criterion. This criterion was also scored by the hurdle method, in which satisfactory status on all criteria was required for passing, and by the compensation method, which required that a passing subject obtain

TABLE 15

DISTRIBUTION OF [illegible] COMPOSITE
[illegible] STATE UNIVERSITY AND UNIVERSITY OF MICHIGAN (Spring 1942)
(N = 35*)

	Desirable						Undesirable									
Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Criterion I	+	-	+	-	+	+	-	+	+	-	+	-	+	-	-	-
Criterion II	+	-	-	+	+	+	+	-	+	+	-	-	-	-	+	-
Criterion III	+	+	+	+	+	-	+	-	-	-	+	-	-	+	-	-
Criterion IV	+	+	+	+	-	+	-	+	-	+	-	+	-	-	-	-
N	25	1		2			2		2	1					2	
%	71.4	2.8		5.7			5.7		5.7	2.8					5.7	
	N = 28		% = 80				N = 7			% = 20						

*There were 9 additional failing cases for this group on which complete data were not available.

TABLE 15

DISTRIBUTIONS OF FOUR CRITERIA COMPOSITE
OHIO STATE UNIVERSITY AND UNIVERSITY OF DAYTON (Summer 1942)
(N = 33*)

	Desirable						Undesirable									
Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Criterion I	+	-	+	-	+	+	-	+	+	-	+	-	+	-	-	-
Criterion II	+	-	-	+	+	+	+	-	+	+	-	-	-	-	+	-
Criterion III	+	+	+	+	+	-	+	-	-	-	+	-	-	+	-	-
Criterion IV	+	+	+	+	-	+	-	+	-	+	-	+	-	-	-	-
N	23		1		5	1	1		1					1		
%	70		3		15	3	3		3					3		
	N = 30		% = 91				N = 3			% = 9						

*There were 2 additional failing cases from the Dayton group on which complete data were not available.

TABLE 17

SUMMARY DISTRIBUTION OF FOUR CRITERIA COMPOSITE
(Spring and Summer)
(N = 68)

Pattern	Desirable 1	2	3	4	5	6	Undesirable 7	8	9	10	11	12	13	14	15	16
Criterion I	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Criterion II	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Criterion III	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Criterion IV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
N	48	1	1	0	7	1	3		3	2				1	2	
%	70.6	1.5	1.5	0	10.3	1.5	4.4		4.4	3.4				1.4	2.9	
	N = 58		% = 85				N = 10			% = 15						

TABLE 18

DISTRIBUTION OF FOUR CRITERIA COMPOSITE
PURDUE UNIVERSITY (Spring)
(N = 36)

Pattern	Desirable 1	2	3	Undesirable 5	6	7	8
Criterion I	+	+	+	+	+	+	+
Criterion II	+	+	+	+	+	+	+
Criterion III	+	+	+	+	+	+	+
N	1	1		2	2	4	
%	5	10		10		22	
	N = 12			N = 8		N = 26	

* There was [illegible] in each group on which complete data were [illegible]

DISTRIBUTION OF THREE CRITERIA COMPOSITE
OHIO STATE UNIVERSITY AND UNIVERSITY OF MICHIGAN (Spring)
(N = 41*)

Pattern	Desirable 1	2	3	Undesirable 4	5	6	7	8
Criterion I	+	-	+	+	-	-	-	+
Criterion II	+	+	-	+	+	-	-	-
Criterion III	+	+	+	-	-	-	+	-
N	31	3		3	3		1	
%	76	7		7	7		2	
	N = 34	% = 83		N = 7	% = 17			

*There were 9 additional failing cases for this Spring group on which complete data were not available.

TABLE 20

DISTRIBUTION OF THREE CRITERIA COMPOSITE
OHIO STATE UNIVERSITY AND UNIVERSITY OF DAYTON (Summer)
(N = 34*)

Pattern	Desirable 1	2	3	Undesirable 4	5	6	7	8
Criterion I	+	-	+	+	-	-	-	+
Criterion II	+	+	-	+	+	-	-	-
Criterion III	+	+	+	-	-	-	+	-
N	29	1	1	2			1	
%	85	3	3	6			3	
	N = 31	% = 91		N = 3	% = 9			

*There were 2 additional failing cases from the Dayton group on which complete data were not available.

Table 21

SUMMARY DISTRIBUTION OF THREE CRITERIA COMPOSITE
(Spring and Summer)
(N = 207)

Pattern	Desirable 1	2	3	Undesirable 5	6	7	8
Criterion I	-	-	-	-	-	-	-
Criterion II	-	-	-	-	-	-	-
Criterion III	-	-	-	-	-	-	-
N	70	7	1	5		6	
%	34	3	½	2		3	
	N = 86	% = 42		N = 18		% = 17	

satisfactory status on the Ohio State Flight Inventory, and one other criterion. The distributions of cases on the various patterns of the dichotomous composite are given, for various samples, in Tables 15 to 21.

Continuous Composite Criteria. Since the dichotomous nature of the above composites would require that predictor-criterion correlation coefficients be expressed as biserials, it was considered wise to develop, in addition to the above, a continuous composite. Since all of the above individual criteria were continuous except the Camera Criterion based on three-point ratings, it was decided to substitute for this criterion variable the Flight Scores derived from the photographic records, which were continuous, and which correlated highly with the

[illegible] criterion V, the three-fold [illegible].

As in the case of the dichotomous composite, both four-fold and three-fold composites were set up. It was decided that in the four-fold criterion the OSPI scores and the Geneva Flight Scores were to be weighted twice as heavily as the other two criteria because of their greater objectivity. In the three-fold criterion the Flight Inventory alone carried the double weight, since the Flight Scores were not used.[49]

The procedure for deriving the continuous composite criteria was as follows:

a. Raw scores from the distributions of the four variables listed above were converted to standard scores. Standard scores were computed separately for the maximum number of cases available for the three-fold and for the four-fold composite.

b. Scores were adjusted so that all values were positive, and high scores were "favorable."

c. The resulting scores were weighted as noted above.

d. The appropriate weighted scores were combined by addition to form the four-fold or the three-fold criterion.

The mean of the four-fold composite was approximately 300, and the standard deviation, 30. The three-fold composite yielded a mean and standard deviation of 200 and 17 respectively. The correlations between the four-fold and the three-fold continuous composites were .81 for the Spring sample (N = 32) and .77 for the Summer sample (N = 33).

The intercorrelations among the four criteria used in the composite, based on the cases used in subsequent analyses involving the composite criteria, are summarized in Table 22. It will be noted that the sample on which these coefficients are based is not identical with the samples on which correlations reported on page 44 are based. However, the criterion intercorrelations in Table 22 are similar to those in Tables 13 and 14, except that in Table 22 the correlations between Total Time and Ratings on the Purdue Scale for the Summer group are high.

---

[48]Biserial coefficients of correlation are reported varying from .88 to 1.00 between these two variables, based on two independent samples, and depending upon where on the three-point continuum the dichtomy was made. See: Viteles, M.S., and Thompson, A.S. Op. cit. (Referred to in Footnote 18.)

[49]The size of the N did not warrant the use of statistical methods for determining how the criteria were to be combined or weighted in either the dichtomous or the continuous composite.

TABLE 22

INTERCORRELATIONS OF CRITERIA
MIDWEST PROJECT Spring Four-fold (N = 32)

	Time	Purdue (Item 14)	OSFI	Cam. Aspect
Time	-	.29	.47.	.40
Purdue (Item 14)		-	.09	.22
OSFI			-	.59
Cam. Aspect				-

Summer Four-fold (N = 33)

	Time	Purdue (Item 14)	OSFI	Cam. Aspect
Time	-	.50	.34	.37
Purdue (Item 14)		-	.00	-.02
OSFI			-	.57
Cam. Aspect				-

Spring Four-fold (N = 39)

	Time	Purdue (Item 14)	OSFI
Time	-	.27	.40
Purdue (Item 14)		-	.09
OSFI			-

Summer Three-fold (N = 34)

	Time	Purdue (Item 14)	OSFI
Time	-	.51	.33
Purdue (Item 14)		-	.05
OSFI			-

CORRELATION OF INDIVIDUAL PREDICTORS WITH COMPOSITE CRITERIA

In developing batteries to be used in predicting composite criteria it is important to determine the correlations of individual predictors with the composite criteria.

Correlation of Individual Predictors with Dichotomous Composite.

The correlations of the various predictors with the dichtomous composite criteria, hurdle and compensation scores, are summarized in Table 23. Examination of this table indicates that the correlations between the psychomotor tests and the composite criteria are consistently the highest, with the Mashburn Test showing in general the highest correlations. The Two-Hand Coordination Test and Eye-Hand Coordination Test also show a number of relatively high correlations. Although there is considerable variation among the coefficients for the various groups, there seems to be little to choose between the compensation and hurdle scoring methods as far as predictor-criterion correlations are concerned. Such difference as exists possibly favors the compensation criterion.

The predictor-criterion correlations for the paper-and-pencil tests are in general low, with the possible exception of the Test of Mechanical Comprehension, which correlates as high as .40 with the three-fold criterion. Among the physiological tests, only Vital Capacity and Vital Capacity/Body Surface show sizable correlations with the composite criteria, and this primarily for the four-fold composite on the Spring sample.

The correlations between the Aviation Interview Scores and the composites are in general low, and indicate no significant trend. In no case is a coefficient of greater than .30 evident on the two interdependent samples (i.e., Groups I and II).

The correlations between a number of scores on the Strong Vocational Interest Inventory and the dichotomous composite criteria, for the Midwest sample, Spring group, are presented in Table 24. Examination of these coefficients, which are all biserials, indicates that for the two samples represented by Groups I and II, the highest predictor-criterion correlations involve the Aviator scale. Certain relatively high correlations (in the neighborhood of .60) between criteria, and the Chemist, Engineer, and Physician scales are of interest, although these correlations are found for the Spring sample only. Also of interest are the generally low or negative correlations between the composite criteria and the Strong Vocational Interest Inventory scores on Social Science Teacher, Personnel Manager, Purchasing Agent, and Life Insurance Salesman, and between the criteria and Strong Vocational Interest Inventory scores on Groups V, VIII, IX, and X.

Since, as noted above, the correlation coefficients are biserials, based on small N's and small numbers of failing cases, conclusions must be considered tentative.

TABLE 23

CORRELATION OF INDIVIDUAL PREDICTORS WITH DICHOTOMOUS COMPOSITE CRITERIA
MIDWEST PROJECT 1942

	Predictor	Group*	Four-fold Criterion N	M	σ	rA**	rB**	Three-fold Criterion rA**	rB**	N	M	σ
PHYSIOLOGICAL	Body Surface	I	26	1.9	.1	.19	.26	.03	.23	30	1.9	.1
		II	32	1.9	.1	.35	.35	.24	.38	37	1.9	.1
		III	35	1.9	.1	-.07	.08	-.13	.08	36	1.9	.1
	Vital Capacity	I	26	5019	572	.42	.41	.11	.25	30	4940	572
		II	32	4963	593	.54	.43	.29	.39	37	4916	575
		III	35	4806	776	-.14	.06	-.12	.21	36	4814	767
	Tidal Air	I	26	797	278	.22	.14	.10	-.06	30	780	271
		II	32	766	264	.31	.29	.21	.06	37	759	257
		III	35	906	456	-.18	-.12	-.13	-.04	36	906	450
	Vital Capacity/ Body Surface	I	26	2658	220	.45	.35	.09	.14	30	2627	228
		II	32	2631	224	.46	.24	.17	.20	37	2608	232
		III	35	2526	380	-.16	.04	-.08	.22	36	2533	378
	Tidal Air/ Body Surface	I	26	451.5	149.5	.16	.10	.13	-.10	30	444.7	146.7
		II	32	433.4	141.2	.23	.23	.19	-.01	37	430.0	137.8
		III	35	499.7	237.3	-.19	-.12	-.12	-.04	36	499.7	233.9
	Pulse Rate	I	26	72.7	12.7	-.07	-.13	-.07	-.16	30	72.5	12.4
		II	32	72.5	11.9	.02	-.02	.03	-.05	37	72.6	11.7
		III	35	77.5	12.0	-.06	.08	.01	.06	36	77.6	11.9
PSYCHOMOTOR	Two-Hand Coord. Test: Trial 6	I	26	57.2	14.1	.69	.64	.45	.60	30	57.3	13.4
		II	42	59.1	14.8	.34	.34	.26	.32	48	59.1	13.9
		III	35	59.1	13.8	.41	.29	.64	.41	36	59.2	13.8
	Two-Hand Coord. Test: Mean	I	26	53.2	11.3	.58	.41	.46	.53	30	53.1	11.1
		II	42	53.3	13.6	.28	.28	.25	.27	48	53.4	13.2
		III	35	53.2	14.1	.28	.29	.45	.31	36	53.1	13.9
	Mashburn Test: Through Trial 39	I	26	339.3	57.4	-.88	-.60	-.68	-.72	30	295.2	58.0
		II	42	309.2	55.3	-.65	-.61	-.61	-.61	48	306.2	55.4
		III	35	295.1	47.4	.87	.31	-.88	-.67	36	296.6	47.3
	Eye-Hand Coord. Test: Pattern A	I	26	56.6	7.6	.47	-.25	-.22	-.34	30	56.3	8.2
		II	42	58.1	10.4	-.27	-.26	-.23	-.16	48	58.4	11.2
		III	35	58.1	10.0	-.48	-.63	-.64	-.80	36	59.3	16.6
	Eye-Hand Coord. Test: Pattern B	I	26	37.2	8.2	.40	-.13	-.30	-.31	30	37.1	6.7
		II	42	39.0	10.3	-.15	-.25	-.11	-.12	48	38.8	10.3
		III	35	36.2	9.4	-.50	-.58	-.71	-.70	36	36.1	4.8
	Eye-Hand Coord. Test: Pattern C	I	26	43.9	6.1	.70	.64	-.46	-.54	30	43.8	4.3
		II	42	44.4	9.6	.47	.43	-.35	-.42	48	43.7	8.6
		III	35	36.1	7.2	-.31	-.44	-.71	-.69	36	36.1	7.2

