AN APPRAISAL OF CERTAIN TESTS OF PILOT APTITUDE

by

Claude E. Buxton

and

Kenneth W. Spence

A report of research conducted by members of the staff of the Department of Psychology, State University of Iowa, Iowa City, Iowa, with the cooperation of the Department of Psychology, Army Air Forces School of Aviation Medicine, Randolph Field, Texas, which furnished equipment and procedures and the National Research Council Committee on Selection and Training of Aircraft Pilots which provided a grant-in-aid from funds furnished by the Civil Aeronautics Administration.

July 1946

CIVIL AERONAUTICS ADMINISTRATION
Division of Research
Report No. 64
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. C. Flanagan

J. L. Holland

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

July 12, 1946

Dr. Dean R. Brimhall
Director of Research
Civil Aeronautics Administration
Room 3895, Commerce Building
Washington 25, D. C.

Dear Dr. Brimhall:

Attached is a report entitled An Appraisal of Certain Tests of Pilot Aptitude, by C. E. Buxton and K. W. Spence. This report 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 report describes the last and most comprehensive of a series of studies conducted at the State University of Iowa in cooperation with the Army Air Forces. The study provides important basic information on psychomotor tests used in the Army Selection and Classification Program. The findings represent a significant contribution to the fund of information concerning the reliability and validity of such instruments which has been accumulated during approximately six years of research conducted not alone at the State University of Iowa, but at the Naval Training Station, Pensacola, Florida; Harvard University, the Ohio State University, and at other Committee research centers.

The study described in this report, as is true of much of the work carried on by the Committee on Selection and Training of Aircraft Pilots, reflects the benefits of close cooperation with educational institutions and military organizations. Such cooperation has been beneficial not only from the viewpoint of enriching research results, but in terms of considerably reducing the costs involved in conducting research.

Cordially yours,

M. S. Viteles

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

MSV:rm

EDITORIAL FOREWORD

In 1943, the Committee on Selection and Training of Aircraft Pilots with the cooperation of the Army Air Forces and the U. S. Navy, undertook further investigation of the reliability and relative validity of certain of the psychomotor tests which were being currently employed in the selection of applicants for flight training. These studies were undertaken with the view to determining (1) the effects of extending the time limits (or increasing the number of trials) on the reliability and validity of these instruments; (2) the effects of certain other modifications in the procedures of administration on the reliability and validity; and (3) the interrelationships of these psychomotor tests and certain psychomotor and pencil-and-paper tests which were not being employed in selection batteries at the time this research was undertaken. The over-all goal was, of course, the construction of a selection test battery of greater validity than any already in use.

Prior to the initiation of the program sponsored by the Committee on Selection and Training of Aircraft Pilots the State University of Iowa was already cooperating with the Army Air Forces in an experimental analysis of certain psychomotor tests used in the Aviation Psychology Program of the Army Air Forces. Other reports of these investigations have appeared in the CAA Division of Research Series as Reports Nos. 29, 44, 45, and 53.[1] The present report describes another and final study to grow out of this earlier program.

Other studies sponsored by the Committee on Selection and Training of Aircraft Pilots have been concerned with the reliability, validity, and general utility of prediction instruments. In this connection the reader is referred particularly to the Pensacola study[2] and to the Boston-Midwest report.[3] For example, in terms of the most direct comparisons,

---

[1]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.

Spence, K. W., Buxton, C. E., and Melton, A. W. The effect of massing and distribution of practice on rotary pursuit test scores. Washington, D. C.: CAA Division of Research, Report No. 44, March 1945.

Spence, K. W., Buxton, C. E., and Melton, A. W. The effect of massing and distribution of practice on two-hand coordination test scores. Washington, D. C.: CAA Division of Research, Report No. 45, April 1945.

Spence, K. W., Buxton, C. E., and Melton, A. W. The effect of massing and distribution of practice on the S.A.M. complex coordination test. Washington, D. C.: CAA Division of Research, Report No. 53, December 1945.

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

[3]NRC Committee on Selection and Training of Aircraft Pilots. Report on the Boston-Midwest project. Washington, D. C.: CAA Division of Research Report No. 52, November 1945.

it is found that the reliability coefficients for the serial coordination test obtained in the Boston-Midwest study, on the basis of test-retest, are somewhat lower than the coefficients obtained by the split-halves procedure in the present investigation. On the other hand, the intercorrelations among the mechanical comprehension test, the serial coordination test, and the two-hand coordination test are of approximately the same magnitude as the intercorrelations among these tests obtained in this study. Similarly, the criterion correlations for the mechanical comprehension test reported in the present study are generally of the same order as those obtained in the Boston-Midwest investigation. Reference to the reports cited above provides other comparisons of interest to those engaged in research on psychomotor tests.

TABLE OF CONTENTS

SUMMARY

This study was concerned with the investigation of the relationship among four psychomotor tests and certain pencil-and-paper tests of perceptual motor ability, and with an evaluation of these instruments as predictors of flight success. More specifically, research was directed toward (1) determining the minimum length of each psychomotor test for adequate reliability, (2) investigating the intercorrelations among all tests employed, (3) evaluating several alternative methods of scoring test performance, (4) searching for a usable and meaningful criterion measure, and (5) evaluating all tests in terms of their predictive efficiency.

The subjects were secured from the War Training Service program for naval cadets. All were in the elementary stage of flight training. Groups A and B, studied successively, consisted of 90 and 63 subjects, respectively. Group C, consisting of 56 subjects, was treated separately since the criterion data for Group C were not comparable to those of Groups A and B.