APPENDIX

TABLE 23 (Continued)

CORRELATION OF INDIVIDUAL PREDICTORS WITH DICHOTOMOUS COMPOSITE CRITERIA

Predictor	Group*	Four-fold Criterion N	M	σ	rA**	rB**	Three-fold Criterion rA**	rB**	N	M	σ
Biographical Inventory 1% A	I	26	7.4	2.0	-.10	-.04	.24	.24	30	7.6	2.0
	II	41	7.9	2.4	-.16	-.11	.12	-.12	47	8.1	2.3
Otis Test	III	26	51.2	9.5	.11	.04	.28	.21	30	51.5	9.8
	II	41	52.5	8.8	-.12	-.13	.04	-.03	47	52.9	8.9
Mechanical Comprehension	I	26	39.5	9.5	.30	.23	.40	.34	30	38.6	9.7
	II	41	40.6	10.6	.30	.29	.29	.27	47	40.1	10.5
Aviation Interview											
A	I	26	14.7	5.2	-.05	.01	.09	.04	30	14.8	5.3
	II	39	15.1	5.2	-.01	.08	.15	.25	74	15.5	4.7
B	I	26	15.8	3.5	-.12	-.03	.23	-.10	30	16.0	3.4
	II	39	15.3	3.4	.17	.18	.21	.27	74	16.2	3.5
C	I	26	16.4	3.3	-.30	-.31	-.26	-.27	30	16.4	3.3
	II	39	15.3	3.7	.04	-.04	.30	.26	74	15.6	4.0
D	I	26	14.8	5.1	.09	.09	.18	.07	30	14.9	5.0
	II	39	15.2	4.9	.18	.20	.25	.24	74	15.5	4.5
E	I	26	15.4	4.0	.13	.10	.24	17	30	15.5	4.1
	II	39	15.9	3.7	.03	.10	.19	.17	74	16.2	3.6
F	I	26	15.7	2.9	-.03	-.04	-.06	-.07	30	15.6	3.1
	II	39	14.5	5.3	.18	.16	.31	.40	74	15.1	3.8
G	I	26	16.1	3.5	.06	.09	.03	.05	30	16.1	3.7
	II	39	14.9	4.2	.24	.21	.38	.37	74	15.3	4.3
H	I	26	17.0	3.7	-.06	.04	-.08	-.02	30	17.0	3.8
	II	39	15.7	4.1	.12	.13	.23	.29	74	15.9	4.0
I	I	26	13.6	5.5	-.05	.00	-.06	-.06	30	13.9	5.6
	II	39	13.3	5.2	.06	.10	.25	.25	74	14.3	5.0
Aviation Interview Mean	I	26	15.7	3.3	-.06	-.01	.02	-.02	30	15.8	3.4
	II	39	15.2	3.2	.16	.19	.33	.38	74	15.7	3.2

*Group I = Spring cases complete measures on all variables
Group II = Spring cases maximum N for each variable
Group III = Summer cases maximum N for each variable

**rA = Biserial correlation with 4-criteria compensation composite
rB = Biserial correlation with 4-criteria hurdle composite
rA' = Biserial correlation with 3-criteria compensation composite
rB' = Biserial correlation with 3-criteria hurdle composite

[illegible] and the three-fold
[illegible] criteria are given in Tables 25 to 28.
[illegible] data are treated separately.
[illegible] largest number of cases available
[illegible] number of cases is sometimes
[illegible] should be interpreted
[illegible] caution.

[illegible] that, in general, the Mashburn Test scores show the highest correlation with the criteria, which is in agreement with the results on the dichotomous composite. The Mashburn correlation coefficients with the continuous composite are in the neighborhood of -.40. The highest single predictor-criterion correlation is between Eye-Hand Coordination Test Pattern D score, and the three-fold composite in the spring group. This correlation was -.51 (see Table 25). However, this correlation dropped to -.33 in the summer group (see Table 26), and in general, the three Eye-Hand Coordination Test scores correlate about -.35 with the criteria, showing somewhat higher correlations with the four-fold than with the three-fold criterion. [illegible] also suggests that the Pattern D score on the Eye-Hand Coordination Test tends to correlate more highly and more consistently with the criteria than do Pattern A or Pattern B. The Eye-Hand Coordination Test correlates about .30 with the criteria, and the score [illegible] apparently yields a higher correlation with the criteria than [illegible] of the six trials, in both the Spring and Summer groups.

The only paper-and-pencil test which seems to give a serviceable correlation with the criteria is the Test of Mechanical Comprehension, which correlates about .35 with the four-fold criterion and .32 with the three-fold criterion. The Test of Mechanical Comprehension scores were available only for the Spring group. The Biographical Inventory, the Test, and Personal History Inventory do not correlate to any extent with either criteria.

Aviation Interview scores were available only for the Spring group. These scores were correlated with the criterion. Personality as related to Flying (Scale VII), Appearance [illegible], and Physique as related to Flying (Scale VIII), and Fitness for Flight Training (Scale IX). All three scores correlated more highly with the four-fold criterion than with the three-fold criterion, although none of the coefficients are markedly high, the highest r of .35, being the correlation between Scale VII and the four-fold criterion. This scale also showed the highest coefficient with the three-fold criterion.

The correlations of the physiological measures with the criteria are in general quite low. For the Spring group, Body Surface and Vital Capacity seem to be the best measures, while for the Summer group Vital Capacity and Tidal Air yield the highest correlations. However, because of the radical shifts of the correlation coefficients between Spring and Summer groups, these measures, as well as indices computed from them, must be considered of dubious value.

TABLE 24

CORRELATION OF STRONG PREDICTORS WITH FOUR TYPES OF AUTHORITATIVE DICHOTOMOUS CRITERIA FOR FOUR SAMPLES
MIDWEST PROJECT (Spring Group)

			Four Measures				Three Measures				
Predictor	Group*	N	M	σ	rA**	rB**	rA'**	rB'**	N	M	σ
Aviator	I	26	3.69	1.73	.45	.60	.54	.49	30	3.73	1.63
	II	42	3.86	1.63	.37	.49	.36	.44	48	3.87	1.55
Chemist	I	26	2.62	2.10	.52	.62	.57	.57	30	2.63	1.97
	II	42	3.00	2.14	.14	.25	.12	.19	48	2.96	2.02
Engineer	I	26	2.58	1.98	.48	.59	.53	.49	30	2.67	1.97
	II	42	2.93	2.03	.18	.29	.14	.23	48	2.87	2.01
Math-Science	I	26	4.19	1.66	.09	.23	.13	.20	30	4.13	1.59
	II	42	4.17	1.53	.03	.09	.05	.14	48	4.15	1.49
Psychologist	I	26	1.35	.96	.28	.32	.25	.27	30	1.30	.90
	II	42	1.52	1.18	-.25	-.15	-.22	-.24	48	1.48	1.12
Physician	I	26	2.08	1.41	.49	.59	.55	.53	30	2.17	1.42
	II	42	2.29	1.52	.19	.34	.20	.19	48	2.33	1.50
Group I	I	26	3.04	1.72	.53	.66	.68	.58	30	3.03	1.62
	II	42	3.33	1.66	.20	.34	.25	.20	48	3.31	1.60
Social Science	I	26	3.92	1.92	-.49	-.61	-.66	-.53	30	3.77	1.91
	II	42	3.55	1.89	-.17	-.30	-.22	-.25	48	3.50	1.84
Personnel Manager	I	26	3.46	1.86	-.35	-.32	-.04	-.01	30	3.47	1.84
	II	42	3.29	1.88	-.28	-.21	-.08	-.10	48	3.29	1.81
Group V	I	26	4.62	1.62	-.49	-.61	-.48	-.42	30	4.57	1.58
	II	42	4.26	1.79	-.21	-.30	-.13	-.22	48	4.31	1.73
Purchasing Agent	I	26	3.42	1.55	.12	-.05	-.04	.00	30	3.50	1.50
	II	42	3.19	1.75	.18	.12	.01	.16	48	3.21	1.72
Group VIII	I	26	3.85	1.63	-.30	-.42	-.31	-.29	30	3.93	1.59
	II	42	3.60	1.69	-.02	-.20	-.17	-.07	48	3.62	1.68
Life Insurance	I	26	3.00	1.71	-.24	-.70	-.48	-.51	30	3.10	1.64
	II	42	2.71	1.65	-.23	-.34	-.15	-.27	48	2.79	1.59
Group IX	I	26	4.04	1.76	-.41	-.60	-.36	-.46	30	4.07	1.65
	II	42	3.83	1.66	-.18	-.29	-.11	-.23	48	3.87	1.56
Group X	I	26	2.88	1.40	.21	.00	.27	.17	30	2.90	1.33
	II	42	3.17	1.43	-.10	-.12	.01	-.16	48	3.15	1.41

*Group I = Spring cases with complete measures on all variables
Group II = Spring cases maximum N for each variable including all who were washed out in the Civilian Pilot Training Course.

**rA = Biserial correlation with four-fold compensation composite
rB = Biserial correlation with four-fold hurdle composite
rA' = Biserial correlation with three-fold compensation composite
rB' = Biserial correlation with three-fold hurdle composite

TABLE 25

CORRELATIONS OF VARIABLES WITH FOUR-FOLD CONTINUOUS COMPOSITE MIDWEST PROJECT (Spring Group)

	Variable	r	$S.E._r$	M	σ	N
1.	Four-fold Composite	--	--	300.8	31.4	32
2.	Three-fold Composite	.81	.06	200.0	16.8	32
3.	Two-Hand: trial 6	.38	.15	60.2	14.7	32
4.	Two-Hand: Mean	.29	.16	53.9	11.2	32
5.	Washburn: Through trial 39	-.46	.14	298.8	51.8	32
6.	Eye-Hand: Pattern A	-.27	.16	58.3	7.9	32
7.	Eye-Hand: Pattern B	-.15	.17	39.5	10.2	32
8.	Eye-Hand: Pattern D	-.45	.14	43.8	7.7	32
9.	M.C. (Rights minus one-half number wrong)	.58	.15	40.1	10.0	32
10.	M.C. (Number right)	.39	.15	51.8	6.6	32
11.	Personal History Inventory	.00	.18	24.5	7.7	32
12.	Aviation Interview VII	.32	.17	15.2	3.7	29
13.	Aviation Interview VIII	.19	.13	16.0	4.0	29
14.	Aviation Interview IX	.25	.18	13.0	5.4	29
15.	Biographical Inventory	-.22	.17	8.1	2.4	31
16.	Otis Test	.05	.15	51.2	8.9	32
17.	Aviation Information	...	...	...	...	(No Cases)
18.	Body Surface	.33	.17	186.6	11.7	27
19.	Vital Capacity	.32	.17	49.7	5.4	27
20.	Tidal Air	.07	.19	8.0	2.7	27
21.	Tidal Air/Body Surface	-.02	.19	45.4	14.5	27
22.	Vital Capacity/Body Surface	.15	.19	26.3	2.2	27
23.	Pulse Rate	-.09	.19	70.7	11.1	27

CORRELATIONS OF VARIABLES WITH THREE-FOLD CONTINUOUS COMPOSITE MIDWEST PROJECT (Spring Group)

	Variable	r	S_{E_r}	M	σ	N
1.	Three-fold Composite	--	--	200.1	15.5	39
2.	Four-fold Composite	.31	.06	300.8	31.4	32
3.	Two-Hand: trial 6	.25	.15	60.2	13.5	39
4.	Two-Hand: Mean	.16	.16	54.1	10.7	39
5.	Mashburn: Through trial 39	-.35	.14	296.6	51.2	39
6.	Eye-Hand: Pattern A	-.14	.16	57.7	9.3	39
7.	Eye-Hand: Pattern B	-.09	.16	38.6	10.2	39
8.	Eye-Hand: Pattern D	-.27	.11	43.1	7.8	39
9.	M.C. (Rights minus one-half number wrong)	.33	.14	40.1	10.2	39
10.	M.C. (Number right)	.33	.14	51.8	6.8	39
11.	Personal History Inventory	.09	.16	24.9	7.2	39
12.	Aviation Interview VII	.23	.16	15.3	3.9	35
13.	Aviation Interview VIII	.11	.17	16.0	4.0	35
14.	Aviation Interview IX	.15	.17	13.2	5.3	35
15.	Biographical Inventory	-.15	.16	8.2	2.3	38
16.	Otis Test	.07	.16	51.8	9.1	39
17.	Aviation Information	--	--	--	--	(No Cases)
18.	Body Surface	.16	.17	187.0	12.1	33
19.	Vital Capacity	.12	.17	49.5	5.6	33
20.	Tidal Air	-.13	.17	7.9	2.6	33
21.	Tidal Air/Body Surface	-.17	.17	44.4	13.9	33
22.	Vital Capacity/Body Surface	-.01	.17	26.2	2.3	33
23.	Pulse Rate	-.14	.17	72.9	11.8	33

TABLE 27

CORRELATIONS OF VARIABLES WITH FOUR-FOLD CONTINUOUS COMPOSITE MIDWEST PROJECT (Summer Group)

Variable	r	S.E.r	M	σ	N
1. Four-fold Composite	--	--	300.2	30.0	33
2. Three-fold Composite	.77	.07	200.5	17.7	33
3. Two-Hand: trial 6	.38	.15	60.4	15.2	33
4. Two-Hand: Mean	.34	.15	54.0	13.3	33
5. Mashburn: Through trial 39	-.36	.15	230.2	43.6	33
6. Eye-Hand: Pattern A	-.54	.12	57.1	14.8	33
7. Eye-Hand: Pattern B	-.49	.13	35.6	9.2	33
8. Eye-Hand: Pattern D	-.42	.14	35.7	7.1	33
9. M.C. (Rights minus one-half number wrong)	--	--	--	--	(No Cases)
10. M.C. (Number right)	--	--	--	--	(No Cases)
11. Personal History Inventory	--	--	--	--	(No Cases)
12. Aviation Interview VII	--	--	--	--	(No Cases)
13. Aviation Interview VIII	--	--	--	--	(No Cases)
14. Aviation Interview IX	--	--	--	--	(No Cases)
15. Biographical Inventory	--	--	--	--	(No Cases)
16. Otis Test	--	--	--	--	(No Cases)
17. Aviation Information	--	--	--	--	(No Cases)
18. Body Surface	-.09	.17	287.9	9.2	33
19. Vital Capacity	-.22	.1	45.1	7.7	33
20. Tidal Air	-.37	.1	9.2	4.6	33
21. Tidal Air/Body Surface	-.36	.1	90.3	24.1	33
22. Vital Capacity/Body Surface	-.21	.17	25.4	3.8	33
23. Pulse Rate	-.10	.1	77.8	12.2	33

CORRELATIONS OF VARIABLES WITH THREE-FOLD CONTINUOUS COMPOSITE MIDWEST PROJECT (Summer Group)

		r	S.E. r	M	σ	N
1.	Three-fold Composite	--	--	200.1	17.6	34
2.	Four-fold Composite	.77	.07	300.2	30.0	33
3.	Two-Hand: trial 6	.34	.15	60.2	15.0	34
4.	Two-Hand: Mean	.29	.16	53.9	13.1	34
5.	Mashburn: Through trial 39	.30	.13	281.6	43.7	34
6.	Eye-Hand: Pattern A	-.32	.15	58.0	15.5	34
7.	Eye-Hand: Pattern B	-.33	.15	36.2	9.6	34
8.	Eye-Hand: Pattern D	-.33	.15	35.9	7.1	34
9.	M.C. (Rights minus one-half number wrong)	--	--	--	--	(No Cases)
10.	M.C. (Number right)	--	--	--	--	(No Cases)
11.	Personal History Inventory	--	--	--	--	(No Cases)
12.	Aviation Interview VII	--	--	--	--	(No Cases)
13.	Aviation Interview VIII	--	--	--	--	(No Cases)
14.	Aviation Interview IX	--	--	--	--	(No Cases)
15.	Biographical Inventory	--	--	--	--	(No Cases)
16.	Otis Test	--	--	--	--	(No Cases)
17.	Aviation Information	--	--	--	--	(No Cases)
18.	Body Surface	-.17	.17	187.7	9.2	34
19.	Vital Capacity	-.24	.16	48.4	7.6	34
20.	Tidal Air	-.41	.14	9.2	4.6	34
21.	Tidal Air/Body Surface	-.38	.15	50.9	23.7	34
22.	Vital Capacity/Body Surface	-.19	.17	25.5	3.7	34
23.	Pulse Rate	-.22	.16	77.9	12.1	34

Evaluation of Predictors in Terms of Correlations with Composite Criteria.

On the basis of their relationships with both dichotomous, and continuous, composite criteria it is evident that in general the psychomotor tests tend to predict the composite criteria most efficiently with the Mashburn Test showing the highest relationships with the dichotomous composite, and the Mashburn Test and Pattern D of the Eye-Hand Coordination Test yielding the highest coefficients with the continuous composite. Following the psychomotor tests, the Test of Mechanical Comprehension is probably next in rank of efficiency, in terms of either dichotomous or continuous composite, with the physiological measures and the interview being less efficient. The Biographical Inventory, the Otis Test, and the Personal History Inventory show little or no correlation with either composite criteria.

DEVELOPMENT OF PREDICTION BATTERIES

In developing prediction batteries, it is of importance to examine the intercorrelations among the various predictors as well as their correlations with the criteria.

Intercorrelations of Physiological Measures.

Intercorrelations among the complete list of physiological measures[50] are available on the Boston Fall group only and are presented in Table 41, Appendix A. It will be noted that the coefficients are in general low, the only sizable correlations arising when the two correlated variables have a common element, e.g., ratios such as Tidal Air over Body Surface vs. Tidal Air, or Pulse Rate (lying) vs. Maximum Pulse Rate.

Data on the intercorrelations among selected physiological measures, in general those showing highest reliability, are available from all four samples. These intercorrelations are presented in Tables 42 to 45, Appendix A.[51] Examination of these tables indicates that while considerable relationship is evident between Body Surface and Vital Capacity (.36 to .63), the only other high correlations are between variables which have a common element, and may therefore be regarded with some question. However, the intercorrelations, for the most part, are positive. The variable Pulse Rate, for which data are available from the Midwest Project only (Tables 44 and 45, Appendix A), shows no high correlation with the other variables presented in the tables. The only variables with which

---

[50] With the exception of the physiological measure, Body Surface.

[51] It should be noted that the coefficients and N's in Table 42, Appendix A, are not in exact agreement with the coefficients and N's presented in Table 41, Appendix A, although both tables present data from the Boston Fall group. For the smaller number of variables a larger sample of subjects was available, thus affecting the coefficients to a slight degree.

Pulse Rate (lying) correlates to any degree on both of the Midwest samples are Tidal Air and Tital Air/Body Surface, the coefficients being in the neighborhood of .25. However, reference to Table 41, Appendix A, shows that on the Boston Fall sample the correlations between Pulse Rate and each of these variables were also low, the only sizable coefficients involving Pulse Rate resulting when this measure was correlated with some other Pulse Rate variable.

Intercorrelations of Psychomotor Measures.

Intercorrelations among the scores on the Two-Hand Coordination Test and on the Mashburn Test, for the Boston Fall sample, are presented in Table 46, Appendix A. The correlations among trials on the Two-Hand Coordination Test are relatively high, varying between .37 and .74, over half of the coefficients being greater than .50. The intercorrelations among Mashburn Test scores are considerably higher than those of the Two-Hand Coordination Test scores, varying between .70 and .98, with more than half of these coefficients greater than .90. The high intercorrelations among scores on trials 13, 26, 39 and 40 are due, in large part, to the fact that the measures are time scores computed cumulatively, and thus were not independently obtained. The intercorrelations among independently computed Mashburn scores (Table 46, Items 10, 14, and 15), however, are also rather high, ranging from .70 to .77.

In general, the mean score and highest score on the Two-Hand Coordination Test show the highest correlation with other Two-Hand Test scores. Further, these measures correlate higher with the Mashburn Test scores than do the other Two-Hand measures. It should be noted in this connection that these Two-Hand measures were also the most reliable. The negative correlations between Two-Hand Coordination Test scores and Mashburn Test scores result from the fact that on the Two-Hand Test, a high score indicates "good" performance, while on the Mashburn Test, a high score indicates "poor" performance.

In Table 47, Appendix A, are presented intercorrelations (for the Boston Spring group) among the mean score for the Two-Hand Coordination Test and the Mashburn Test scores. In addition, intercorrelations among scores for the Eye-Hand Coordination Test are given in this table.[52] Examination of this table indicates that the correlations between Two-Hand Test scores and Mashburn Test scores are comparable, in relative terms, to those on the Fall sample discussed above, although the coefficients are in general somewhat higher. The intercorrelations among the Eye-Hand Test scores are relatively high, but these scores show relatively low intercorrelations with the Two-Hand Test scores and Mashburn Test scores, the coefficients ranging from .19 to .34.

---

[52]The Eye-Hand Coordination Test was not administered to the Boston Fall group.

Intercorrelations between two Two-Hand Coordination Test scores, three Mashburn Test scores, and the Eye-Hand Coordination Test scores on Patterns A, B, and D are presented for the Midwest Spring and Summer samples, respectively, in Tables 48 and 49, Appendix A. Also included in these tables are the correlations between the above psychomotor scores and selected physiological measures. It will be noted that the correlations with the Mashburn Test scores presented are generally in line with the coefficients obtained from the Boston samples. The correlations between Two-Hand Coordination Test mean score and the Two-Hand Test score on trial 6 are in the neighborhood of .80, as they were for the Boston Fall sample. Similarly, in both the Midwest Spring and Summer groups, the Eye-Hand Coordination Test scores on Patterns A, B, and D are relatively highly intercorrelated, but show lower correlations with other psychomotor measures, being in general below .30.[53]

Summarizing the several psychomotor scores as predictors of flight proficiency, it can be stated that, in general, the measures are positively related, the negative coefficients resulting, for the most part, from differences in scoring procedure on the various tests, i.e., from the fact that on the Two-Hand Coordination Test a high score denoted "good" performance, while on the Mashburn Test and Eye-Hand Coordination Test a high score denoted "poor" performance. Various scores on the same tests are relatively highly correlated, the correlations between scores on different tests being in general lower. None of the correlations between scores on different tests appears sufficiently high as to constitute prima facie evidence that it would be uneconomical to include the various tests in a single test battery.

Correlations Between Psychomotor and Physiological Measures.

Correlations between psychomotor and physiological measures, for the Midwest Spring and Summer groups respectively, are also presented in Tables 48 and 49, Appendix A. Inspection of these tables indicates that the coefficients tend to be low. There is some hint in the Midwest Summer sample of slight relationship between the Two-Hand Coordination Test scores and physiological measures involving Tidal Air, the coefficients varying between -.29 and -.34. This trend, however, is not substantiated by data from the Spring sample.

Intercorrelations Among Paper-and-Pencil Tests.

In Tables 50 and 51, Appendix A, are presented intercorrelations among Biographical Inventory scores, scores on the Test of Mechanical

---

[53] On the basis of the Pensacola data, the correlations of the Eye-Hand Coordination Test with the Two-Hand Coordination Test and the Mashburn Test, respectively, were .22 and .21. See: McFarland, R. A., and Franzen, R. Op. cit. (Referred to in Footnote 1.)

Comprehension, and Otis Test scores for the Boston Fall and Spring groups, respectively. In both samples the intercorrelations among the four B.I. scores are high (.77 to .94). It is to be noted that the correlation between "plus" and "plus and minus" scoring procedures[54] are in all cases greater than .80. Furthermore, the correlations between scores on Part A and scores on Parts A plus B are high.

The correlations between Biographical Inventory scores and scores on the Test of Mechanical Comprehension are moderately high (.29 to .43), while low correlations (-.07 to .17) are evident between the former and the Otis Test. The Test of Mechanical Comprehension and the Otis Test correlate in the neighborhood of .45 on both Boston samples.

In Tables 52 and 53, Appendix A, are presented intercorrelations among scores of the Biographical Inventory, the Otis Test, and the Test of Mechanical Comprehension for the Midwest Spring and Summer samples. In Table 53, Appendix A, the Test of Aviation Information is also included, this test being administered only to the Summer group. It will be noted that the data presented from the Midwest Project include only Biographical Inventory scores in terms of the "plus" scoring procedures, but that the score on Part B of the test is included.

Examination of Tables 52 and 53, Appendix A, indicates that score on Part A and score on Parts A plus B are highly correlated (.90 and .88). This substantiates the finding from the Boston sample, although it should be noted that the two measures are not independent, score on Part A contributing to the score on Parts A plus B. The correlation between Part A and Part B of the B.I. is low, on both Spring and Summer groups (.02 and .11). The correlations between B.I. (scores on Part A, and Parts A plus B) and the Otis Test are relatively low, as was indicated in the Boston samples, although the correlations between the Biographical Inventory and the Test of Mechanical Comprehension are considerably lower than on the Boston samples, particularly when the Midwest Summer group is considered. While for the Midwest Spring group the correlation between the Otis Test and the Test of Mechanical Comprehension is relatively high (.33), in the Boston groups for the Midwest Summer group the correlation between these tests drops to .01.

Intercorrelations of the scores on the Strong Vocational Interest Inventory, and the correlations between Strong Vocational Interest Inventory scores and other measures are available only for the Boston Fall group. Intercorrelations among Strong Vocational Interest Inven-

---

[54] In the "plus" scoring procedure, only positively weighted items were considered in arriving at the total score. In the "plus and minus" procedure, positive and negative weights were added algebraicly. The expression "1%" indicates that items at the .01 level of significance were assigned unit weights.

tory scores, and between these scores and the Otis Test, the Test of Mechanical Comprehension, the Two-Hand Coordination Test, and a number of scores from the Washburn Test are presented in Table 54, Appendix A.

Examination of this table indicates that the usual high positive correlations are found between similar occupations, (e.g., chemist and engineer) and high negative correlations are evident between dissimilar occupations, e.g., engineer and life insurance salesman. The Aviator score shows relatively high correlations with engineer, chemist, and math-science teacher, and a relatively high negative correlation with life insurance salesman.

The correlations between the Aviator score and the psychomotor tests are not high (.13 to .23), but the coefficients are positive. There appears to be a general tendency for these coefficients of correlation to be higher than the coefficients between psychomotor tests and Strong Vocational Interest Inventory scores for other occupations.

It should further be noted that, in this sample, the correlations between the Test of Mechanical Comprehension and the psychomotor tests are higher than are the correlations between the Otis Test and the psychomotor tests. The highest correlation is .53, between the Test of Mechanical Comprehension and the Two-Hand Coordination Test mean score.

In Table 55, Appendix A, are presented the correlations of paper-and-pencil tests with physiological measures and psychomotor test scores for the Midwest Summer group. Although the sample on which the correlations involving the physiological measures are based is small, certain relationships are suggestive. The correlation between scores on the Biographical Inventory, Parts A plus B, and Vital Capacity is -.53 and the correlation with Vital Capacity/Body Surface is -.55. These Vital Capacity measures correlated positively with the Otis Test, the coefficients being .37 and .36 respectively. The correlation between the Otis Test and Pulse Rate (lying) is relatively high (.55), as is the correlation between the Test of Aviation Information and Pulse Rate (53).

Although in general the coefficients are not high, it may also be noteworthy that Tidal Air and Tidal Air/Body Surface are positively related to the Biographical Inventory score on Part A (the coefficients being .22 and .23 respectively), while these physiological measures correlated negatively with the score on Part B of this test, the coefficients in both cases being-.32. Measures involving Tidal Air also correlate positively with the Test of Aviation Information, the coefficients being .32 and .34. Additional data from other samples bearing on these relationships would be of considerable interest.

The correlations between the paper-and-pencil tests and the psychomotor measures are based on a larger sample. Scores on the Two-Hand Coordination Test are all positively related to the paper-and-pencil

tests, while scores on the Mashburn Test and Eye-Hand Coordination Test are negatively correlated with the paper-and-pencil tests. On these latter psychomotor tests, however, a high score denotes "poor" performance.

The correlations of the paper-and-pencil tests with the Two-Hand Coordination Test scores and the Mashburn Test scores are in general between .10 and .30. On this sample, the trend shown in the Boston sample, i.e., higher correlations between the Test of Mechanical Comprehension and the Two-Hand or Mashburn Tests than between the Otis Test and these psychomotor measures, does not appear. The Test of Mechanical Comprehension, however, does correlate markedly higher with the Eye-Hand Coordination Test scores on all three patterns than do the other paper-and-pencil tests, the coefficients varying between -.51 and -.55. The Eye-Hand Coordination Test, Patterns A and D, correlate relatively high with the Test of Aviation Information, the coefficients being -.36 and -.35 respectively. As noted above, these relationships can be considered only suggestive, pending analysis of data from other samples.

Table 56, Appendix A, presents intercorrelations among the Aviation Interview scales, and correlations between the Interview scales and the Biographical Inventory scores, the Otis Test, and the Test of Mechanical Comprehension for the Boston Spring group. The intercorrelations among Interview scales are positive, and in general are relatively high, particularly among Scales C, G, H, and I.[55] Correlations between the paper-and-pencil tests and the Interview Scales are in general positive, and vary from -.04 (the only negative coefficient) between the Test of Mechanical Comprehension and Scale H (appearance and mannerisms) to .52 between the Biographical Inventory and Scale E (Desire to Fly). In general, Scales D and E (Desire to Fly and Hobbies) correlate somewhat higher with the paper-and-pencil tests than does Scale I (Fitness for Flight Training). Finally, the Biographical Inventory scores tend to correlate higher with scores on the Interview Scales than do the Otis Test or the Test of Mechanical Comprehension.

In Table 57, Appendix A, are presented the intercorrelations for the Midwest Spring sample, among the Aviation Interview scores, and the correlations between Aviation Interview scores and the paper-and-pencil tests. These correlations are in agreement with those from the Boston sample to the extent that Interview Scales D and E in general correlate somewhat higher with the paper-and-pencil tests than does Interview Scale I, and to the degree that the Biographical Inventory Part A correlates higher with the Interview Scales than do other paper-and-pencil tests. This trend is not as marked in the Midwest as in the Boston sample, however. Despite the agreement in general trends, many of the coefficients from the Midwest sample differ markedly from their counterparts on the Boston

---

[55]These scales are identified as follows: C (General Social Adjustment), G (Personality), H (Appearance and Mannerisms), I (Fitness for Flight Training).

sample, particularly noteworthy being the low but negative correlations of the Biographical Inventory, Part A, and Parts A plus B, with the Interview Scale F (Athletic Activities). This negative relationship is particularly odd because of the fact that the Biographical Inventory, Part A, is heavily loaded with items pertaining to athletic activities. It should further be noted that the score on Part B[56] of the Biographical Inventory correlates zero, or extremely low, with the Interview Scales.

The intercorrelations among the Interview Scales are perhaps somewhat higher on the Midwest than on the Boston samples. However, as on the Boston sample, the intercorrelations involving Scales C, G, H, and I are in general the highest.

Selection of Variables for Predictor Battery.

In view of their reliability, their predictor-criteria correlations, and their interrelationships, the following variables were selected for inclusion in a number of predictor batteries: Two-Hand Coordination Test, trial 6; the Two-Hand Coordination Test, mean score; the Mashburn Test, through trial 39; the Eye-Hand Coordination Test, Patterns A, B, and D; and the Test of Mechanical Comprehension. The variable Tidal Air was also included in one battery to determine the contribution of a physiological variable. The intercorrelations of these variables in terms of the cases actually used in the various comparison between batteries and dichotomous and continuous composite criteria are presented in Tables 29 to 32. It will be noted that the intercorrelations based on these cases are in general similar to the intercorrelations presented in Appendix A and discussed above. The high intercorrelation between the Two-Hand Coordination Test, mean score, and trial 6 score for the same test, however, indicates that these variables should be included in different batteries. The correlation of .99 between the two Mechanical Comprehension Test scores (Rights, and Rights minus one-half wrongs, respectively) indicates that it would make little difference which score were used and the "Rights" score was actually used in subsequent computations.

Pattern D of the Eye-Hand Coordination Test was selected for inclusion with other tests in the batteries since it was the best predictor, and did not correlate higher with the Two-Hand Test and the Mashburn Test than did the other Eye-Hand patterns. However, all three Eye-Hand patterns were included in one battery to determine how well such an apparently simple and economical battery functioned.

---

[56] This variable was not used in the comparisons from the Boston sample.

TABLE 29

INTERCORRELATIONS OF VARIABLES FOR CASES USED AGAINST THREE-FOLD COMPOSITE
MIDWEST PROJECT (Spring Group)

Psychomotor and Paper-and-Pencil Tests	1	2	3	4	5	6	7	8	9	10
1. Two-Hand: Trial 6	—	.77	-.24	-.08	-.22	-.45	.34	.34	.07	.05
2. Two-Handr Mean		—	-.33	-.36	-.30	-.51	.45	.46	.21	.17
3. Eye-Hand: Pattern A			—	.63	.57	.38	-.22	-.21	.00	.05
4. Eye-Hand: Pattern B				—	.62	.40	-.19	-.20	-.15	-.20
5. Eye-Hand: Pattern D					—	.47	-.31	-.32	-.15	.07
6. Mashburn, through trial 39						—	-.32	-.32	-.13	-.16
7. M.C. (Rights minus one-half number wrong)							—	.99	.48	.17
8. M.C. (Number right)								—	.50	.16
9. Personal History Inventory									—	.08
10. Otis Test										—
Mean	60.2	54.1	57.7	38.6	43.1	296.6	40.1	51.8	24.9	51.8
Sigma	13.5	10.7	9.3	10.2	7.8	51.2	10.2	6.8	7.2	9.1
N = 39										

Psychological Tests	1	2	3	4	5	6
1. Body Surface (BS)	—	.62	.00	-.22	.08	-.11
2. Vital Capacity (VC)		—	.02	-.13	.82	-.09
3. Tidal Air (TA)			—	.97	.03	.15
4. TA/BS				—	.00	.17
5. VC/BS					—	-.03
6. Pulse Rate						—
Mean	137.0	49.5	7.9	44.4	26.2	72.9
Sigma	12.1	5.6	2.6	13.9	2.3	11.8
N = 33						

Aviation Interview	1	2	3
1. Interview VII	—	.91	.89
2. Interview VIII		—	.82
3. Interview IX			—
Mean	15.3	16.0	13.2
Sigma	3.9	4.0	5.3
N = 35			

TABLE 30

INTERCORRELATIONS OF VARIABLES FOR CASES USED AGAINST FOUR-FOLD COMPOSITE
MIDWEST PROJECT (Spring Group)

Psychomotor and Paper-and-Pencil Tests	1	2	3	4	5	6	7	8	9	10
1. Two-Hand: Trial 6	--	.81	-.28	-.09	-.26	-.45	.41	.42	.07	-.01
2. Two-Hand: Mean		--	-.30	-.33	-.26	-.51	.45	.45	.20	.20
3. Eye-Hand: Pattern A			--	.61	.54	.48	-.30	-.30	.02	.03
4. Eye-Hand: Pattern B				--	.59	.45	-.19	-.20	-.11	-.31
5 Eye-Hand: Pattern D					--	.51	-.33	-.34	-.15	.04
6. Washburn, through trial 39						--	-.40	-.40	-.12	-.05
7. M.C. (Rights minus one-half number wrong)							--	.99	.52	.25
8. M.C. (Number right)								--	.54	.25
9. Personal History Inventory									--	.07
10. Otis Test										--
Mean	60.2	53.9	58.3	39.5	43.8	298.8	40.1	51.8	24.5	51.2
Sigma	14.7	11.2	7.9	10.2	7.7	51.8	10.0	6.6	7.7	8.9
N = 32										

Psychological Tests	1	2	3	4	5	6
1. Body Surface (BS)	--	.62	.02	-.18	.08	-.02
2. Vital Capacity (VC)		--	-.04	-.18	.83	.02
3. Tidal Air (TA)			--	.97	-.06	.10
4. TA/BS				--	-.10	.12
5. VC/BS					--	.03
6. Pulse Rate						--
Mean	186.6	49.7	8.0	45.4	26.3	73.7
Sigma	11.7	5.4	2.7	14.5	2.2	11.1
N = 27						

Aviation Interview	1	2	3
1. Interview VII	--	.91	.87
2. Interview VIII		--	.81
3. Interview IX			--
Mean	15.2	16.0	13.0
Sigma	3.7	4.0	5.4
N = 29			

TABLE 31

INTERCORRELATIONS OF VARIABLES FOR CASES USED AGAINST THREE-FOLD COMPOSITE
MIDWEST PROJECT (Summer Group)

		1	2	3	4	5	6	7	8	9	10	11	12
1.	Two-Hand: Trial 6	—	.87	-.23	-.23	-.35	-.54	-.05	-.07	-.44	-.44	-.03	.10
2.	Two-Hand: Mean		—	-.12	-.15	-.30	-.56	-.02	.08	-.42	-.41	.71	.12
3.	Eye-Hand: Pattern A			—	.75	.80	.30	-.09	.01	.16	.14	.02	.11
4.	Eye-Hand: Pattern B				—	.63	.44	.01	.29	.32	.31	.31	-.06
5.	Eye-Hand: Pattern D					—	.38	-.05	-.07	.25	.22	-.08	-.15
6.	Mashburn, through trial 39						—	.24	.39	.37	.36	.34	.02
7.	Body Surface (BS)							—	.47	.16	.06	.19	-.11
8.	Vital Capacity (VC)								—	.28	.25	.06	.19
9.	Tidal Air (TA)									—	.99	.27	.13
10.	TA/BS										—	.27	.12
11.	VC/BS											—	.27
12.	Pulse Rate												—
	Mean	60.2	53.9	58.0	36.2	35.9	281.6	187.7	43.4	9.2	50.9	25.5	77.9
	Sigma	15.0	13.1	15.5	9.6	7.1	43.7	9.2	7.6	4.6	23.7	3.7	12.1
	N = 34												

TABLE 32

INTERCORRELATIONS OF VARIABLES FOR CASES USED AGAINST FOUR-FOLD COMPOSITE
MIDWEST PROJECT (Summer Group)

		1	2	3	4	5	6	7	8	9	10	11	12
1.	Two Hand: Trial 6	--	.87	-.21	-.22	-.35	-.54	-.07	.06	-.45	-.44	.03	.11
2.	Two-Hand: Mean		--	-.11	-.14	-.30	-.56	-.03	.08	-.42	-.41	.72	.12
3	Eye-Hand: Pattern A			--	.72	.81	.26	-.05	-.01	.17	.15	-.02	-.15
4.	Eye-Hand: Pattern B				--	.62	.42	.06	.29	.35	.33	.29	.09
5	Eye Hand: Pattern D					--	.37	-.04	.08	.25	.22	-.10	-.17
6.	Washburn, through trial 39						--	.27	.38	.38	.37	.33	.00
7	Body Surface (BS)							--	.48	.16	.06	.21	-.10
8	Vital Capacity (VC)								--	.28	.25	.95	.19
9.	Tidal Air (TA)									--	.99	.27	.12
10	TA BS										--	.27	.13
11	VC/BS											--	.26
12	Pulse Rate												--
	Mean	60.4	54.0	57.1	35.6	35.7	280.2	187.9	48.3	9.2	50.9	25.4	77.8
	Sigma	15.2	13.3	14.8	9.2	7.1	43.6	9.2	7.7	4.6	24.1	3.8	12.2
	N = 33												

Prediction of Dichotomous Composite

The following test batteries were used in the prediction of the dichotomous composite criteria:

Battery 1: Two-Hand Coordination Test, trial 5
Mashburn Serial Reaction Test, time through trial 39
Eye-Hand Coordination Test, Pattern C
Test of Mechanical Comprehension

Battery 2: Two-Hand Coordination Test, mean score
Mashburn Serial Reaction Test, time through trial 39
Eye-Hand Coordination Test, Pattern D
Test of Mechanical Comprehension

Battery 3: Eye-Hand Coordination Test, Pattern A
Eye-Hand Coordination Test, Pattern B
Eye-Hand Coordination Test, Pattern D

Battery 4: Two-Hand Coordination Test, trial 5
Mashburn Serial Reaction Test, time through trial 39
Eye-Hand Coordination Test, Pattern D

Battery 5: Two-Hand Coordination Test, mean score
Mashburn Serial Reaction Test, time through trial 39
Eye-Hand Coordination Test, Pattern D

The multiple correlation coefficients of these batteries with the four-fold and three-fold dichotomous composite criteria, hurdle and compensation types, are summarized in Table 33. The regression weights computed on one sample (e.g., the Spring Group) were also applied to the other sample (e.g., the Summer Group) and multiple coefficients computed using weights derived on the independent sample. Thus in Table 33, under the heading Spring Weights the coefficients in the row Summer cases represent the coefficient based on a sample other than that on which the regression weights were derived.[57]

In evaluating these coefficients it should be kept in mind that the number of cases is small, resulting in a standard error of an r of zero being in the neighborhood of .25, and that the multiple coefficients are undoubtedly inflated due to the use of biserial coefficients in computing

---

[57]Coefficients for Batteries 1 and 2 could not be computed on the Summer data, since data on the Test of Mechanical Comprehension were used in this analysis.

TABLE 33

MULTIPLE BISERIAL COEFFICIENTS: TEST BATTERIES VS. DICHOTOMOUS COMPOSITE CRITERIA
MIDWEST PROJECT

No. of Cases	Failures	Failures		Battery 1		Battery 2		Battery 3		Battery 4		Battery 5	
Total	Comp.	Hurd.						Four-fold Dichotomous Composite					
26	6	8											
41	16	18											
35	5	12		Comp.	Hurd.	Comp.	Hurd.	Comp.	Hurd.	Comp.	Hurd.	Comp.	Hurd.
Spring Heights			26										
Spring cases			26	.59	.83	.96	.73	.71	.77	.98	.83	.94	.72
" "			41	.72	.62	.71	.61	.51	.60	.72	.62	.73	.62
Summer cases			35					.29	- .03	.70	.43	.72	.46
Spring Heights			41										
Spring cases			41	.76	.67	.76	.66	.52	.60	.76	.67	.76	.65
" "			26	.74	.78	.90	.67	.71	.77	.95	.77	.92	.69
Summer cases			35					.16	.01	.91	.44	.83	.42
Summer Heights			35	N_{total}	$N_{failers\ A}$	$N_{failers\ B}$							
Summer cases			35	26	6	8		.61	.67	.88	.47	.90	.47
Spring cases			41	41	16	18		.38	.21	.86	.75	.83	.71
" "			26	35	5	12		.30	- .05	.71	.62	.71	.59
								Three-fold Dichotomous Composite					
Spring Heights			30										
Spring cases			30	.72	.82	.71	.78	.46	.54	.71	.82	.71	.78
" "			47**	.64	.65	.65	.65	.38	.46	.65	.65	.65	.64
Summer cases			36					.69	.51	.95	.69	.88	.66
Spring Heights			47**										
Spring cases			47**	.68	.70	.69	.70	.39	.47	.68	.69	.68	.69
" "			30	.70	.78	.69	.75	.45	.54	.69	.78	.68	.76
Summer cases			36					.57	.25	.93	.74	.93	.74
Summer Heights			36	N_{total}	$N_{failers\ A'}$	$N_{failers\ B'}$							
Summer cases			36	30	7	8		.78	.81	.96	.77	.96	.77
Spring cases			30	47	16	19		.42	.37	.70	.74	.67	.73
" "			47**	36	5	7		.33	.23	.65	.66	.66	.66

*These cases include the N of 26, as well as all students failing the [illegible]

predictor-criteria zero-order correlations.[58] However, it is noteworthy that while the coefficients are remarkably high (maximum .99), the correlations in general hold up when the regression weights are applied to an independent sample, particularly on the basis of the compensation criteria. While these coefficients cannot be interpreted at their face value, they are of interest in demonstrating the relative "predictability" of the two types of composite, and in indicating the relative predictive value of the various batteries.

The compensation criteria consistently obtain higher coefficients, except as predicted by the Eye-Hand Battery (Battery 3). The Eye-Hand Battery is apparently the least predictive. Comparison of Batteries 1 and 4, and 2 and 5, which differ only in the inclusion of the Test of Mechanical Comprehension in Batteries 1 and 2, indicates that this test contributes little to the batteries.

In Tables 34 and 35 are given the regression coefficients computed from the four-fold and the three-fold composite criteria, respectively. Examination of these tables indicates that, in general, the Mashburn Test contributes the most to the various batteries, and the Test of Mechanical Comprehension, the least.

Prediction of Continuous Composite

The batteries used in predicting the continuous composite were the same as those used in connection with the dichotomous composites except for the elimination of the Eye-Hand Battery (Battery 3), and the addition of a battery consisting of: Eye-Hand Coordination Test, trial 6; Mashburn Serial Reaction Test, time through trial 39; Eye-Hand Coordination Test, Pattern D; and Tidal Air. The multiple correlations between Batteries 1, 2, 4, and 5 and the four-fold and three-fold continuous composite criteria, respectively, are presented in Tables 36 to 38 for the Spring and Summer groups. In these tables the multiple correlation coefficients from the hurdle and compensation type dichotomous composite criteria are also presented for comparison purposes. It should be noted that the multiple coefficients for these batteries against the continuous composites range, in general, between .50 and .60, and that

---

[58] The spurious nature of these multiple coefficients probably results from the following situation. The biserial predictor-criterion correlations are undoubtedly spuriously high, as is indicated by the fact that the Pearsonian predictor-criterion correlations computed against a continuous composite are markedly lower (compare Tables 23 to 24 with Tables 25 to 28), and the drop cannot be entirely accounted for by the substitution of Flight Scores for Camera Criterion V in the continuous composite, since these two variables intercorrelate highly. Furthermore, there is no reason to suspect that the Pearson r's are spuriously low. In determining the multiple correlation, while the predictor-criterion correlations were spuriously high, the intercorrelations between predictors were Pearsonians, and thus markedly lower, and not spurious. This combination of high predictor-criterion coefficients, and relatively low intercorrelations between predictors, could be expected to result in a marked inflation of the multiple correlation coefficients.

TABLE 34

REGRESSION COEFFICIENTS: TEST BATTERIES VS. FOUR-FOLD DICHOTOMOUS COMPOSITE CRITERIA
MIDWEST PROJECT

			Battery 1		Battery 2		Battery 3		Battery 4		Battery 5	
Sample	N	Test	Comp.	Hurd.	Comp.	Hurd.	Comp.	Hurd.	Comp.	Hurd.	Comp.	Hurd.
Spring	26	Two-Hand: 6	.3865	.4752					.3629	.4708		
Spring	41**		.0677	.1620					.0812	.1741		
Summer	35								-.0784	.0883		
Spring	26	Two-Hand: Mean			.2675	.1817					.2106	.1473
Spring	41**				-.0951	-.0645					-.0247	-.0134
Summer	35										-.2842	.1127
Spring	26	Mash. 39*	-.5767	-.1138	-.6630	-.3273			-.5322	-.1347	-.6212	-.3020
Spring	41**		-.5888	-.3804	-.6276	-.4490			-.5924	-.3837	-.6367	-.4620
Summer	35								-.9234	-.1083	-1.0273	-.0930
Spring	26	Eye-Hand: A*					-.1304	.0499				
Spring	41**						-.0262	.0255				
Summer	35						-.1270	-.6246				
Spring	26	Eye-Hand: B*					.2057	.5768				
Spring	41**						.1872	.4436				
Summer	35						-.5409	-.2890				
Spring	26	Eye-Hand: C*	-.3485	-.5121	-.3213	-.4443	-.7625	-1.0685	-.3345	-.4603	-.3097	-.4373
Spring	41**		-.2493	-.3270	-.2440	-.3131	-.6296	-.8505	-.2523	-.3297	-.2480	-.3189
Summer	35						.0517	.2538	.0257	-.3581	.0023	-.3605
Spring	26	M. C.	-.1538	-.0210	-.1795	-.1086						
Spring	41**		.0317	.0285	.0823	.1184						

*Negative sign should be interpreted as positive.
**These N cases include the N of 26, as well as all students failing the course.

TABLE 35

REGRESSION COEFFICIENTS: TEST BATTERIES VS. THREE-FOLD DICHTOMOUS COMPOSITE CRITERIA
MIDWEST PROJECT

			Battery 1		Battery 2		Battery 3		Battery 4		Battery 5	
Sample	N	Test	Comp.	Hand.	Comp.	Hand.	Comp.	Hand.	Comp.	Hand.	Comp.	Hand.
Spring	30	Two-Hand: G	.1535	.3338					.1724	.3419		
Spring	47**		-.0567	.0457					-.0044	.0718		
Summer	36								.1393	-.0363		
Spring	30	Two-Hand: Mean			.1149	.1814					.1702	
Spring	47**				-.1161	-.0640					-.0502	
Summer	36										-.1133	
Spring	30	Mash. 39*	-.4775	-.4341	-.5122	-.5325			-.4839	-.3973	-.5123	
Spring	47**		-.6112	-.5541	-.6307	-.5903			-.6299	-.5594	-.6397	
Summer	36								-.6393	-.4039	-.7594	
Spring	30	Eye-Hand: A*					-.0291	.0094				
Spring	47**						-.0913	.0994				
Summer	36						.1239	-.6166				
Spring	30	Eye-Hand: B*					.0439	.1161				
Spring	47**						.1664	.1479				
Summer	36						-.4773	-.2152				
Spring	30	Eye-Hand: D*	-.1502	-.2472	-.1322	-.2129	-.4707	-.6265	-.2267	-.3772	-.1430	
Spring	47**		-.1117	-.1965	-.1231	-.1949	-.4430	-.6117	-.1263	-.2038	-.1306	
Summer	36						-.4953	-.0245	-.3731	-.4228	-.4155	
Spring	30	M. C.	.1492	.0336	.1293	.0110						
Spring	47**		.1299	.0650	.1522	.1057						

*Negative sign should be interpreted as positive.
**These 47 cases include the N of 26, as well as all students failing the course.

TABLE 36

MULTIPLE CORRELATIONS; PREDICTOR BATTERIES 1, 2, AND 3 TYPES OF COMPOSITE CRITERIA
MIDWEST PROJECT BATTERY 1

(Two-Hand Coordination Test, trial 6; Mashburn Test, time through trial 39; Eye-Hand Coordination Test, Pattern D; Test of Mechanical Comprehension)

FOUR-FOLD COMPOSITE	N	% Passing	R	S.E.$_R$	S.E.$_{R \text{ of } 0}$
Compensation	26	77	.39*	.079	.272
Hurdle	26	69	.83*	.122	.257
Continuous	22		.57	.119	.177
THREE-FOLD COMPOSITE					
Compensation	30	77	.52*	.158	.252
Hurdle	30	73	.52*	.123	.246
Continuous	29		.60	.102	.160

BATTERY 2

(Two-Hand Coordination Test, trial 6; Mashburn Test, time through trial 39; Eye-Hand Coordination Test, Pattern D; Test of Mechanical Comprehension)

FOUR-FOLD COMPOSITE					
Compensation	[illegible]	77	.06*	.091	.272
Hurdle	[illegible]	69	.[illegible]*	.153	.257
Continuous	[illegible]		.[illegible]	.121	.177
THREE-FOLD COMPOSITE					
Compensation	[illegible]	77	.[illegible]*	.160	.252
Hurdle	[illegible]	73	.[illegible]*	.135	.246
Continuous	[illegible]		.[illegible]	.101	.160

*[illegible] correlation [illegible] correlation.

MULTIPLE CORRELATIONS: PROJECTED BATTERY A AND 3 TYPES OF COMPOSITE CRITERIA
[illegible] PROJECT

BATTERY A
(Two-Hand Coordination Test, trial 6, Washburn Test, time through trial 39, Eye-Hand Coordination Test, Pattern D)

	N	% Passing	R	S.E.R	S.E.R of 0
FOUR-FOLD COMPOSITE					
Compensation					
Spring	26	[illegible]	.99*	.083	.272
Summer	35	[illegible]	.88*	.131	.262
Hurdle					
Spring	26	[illegible]	.85*	.122	.257
Summer	35	[illegible]	.47*	.181	.218
Continuous					
Spring	32		.55	.123	.177
Summer	33		.50	.131	.174
THREE-FOLD COMPOSITE					
Compensation					
Spring	30	77	.71*	.160	.252
Summer	36	86	.96*	.107	.260
Hurdle					
Spring	30	77	.82*	.123	.246
Summer	36	81	.77*	.141	.240
Continuous					
Spring	39		.59	.104	.160
Summer	34		.53	.123	.171

*R computed from biserial coefficients of correlation.

TABLE 38

MULTIPLE CORRELATIONS: PREDICTOR BATTERY 5 AND 3 TYPES OF COMPOSITE CRITERIA
MIDWEST PROJECT

BATTERY 5
(Two-Hand Coordination Test, Mean score; Washburn Test, time through trial 39; Eye-Hand Coordination Test, Pattern D)

FOUR-FOLD COMPOSITE	N	% Passing	R	$S.E._R$	$S.E._R$ of 0
Compensation					
Spring	26	77	.94*	.098	.272
Summer	25	81	.90*	.125	.262
Hurdle					
Spring	26	59	.72*	.156	.257
Summer	25	46	.47*	.181	.218
Continuous					
Spring	27		.59	.127	.177
Summer	23		.53	.132	.174
THREE-FOLD COMPOSITE					
Compensation					
Spring	[illegible]	[illegible]	[illegible]	[illegible]	.252
Summer	[illegible]	[illegible]	[illegible]	.107	.259
Hurdle					
Spring	[illegible]	[illegible]	[illegible]	[illegible]	.246
Summer	[illegible]	[illegible]	.7[illegible]	[illegible]	.246
Continuous					
Spring	[illegible]		[illegible]	[illegible]	.255
Summer	[illegible]		[illegible]	[illegible]	.175

* Computed from [illegible]

these coefficients are markedly lower than were the coefficients computed for the dichotomous composites.[59]

In Tables 39 and 40 are summarized the increments in prediction obtained as successive tests are added to form various batteries. Table 39 shows these relationships for Battery 1 with the four-fold and three-fold continuous composites, respectively, from data based on the Spring Group. Table 40, based on Summer group data, presents results from a comparable treatment of other combinations of tests. In general, the multiple R was practically maximized after the two best measures had been combined. In fact, in most cases, the predictive value of the battery was little better than the predictive value of the best single test in the battery. These relationships may be summarized as follows:

The increments of the highest multiple R obtained over the highest zero order r for the various groups are as follows:

1. Spring Group, Four-fold criterion: from .46 (Mashburn Test, time through trial 39) to .57 (Table 39).[60]

2. Spring Group, Three-fold criterion: from .57 (Eye-Hand Coordination Test, Pattern D) to .60 (Table 39).

3. Summer Group, Four-fold criterion: from .54 (Eye-Hand Coordination Test, Pattern A) to .64 (Table 40).

4. Summer Group, Three-fold criterion: from .50 (Mashburn Test, time through trial 39) to .57 (Table 40).

The increments of the highest multiple R obtained over the <u>first order</u> multiple R for the various groups are as follows:

---

[59]The continuous composite differed from the dichotomous composite in the following respects:

1. The Camera Criteria were not identical in the two composites, although they were highly correlated.
2. The numbers of cases were not always identical.
3. In the dichotomous composite, hurdle type, all students who did not complete the course for any reason, whether complete criterion data were available or not, were counted as failures. However, to the degree that these cases failed, or dropped out for reasons other than those associated with flight competence, the predictor-criterion correlations on the hurdle type criteria might be expected to be attenuated.

[60]The zero order correlation coefficients in Tables 39 and 40, involving the Mashburn Serial Reaction Test, the Eye-Hand Coordination Test, Tidal Air, and Tidal Air/Body Surface should properly carry negative signs, since in terms of these measures a high score indicated poor performance. However, since zero order coefficients and multiple correlation coefficients are compared, the increment can most clearly be represented as between two positive coefficients, e.g., as between .46 and .57 rather than as between -.46 and .57.

TABLE 39

MULTIPLE CORRELATIONS WITH FOUR-FOLD CONTINUOUS CRITERION
MIDWEST PROJECT (Spring Group)

Variable	r	R	S.E.R	N
Mashburn: time through trial 39	.46	.52	.13	32
Eye-Hand: Pattern D	.45			
Two-Hand: trial 6	.38	.55	.12	32
Mashburn: time through trial 39	.46			
Eye-Hand: Pattern D	.45			
Two-Hand: trial 6	.38	.57	.12	32
Mashburn: time through trial 39	.46			
Eye-Hand: Pattern D	.45			
M.C. (number right)	.39			

MULTIPLE CORRELATIONS WITH THREE-FOLD CONTINUOUS CRITERION
MIDWEST PROJECT (Spring Group)

Variable	r	R	S.E.R	N
Mashburn: time through trial 39	.35	.58	.11	39
Eye-Hand: Pattern D	.57			
Eye-Hand: Pattern D	.57	.59	.10	39
M.C. (number right)	.33			
Mashburn: time through trial 39	.35	.60	.10	39
Eye-Hand: Pattern D	.57			
M.C. (number right)	.33			
Two-Hand: trial 6	.25	.60	.10	39
Mashburn: time through trial 39	.35			
Eye-Hand: Pattern D	.57			
M.C. (number right)	.33			

TABLE 40

MULTIPLE CORRELATIONS WITH FOUR-FOLD CONTINUOUS CRITERION
MIDWEST PROJECT (Summer Group)

Variable	r	R	S.E.R	N
Two-Hand: trial 6	.38	.61	.11	33
Eye-Hand: Pattern A	.54			
Two-Hand: trial 6	.38	.63	.10	33
Eye-Hand: Pattern A	.54			
Tidal Air	.37			
Two-Hand: trial 6	.38	.64	.10	33
Eye-Hand: Pattern A	.54			
Tidal Air	.37			
Washburn: time through trial 39	.36			

MULTIPLE CORRELATIONS WITH THREE-FOLD CONTINUOUS CRITERION
MIDWEST PROJECT (Summer Group)

Variable	r	R	S.E.R	N
Washburn: time through trial 39	.50	.56	.12	34
Tidal Air	.41			
Two-Hand: trial 6	.34	.54	.12	34
Washburn: time through trial 39	.50			
Tidal Air/Body Surface	.38			
Two-Hand: trial 6	.34	.56	.12	34
Washburn: time through trial 39	.50			
Tidal Air	.41			
Two-Hand: trial 6	.34	.57	.12	34
Washburn: time through trial 39	.50			
Tidal Air	.41			
Eye-Hand: Pattern D	.33			

1. Spring Group, Four-fold criterion: from .52 (Mashburn Test, time through trial 39, Eye-Hand Coordination Test, Pattern D) to .57 (Table 39).

2. Spring Group, Three-fold criterion: from .58 (Mashburn Test, time through trial 39, Eye-Hand Coordination Test, Pattern D) to .60 (Table 39).

3. Summer Group, Four-fold criterion: from .61 (Two-Hand Coordination Test, trial 6, Eye-Hand Coordination Test, Pattern A) to .64 (Table 40).

4. Summer Group, Three-fold criterion: from .56 (Mashburn Test, time through trial 39, Tidal Air) to .57 (Table 40).

Interpretation of these results must be guarded, in view of the small number of cases involved. On a larger and more stable population a more detailed determination of optimum batteries would be of interest. Other continuous composite criteria might be developed from the same criteria, for example, which would yield higher correlations with the predictors, although final selection of a criterion must be decided on bases other than the sizes of its correlations with predictors.

The data, however, suggest that test batteries can be set up which will yield relatively high correlations with criteria representing various aspects of flight performance.

SUMMARY.

Reliability of Predictor Tests

Physiological Tests. On the basis of the two independent samples for which data were available, the three physiological variables Body Surface, Vital Capacity, and Vital Capacity/Body Surface yielded the highest reliabilities of the several predictor variables, the reliability coefficients ranging from .76 to .94.

Psychomotor Tests. The reliabilities of the psychomotor tests were somewhat lower, although scores from the Two-Hand Coordination Test and the Mashburn Serial Reaction Test yielded reliability coefficients of .74 or greater on two out of the three independent samples on which data were available. Among the psychomotor tests those scores based on the work sample of maximum length, or on mean measures, were the most reliable. The reliability of the Link Trainer scores proved unsatisfactory.

Paper-and-Pencil Tests. Of the paper-and-pencil tests, only the reliability of the Biographical Inventory was determined on the basis of samples involved in this investigation. No unequivocal conclusions on the basis of data presented can be drawn since the coefficients vary from

.60 on a sample in which test and retest were separated by an interval of three months, to .94 on a sample in which test and retest were separated by an interval of four hours. While certain items may have been recalled after the shorter interval, the fact that during the longer interval of three months the subjects engaged in flight training may have so altered their interests and personal histories as to render the coefficients in the neighborhood of .60 too low.

Relationships Between Individual Predictors and Individual Criteria

None of the predictors on which data were available over all four samples predicted Pass-Fail, Time for Stage A, and Total Time consistently on all samples, and in general the coefficients were not high. The best prediction, in terms of more than one sample, was between the Two-Hand Coordination Test and Time for Stage A, and between the Mashburn Test and Time for Stage A. This relationship was evident for the two Midwest samples only, however.

The correlations between the predictors and the Purdue Rating Scale measures, while being in the expected direction, were in general low, and no consistent trends over the three samples on which data were available were evident.

While in most cases the coefficients were in the expected direction in regard to sign, the correlations between the predictors and the measures derived from the Ohio State Flight Inventory and from the photographic records showed few meaningful trends. However, measures on the Two-Hand Coordination Test, the Mashburn Serial Reaction Test, and the Test of Mechanical Comprehension appeared to show the most promise.

Relationships Between Individual Criteria and Development of Composite Criteria

The various criteria, as measures of flight proficiency,[61] were in general positively related, although the coefficients were not high except when two variations of the same criterion measure were considered. The highest correlations among independent criteria were between the scores on the Ohio State Flight Inventory and scores derived from the photographic records.

Two types of composite criteria were set up, the three-fold composite and the four-fold composite. The three-fold composite criterion included Total Time, the Purdue Rating Scale, and the Ohio State Flight Inventory. The four-fold composite included the foregoing criteria with the addition of measures derived from the photographic records.

---

[61] Certain of the correlation coefficients, however, were negative in sign due to the fact that in terms of some of the measures a low score denoted proficiency while in terms of others a high score denoted proficiency.

There were two variations of each of the above criteria, the dichotomous composite and the continuous composite. In the dichotomous composite cut-off points in terms of each measure in the composite were established and the criterion measures were expressed in terms of Pass-Fail. In both the three-fold and four-fold continuous composite criteria scores were distributed along a continuum.

Relationships Between Individual Predictors and Composite Criteria

On the basis of relationships with both dichotomous and continuous composites, the psychomotor tests, in general, predicted the composite criteria most efficiently. The Mashburn Test showed the highest relationships with the dichotomous composite, while the Mashburn Test and Pattern D of the Eye-Hand Coordination Test yielded the highest correlations with the continuous composite. Following the psychomotor tests, the Test of Mechanical Comprehension is probably next in rank of efficiency as a predictor in terms of either dichotomous or continuous composites, with the physiological measures and the Interview being less efficient. The Biographical Inventory, the Otis, and the Personal History Inventory showed little or no correlation with either composite criteria.

Prediction of Composite Criteria by Batteries of Predictors

In view of their reliabilities, their predictor-criteria correlations, and their interrelationships the following variables were selected for inclusion in a number of predictor batteries: the Two-Hand Coordination Test (score on trial 6 and mean score), the Mashburn Serial Reaction Test (time through trial 39), the Eye-Hand Coordination Test (Patterns A, B, and D), and the Test of Mechanical Comprehension. Tidal Air was also included in one battery to determine the contribution of a physiological variable. Not all of the tests were included in every battery.

The multiple correlation coefficients computed against the dichotomous composite were spuriously high due to the fact that biserial coefficients were employed in determining predictor-criterion correlations, while Pearson coefficients were used in computing the intercorrelations between predictors. However, against the continuous composite the multiple correlations between batteries and criteria ranged in general between .50 and .60. Further analysis indicated that in general the multiple correlation was practically maximized after the best two measures had been combined. In many cases the predictive value of the battery was little better than the predictive value of the best single test in the battery.

Interpretation of these results must be guarded in view of the small number of cases involved. Certainly a detailed determination of the most satisfactory batteries on a larger and more stable population would be of value.

APPENDIX A

INTERCORRELATIONS AMONG PREDICTORS

TABLE 41

INTERCORRELATIONS OF PHYSIOLOGICAL MEASURES
BOSTON PROJECT (Fall Group)
(N = 85)

	Variable	Mean	Sigma
1.	Vital Capacity (VC)	5067.2	622.4
2.	Tidal Air (TA)	727.9	214.8
3.	Tidal Air/Body Surface (TA/BS)	387.3	111.0
4.	Vital Capacity/Body Surface (VC/BS)	2693.5	324.1
5.	Pulse Rate (lying)	68.9	9.5
6.	Systolic Blood Pressure (lying)	121.7	10.8
7.	Diastolic Blood Pressure (lying)	73.1	8.4
8.	Smallest Pulse Pressure	18.9	8.5
9.	Time to Smallest Pulse Pressure	9.7	6.8
10.	Systolic Blood Pressure, initial change	6.1	8.4
11.	Systolic Blood Pressure, maximum change	15.4	8.4
12.	Systolic Blood Pressure, time to max. change	9.1	6.5
13.	Diastolic Blood Pressure, initial change	11.2	7.2
14.	Diastolic Blood Pressure, maximum change	18.7	6.4
15.	Diastolic Blood Pressure, time to max. change	8.5	6.6
16.	Pulse Rate, initial change	19.6	8.0
17.	Pulse Rate, maximum change	24.7	7.6
18.	Pulse Rate, time to max. change	6.2	5.7
19.	Maximum Pulse Rate	93.8	11.4
20.	Pulse Pressure, maximum change	30.0	10.2

TABLE 41 (Continued)

INTERCORRELATIONS OF PHYSIOLOGICAL MEASURES
BOSTON PROJECT (Fall Group)
(N = 85)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
*1.	—	.15	.09	.88	.06	-.13	-.17	.02	.20	-.01	-.11	.04	.08	.16	.13	.16	.25	.04	.15	-.01
2.		—	.98	.08	.08	-.06	.00	.00	-.01	.01	-.08	-.03	-.05	-.04	.08	.04	-.06	-.07	.02	-.09
3.			—	.11	.11	-.06	-.03	.03	-.02	-.01	-.08	.02	-.01	-.04	.08	.02	-.07	-.10	.02	-.08
4.				—	.10	-.13	-.25	.09	.19	-.02	-.13	.14	.02	.18	.11	.15	.23	-.06	.17	-.01
5.					—	.35	.27	-.09	.05	.16	.28	.02	.02	-.15	.11	-.10	-.16	.01	.71	.25
6.						—	.50	.30	.19	.27	.50	.10	.01	-.03	.10	-.04	-.04	-.10	.26	.42
7.							—	-.09	-.04	.10	.17	.03	-.23	-.57	-.06	-.09	-.12	.04	.16	-.21
8.								—	.12	-.33	-.39	-.03	-.08	-.02	.12	-.22	-.22	-.11	-.24	-.41
9.									—	-.16	.13	.47	-.24	.03	.18	-.13	-.05	.00	-.02	.17
10.										—	.63	-.27	-.09	-.15	.00	.28	.27	-.09	.34	.50
11.											—	.15	-.13	-.16	.02	.18	.17	.05	.34	.73
12.												—	-.04	-.05	-.11	-.09	-.08	-.19	-.07	.08
13.													—	.54	-.24	-.10	-.12	-.06	-.05	.24
14.														—	.09	.11	.13	-.15	-.04	.44
15.															—	.13	.12	.07	.14	.05
16.																—	.85	-.27	.45	.17
17.																	—	-.08	.51	.21
18.																		—	-.02	-.02
19.																			—	.35
20.																				—

*See page 89 for the physiological variable corresponding with each of the numbers.

TABLE 42

INTERCORRELATIONS OF SPIROMETRIC MEASURES
BOSTON PROJECT (Fall Group)
(N = 103)

	BS	TA	VC	VC/BS	TA/BS
Body Surface (BS	—	.15	.36	-.10	-.07
Tidal Air (TA)		—	.13	.07	.98
Vital Capacity (VC)			—	.88	.05
VC/BS				—	.01
TA/BS					—

TABLE 43

INTERCORRELATIONS OF SPIROMETRIC MEASURES
BOSTON PROJECT (Spring Group)
(N = 90)

	BS	TA	VC	VC/BS	TA/BS
Body Surface (BS)	—	.29	.47	.05	.12
Tidal Air (TA)		—	.18	.07	.98
Vital Capacity (VC)			—	.90	.12
VC/BS				—	.08
TA/BS					—

TABLE 44

INTERCORRELATIONS OF SELECTED PHYSIOLOGICAL MEASURES
MIDWEST PROJECT (Spring Group)
(N = 50)

	BS	VC	TA	TA/BS	VC/BS	PR
Body Surface (BS)	—	.63	-.10	-.29	.02	-.22
Vital Capacity (VC)		—	-.01	-.13	.78	-.13
Tidal Air (TA)			—	.97	.05	.23
TA/BS				—	.05	.27
VC/BS					—	.02
Pulse Rate (lying) (PR)						—

TABLE 45

INTERCORRELATIONS OF SELECTED PHYSIOLOGICAL MEASURES
MIDWEST PROJECT (Summer Group)
(N = 48)

	BS	VC	TA	TA/BS	VC/BS	PR
Body Surface (BS)	—	.39	.17	.04	.05	-.04
Vital Capacity (VC)		—	.23	.20	.93	.18
Tidal Air (TA)			—	.99	.19	.22
TA/BS				—	.21	.23
VC/BS					—	.22
Pulse Rate (lying) (PR)						—

TABLE 46

INTERCORRELATIONS AMONG TRIALS OF THE TWO-HAND COORDINATION TEST AND THE MASHBURN TEST
BOSTON PROJECT (Fall Group)
(N = 87)

	Two-Hand Coordination Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.	Trial 1	--	.49	.46	.38	.49	.37	.79	.52	.67	-.25	-.24	-.27	-.28	-.19	-.31
2.	Trial 2		—	.55	.46	.57	.53	.70	.67	.77	-.29	-.32	-.30	-.31	-.58	-.27
3.	Trial 3			—	.69	.70	.59	.67	.75	.84	-.29	-.27	-.26	-.26	-.21	-.16
4.	Trial 4				—	.63	.54	.65	.67	.78	-.15	-.14	-.17	-.17	-.11	-.18
5.	Trial 5					—	.74	.66	.85	.87	-.35	-.34	-.37	-.39	-.27	-.36
6.	Trial 6						—	.52	.88	.80	-.40	-.40	-.41	-.41	-.32	-.36
7.	Lowest Score							—	.67	.84	-.29	-.28	-.29	-.29	-.21	-.26
8.	Highest Score								—	.92	-.37	-.38	-.40	-.41	-.32	-.35
9.	Mean, 6 Trials									—	-.37	-.37	-.38	-.39	-.30	-.35
	Mashburn Serial Reaction Test															
10.	Time through Trial 13										—	.95	.93	.93	.77	.73
11.	Time through Trial 26											—	.98	.98	.93	.76
12.	Time through Trial 39												—	.99	.91	.87
13.	Time through Trial 40													—	.91	.87
14.	Time, Trials 14-26, inc.														—	.70
15.	Time, Trials 27-39, inc.															—
	Mean	41.3	48.7	54.3	56.3	58.0	60.3	36.4	66.6	53.3	1.8	3.6	5.2	5.3	1.7	1.6
	Sigma	13.9	14.7	14.2	14.2	14.2	14.4	12.3	12.2	11.3	0.5	0.8	1.1	1.1	0.4	0.3

TABLE 47

INTERCORRELATIONS OF PSYCHOMOTOR TESTS
BOSTON PROJECT (Spring Group)
(N = 88)

	Two-Hand Coordination Test	1	2	3	4	5	6	7	8	9	10	11
1.	Mean (6 Trials)	—	-.38	-.49	-.49	-.49	-.42	-.39	-.22	-.19	-.11	-.19
	Mashburn Test											
2.	Time through Trial 13		—	.81	.81	.81	.41	.60	.21	.23	.14	.21
3.	Time through Trial 26			—	.96	.95	.85	.71	.26	.30	.20	.23
4.	Time through Trial 39				—	.99	.83	.85	.28	.31	.19	.23
5.	Time through Trial 40					—	.83	.85	.29	.32	.20	.29
6.	Time, Trials 14-26, inc.						—	.61	.21	.26	.18	.24
7.	Time, Trials 27-39, inc.							—	.34	.34	.22	.33
	Eye-Hand Coordination Test											
8.	Pattern A								—	.69	.55	.81
9.	Pattern B									—	.60	.83
10.	Pattern D										—	.91
11.	Patterns A+B+D											—
	Mean	55.6	1.8	3.4	5.0	5.1	1.7	1.6	75.0	59.5	67.8	67.6
	Sigma	11.9	.39	.69	.90	.91	.41	.28	12.4	11.7	8.9	8.9

TABLE 48

INTERCORRELATIONS OF SELECTED PHYSIOLOGICAL AND PSYCHOMOTOR MEASURES
MIDWEST PROJECT (Spring Group)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Body Surface (BS)	—	.63	-.10	-.29	.02	-.22	.31	.13	.01	.00	.08	-.05	-.01	-.10
2.	Vital Capacity (VC)		—	-.01	-.13	.78	-.13	.18	.03	.03	.03	.08	.00	.13	-.01
3.	Tidal Air (TA)			—	.97	.05	.23	.23	.20	-.27	-.16	-.18	.07	.01	.15
4.	TA/BS				—	.05	.27	.16	.17	-.25	-.15	-.19	.03	.02	.12
5.	VC/BS					—	.02	-.01	-.01	.04	.12	.06	.03	.13	.07
6.	Pulse rate (lying)						—	-.04	.01	-.11	-.10	-.05	.00	-.16	-.15
7.	Two-Hand: Trial 6							—	.87	.44	-.40	-.41	-.15	-.10	-.06
8.	Two-Hand: Mean Score								—	-.50	-.42	-.46	-.21	-.15	-.09
9.	Mashburn, through Trial 39									—	.92	.88	.29	.31	.27
10.	Mashburn, Time Trials 14-26										—	.68	.29	.31	.31
11.	Mashburn, Time Trials 27-39											—	.27	.27	.18
12.	Eye-Hand: Pattern A												—	.75	.71
13.	Eye-Hand: Pattern B													—	.69
14.	Eye-Hand: Pattern D														—
	Number of Cases (N)	50	50	50	50	50	50	156	156	156	156	156	156	156	156

TABLE 49

INTERCORRELATIONS OF PHYSIOLOGICAL AND PSYCHOMOTOR MEASURES
MIDWEST PROJECT (Summer Group)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Body Surface (BS)	—	.39	.17	.04	.05	-.04	-.16	-.12	.15	.21	.05	-.16	-.07	-.15
2.	Vital Capacity (VC)		—	.23	.20	.93	.18	-.03	-.04	.10	.03	.24	-.17	.04	-.21
3.	Tidal Air (TA)			—	.99	.19	.22	-.34	-.32	.14	.06	.09	.06	.16	.11
4.	TA/BS				—	.21	.23	-.31	-.29	.14	.04	.10	.06	.15	.09
5.	VC/BS					—	.22	.05	.12	.06	-.04	.24	-.14	.06	-.18
6.	Pulse Rate (lying)						—	.14	.14	-.09	-.23	-.05	-.17	-.17	-.20
7.	Two-Hand: Trial 6							—	.82	-.53	-.45	-.48	-.26	-.18	-.26
8.	Two-Hand: Mean Score								—	.50	-.44	-.44	-.27	-.24	-.24
9.	Mashburn, through Trial 39									—	.88	.90	.21	.24	.34
10.	Mashburn, Time Trials 14-26										—	.71	.23	.25	.21
11.	Mashburn, Time Trials 27-39											—	.29	.19	.13
12.	Eye-Hand: Pattern A												—	.72	.71
13.	Eye-Hand: Pattern B													—	.61
14.	Eye-Hand: Pattern D														—
	Number of Cases (N)	48	48	48	48	48	48	98	98	98	98	98	98	98	98

TABLE 50

INTERCORRELATIONS OF PAPER-AND-PENCIL TESTS
BOSTON PROJECT (Fall Group)
(N = 100)

		B.I. ± 1% A	B.I. + 1% A	B.I. ± 1% A+B	B.I. + 1% A+B	M.C.	Otis
1.	B.I.* ± 1% A	-	.93	.89	.88	.33	.14
2.	B.I. + 1% A		-	.81	.92	.35	.16
3.	B.I. ± 1% A+B			-	.94	.29	.16
4.	B.I. + 1% A+B				-	.35	.17
5.	Test of Mechanical Comprehension					-	.49
6.	Otis Test						-

*Biographical Inventory

TABLE 51

INTERCORRELATIONS OF PAPER-AND-PENCIL TESTS
BOSTON PROJECT (Spring Group)
(N = 88)

		B.I. ± 1% A	B.I. + 1% A	B.I. ± 1% A+B	B.I. + 1% A+B	M.C.	Otis
1.	B.I.* ± 1% A	-	.93	.77	.85	.39	-.07
2.	B.I. + 1% A		-	.73	.92	.34	-.07
3.	B.I. ± 1% A+B			-	.80	.43	-.01
4.	B.I. + 1% A+B				-	.43	-.02
5.	Test of Mechanical Comprehension					-	.42
6.	Otis Test						-

*Biographical Inventory

TABLE 52

INTERCORRELATIONS OF PAPER-AND-PENCIL TESTS
MIDWEST PROJECT (Spring Group)

	B.I. + 1% A	B.I. + 1% B	B.I. + 1% A+B	Otis	M.C.
B.I. + 1% A	--	-.02	.90	.11	.34
B.I. + 1% B		--	.41	.11	-.06
B.I. + 1% A+B			--	.15	.28
Otis Test				--	.33
Test of Mechanical Comprehension					--

TABLE 53

INTERCORRELATIONS OF PAPER-AND-PENCIL TESTS
MIDWEST PROJECT (Summer Group)

	B.I. + 1% A	B.I. + 1% B	B.I. + 1% A+B	Otis	M.C.	Aviation Information
B.I. + 1% A	--	.11	.88	-.20	.05	.23
B.I. + 1% B		--	.56	-.05	.21	-.02
B.I. + 1% A+B			--	-.19	.14	.18
Otis Test				--	.01	.30
Test of Mechanical Comprehension					--	.25
Aviation Information						--
N	43	43	42	42	42	42

TABLE 54

INTERCORRELATIONS AMONG PSYCHOMOTOR TESTS AND PAPER-AND-PENCIL TESTS
BOSTON PROJECT (Fall Group)
(N = 87)

Variable	Mean	Sigma
Two-Hand Coordination Test		
1. Mean (6 trials)	53.3	11.3
Washburn Serial Reaction Test		
2. Time through trial 13	1.8	0.5
3. Time through trial 26	3.6	0.8
4. Time through trial 39	5.2	1.1
5. Time through trial 40	5.3	1.1
6. Time Trials 14-26, inclusive	1.7	0.4
7. Time Trials 27-39, inclusive	1.6	0.3
8. Otis Test	52.6	8.9
9. Test of Mechanical Comprehension	43.2	13.3
Strong Vocational Interest Blank		
10. Physician	2.9	1.6
11. Psychologist	1.8	1.3
12. Purchasing Agent	3.0	1.6
13. Social Science Teacher	2.6	1.6
14. Chemist	3.5	1.7
15. Engineer	3.6	1.8
16. Life Insurance Salesman	2.7	1.5
17. Math.-Science Teacher	3.5	1.7
18. Personnel Manager	3.3	1.7
19. Aviator	4.8	1.3
20. Group I	3.8	1.5
21. Group V	3.3	1.6
22. Group VIII	3.1	1.7
23. Group IX	3.8	1.5
24. Group X	4.1	1.4

TABLE 54 (Continued)

INTERCORRELATIONS AMONG PSYCHOMOTOR TESTS AND PAPER AND PENCIL TESTS
BOSTON PROJECT (Fall Group)
(N = 67)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
*1.	-	-37	-37	-38	-39	-30	-35	20	53	22	07	04	-27	27	37	-20	-04	-07	10	13	-22	-08	-17	-10
2.		-	95	95	93	77	73	-07	-33	13	20	-08	01	00	-07	03	-02	07	-16	13	13	-12	-02	26
3.			-	98	98	93	76	-04	-32	18	23	-17	02	01	-10	05	-05	07	-18	19	14	-19	-04	32
4.				-	99	91	87	-08	-34	17	25	-18	05	-01	-13	04	-03	07	-19	18	16	-18	-05	32
5.					-	91	87	-08	-35	16	25	-17	05	-02	-13	05	-03	07	-19	18	16	-18	-05	33
6.						-	70	-01	-27	12	24	-23	01	02	-12	07	-09	06	-17	23	11	-24	-05	37
7.							-	-14	-33	14	26	-19	10	-04	-15	-03	06	05	-16	15	20	-13	-10	24
8.								-	47	26	34	-21	-11	22	11	-14	-05	12	-02	30	-01	-34	-19	24
9.									-	41	30	23	-37	49	48	-46	06	-08	20	44	-19	-33	-46	-07
10.										-	60	-45	-56	70	53	-41	-02	-28	17	85	-33	-67	-55	36
11.											-	-49	-19	53	24	-36	16	10	07	64	10	-55	-53	30
12.												-	-16	-14	20	03	09	-30	16	-54	-45	80	38	-31
13.													-	-59	-77	37	35	53	-24	-54	79	22	28	-25
14.														-	82	-74	30	-24	48	69	-39	-40	-74	-06
15.															-	-67	10	-44	50	50	-64	-16	-53	-12
16.																-	-53	21	-53	-43	20	15	82	34
17.																	-	28	46	-03	42	06	-53	-67
18.																		-	03	-18	73	10	09	-10
19.																			-	19	-22	-03	-32	-38
20.																				-	-26	-75	-58	44
21.																					-	-04	-03	-18
22.																						-	42	-44
23.																							-	22
24.																								-

Note: No decimal points are used.
*See page 99 for the variables corresponding with each of the numbers.

TABLE 55

CORRELATIONS BETWEEN PAPER-AND-PENCIL TESTS AND PHYSIOLOGICAL MEASURES
MIDWEST PROJECT (Summer Group)

	B.I. ± 1% A	B.I. ± 1% B	B.I. ± 1% A+B	Otis	Mech. Comp.	Aviation Information
Body Surface	.03	-.09	-.01	.03	.21	.01
Vital Capacity	-.41	-.42	-.53	.37	.10	.14
Tidal Air	.22	-.32	.06	.20	-.23	.32
Tidal Air/Body Surface	.23	-.32	.06	.19	-.23	.34
Vital Capacity/Body Surface	-.43	-.41	-.55	.36	.00	.13
Pulse Rate (lying)	-.03	.12	.03	.55	.01	.53
Two-Hand: trial 6	.15	.33	.28	.18	.17	.23
Two-Hand: Mean	.15	.32	.27	.22	.28	.16
Mashburn, time through trial 39	-.22	-.13	-.25	-.26	-.09	-.07
Mashburn, time trials 14-26	-.26	-.16	-.30	-.29	-.06	-.05
Mashburn, time trials 27-39	-.27	-.23	-.33	-.21	.00	-.08
Eye-Hand: Pattern A	-.11	-.13	-.15	-.11	-.55	-.36
Eye-Hand: Pattern B	-.14	-.12	-.18	-.16	-.51	-.09
Eye-Hand: Pattern D	-.06	-.12	-.11	.02	-.55	-.35

Number of Cases (N):

Aviation Information vs. Physiological Tests	18
Aviation Information vs. Psychomotor Tests	42
Other Paper-and-Pencil Tests vs. Physiological Tests	19
Other Paper-and-Pencil Tests vs. Psychomotor Tests	43

TABLE 56

INTERCORRELATIONS OF PAPER-AND-PENCIL TESTS AND AVIATION INTERVIEW RATINGS
BOSTON PROJECT (Spring Group)

	Variable	B.I. + 1%	B.I. ± 1%	M.C.	Otis	Interview Scale A	B	C	D	E	F	G	H	I	Mean Scale Rating
1.	B.I. + 1%	-	.80	.38	.10	.25	.20	.30	.46	.50	.38	.25	.15	.36	.39
2.	B.I. ± 1%		-	.46	.14	.32	.28	.31	.49	.52	.29	.22	.11	.33	.40
3.	Test of M.C. (Right minus one-half number wrong)			-	.41	.46	.16	.07	.29	.41	.02	.05	-.04	.10	.24
4.	Otis Test				-	.28	.21	.26	.13	.31	.11	.24	.18	.23	.27
	Aviation Interview														
5.	Scale A					-	.38	.45	.37	.64	.34	.47	.40	.60	.66
6.	Scale B						-	.64	.34	.47	.35	.55	.50	.52	.65
7.	Scale C							-	.53	.73	.53	.87	.83	.81	.87
8.	Scale D								-	.72	.38	.54	.46	.67	.72
9.	Scale E									-	.59	.71	.60	.79	.88
10.	Scale F										-	.61	.49	.62	.67
11.	Scale G											-	.88	.91	.90
12.	Scale H												-	.83	.82
13.	Scale I													-	.93
14.	Mean Scale Rating														-
	Mean	8.2	1.4	49.2	55.5	11.8	14.8	15.4	12.9	14.0	13.4	14.3	14.7	12.4	14.0
	Sigma	2.8	5.0	8.5	9.9	4.8	3.5	3.6	5.3	3.9	3.6	4.4	4.2	4.6	3.3
	N	62	62	62	62	69	69	69	69	69	69	69	69	69	69

TABLE 57

INTERCORRELATIONS OF PAPER-AND-PENCIL TESTS AND AVIATION INTERVIEW RATINGS
MIDWEST PROJECT (Spring Group)

Variable	B.I. + 1% A	B.I. + 1% B	B.I. + 1% A+B	Otis	M.C.	Interview Scale A	B	C	D	E	F	G	H	I	Mean Scale Rating
[illegible] + 1% A	-	-.02	.90	.11	.34	.25	.12	.13	.33	.38	-.13	.04	.06	.24	.21
[illegible] + 1% B		-	.41	.11	-.06	.13	.14	-.05	-.02	-.02	-.11	-.01	.04	-.02	.01
[illegible] + 1% A+B			-	.15	.28	.30	.18	.10	.30	.36	-.17	.04	.07	.23	.21
[illegible] Test				-	.33	.27	.02	.01	.16	.24	-.12	.04	.04	.16	.11
[illegible] of M.C.					-	.42	.11	-.01	.16	.38	-.04	.00	.06	.14	.17
Aviation Interview															
[illegible] A						-	.58	.47	.67	.70	.23	.53	.46	.67	.75
[illegible] B							-	.65	.54	.61	.27	.69	.62	.70	.78
[illegible] C								-	.46	.53	.45	.90	.86	.81	.84
[illegible] D									-	.84	.23	.54	.47	.74	.77
[illegible] E										-	.27	.62	.54	.78	.81
[illegible] F											-	.65	.62	.63	.65
[illegible] G												-	.90	.89	.90
[illegible] H													-	.82	.84
[illegible] I														-	.94
Mean Scale Rating															-

Intercorrelations between interview scales: N = 113

Intercorrelations between paper-and-pencil tests: N = 154-156

Correlations between paper-and-pencil tests and interview scales: N = 69-70