The length of work and rest periods for each of the psychomotor tests employed were as follows: rotary pursuit test, 10 seconds work, 20 seconds rest; serial coordination test, 120 seconds work, 15 seconds rest; two-hand coordination test, 30 seconds work, 30 seconds rest; division-of-attention test, 30 seconds work, 30 seconds rest. For each test the following measures were obtained: initial, central, and final status scores, total score, early gain, later gain, and over-all gain scores. The pencil-and-paper tests in the battery included the following subtests: planning circuits, hidden figures I and II, path distance, visual pursuit, and mechanical comprehension.

Criterion data consisted of ratings on a 5-point scale for every maneuver during Stages A and B of elementary flight training, and an over-all percentage grade for the 5 to 20 maneuvers of each lesson. Four scorings of these data were employed, viz: the mean flight maneuver rating, the mean percentage grade on "last-10" lessons, the mean maneuver rating on "last-10" lessons (all three being corrected for instructor's mean and variability), and the uncorrected mean percentage grade on "last-10" lessons.

The findings of the investigation may be summarized briefly as follows:

1. For three of the psychomotor tests (rotary pursuit, serial coordination, and division-of-attention tests) split-half reliabilities of .90 or better were reached early in the test period. In the cases of the two-hand coordination test the split-half reliability coefficient based on the full length test of 30 trials was no higher than .73.

2. Correlations of successively longer segments with total score showed, however, that the relative unreliability of the two-hand coordination test did not prevent its correlations with total score from reaching the level of .90 as early as for all other tests, i.e., at about half length or less.

3. Split-half reliability coefficients for the various types of psychomotor scores indicated that gain scores appeared to be less reliable than status scores or total score. Reliability coefficients based on total score for all of the psychomotor measures was .95 or above, except for the two-hand coordination test where the reliability coefficient was no higher than .78.

4. To evaluate the several types of measures employed correlations among all of the various psychomotor test scores were computed. In general, status scores tended to be more closely intercorrelated than gain scores.

5. Correlations among the psychomotor measures suggested that (a) none of the tests intercorrelated at more than a moderately low level and (b) inter-test correlations based on gain scores were somewhat lower than correlations based on status scores.

6. Correlations between pencil-and-paper tests and total scores on the psychomotor tests were found to be low and, in general, positive.

7. Psychomotor test criterion correlations showed that (a) when all criterion scores except D were combined and various psychomotor measures correlated with these criteria, track time for Groups A and B combined was first in the number of significant coefficients; (b) the most promising predictors among the types of scores were the total scores, with scores on the central segments of the tests approximating them closely; (c) gain scores gave little promise as predictors of the training criterion; (d) Criterion D was clearly less readily predicted than other criterion measures, while Criterion A was probably predicted best, with C close behind; (e) none of the tests showed a clearly rectilinear relationship with Criteria A and C; and (f) serial position of a test in the battery had no demonstrable effect on its predictive efficiency.

8. Among the pencil-and-paper devices, the mechanical comprehension test, the hidden figures I and mechanical comprehension tests, and the path distance and mechanical comprehension tests offered the most promise as predictors of flight success, and were almost on a par with the best psychomotor predictors.

9. Multiple correlations were computed from the selected combinations of measures and/or scores, on the one hand, and Criteria A and C on the other. These coefficients showed that (a) Criterion A was probably the more accurately predicted; (b) total scores, among those explored for the psychomotor tests, were the best predictors, and (c) combinations of the most promising psychomotor test scores were about equal in predictive value to combinations of the best individual pencil-and-paper sub-tests.

AN APPRAISAL OF CERTAIN TESTS OF PILOT APTITUDE[1]

INTRODUCTION

<u>General Objectives</u>. This project was planned as a study of the relationships among certain complex psychomotor performance tests and certain pencil-and-paper tests of perceptual-motor ability, and as an evaluation of these tests as instruments for the prediction of flight proficiency.

More specifically, the aims of the project were: (1) to determine the minimum length for each psychomotor test necessary for adequate reliability; (2) to investigate the intercorrelations among all the tests and measures employed, both psychomotor and pencil-and-paper; (3) to evaluate several alternative methods of scoring test performances, especially gain scores; (4) to search for a usable and meaningful criterion measure; and (5) to evaluate all tests in terms of their predictive efficiency. Implicit throughout is the comparison of the practical value of the two categories of tests employed.

Three successive groups of subjects were secured with the intention of making identical analyses for all groups. Such a procedure permits evaluation of the test instruments on successive samples even though the samples are relatively small. With psychomotor tests where learning is important, split-half reliabilities tend to be high, giving an overly optimistic impression as to the probable stability of predictive indices for these tests. The successive sample procedure is, therefore, desirable since it provides a means of checking on the stability of such validity coefficients. Unfortunately, because of a change in the form of flight training records, the criterion data on the third group were not comparable to those for the first two groups. The analysis of their data is therefore not included in the body of this report, but is presented in Appendix A.

---

[1]This project, sponsored by the National Research Council Committee on Selection and Training of Aircraft Pilots, has been supported in the following ways: (1) the Committee on Selection and Training of Aircraft Pilots cooperated in the design of the study, furnished financial aid, and some of the equipment; (2) through Lt. Col. A. W. Melton, the Army Air Forces School of Aviation Medicine has furnished the rotary pursuit and two-hand coordination tests, and has given counsel on experimental design and procedures; (3) through Col. J. C. Flanagan, the Army Air Forces have provided appropriate pencil-and-paper test booklets; (4) the State University of Iowa has supplied technical shop assistance and facilities and laboratory space, has paid the salaries of the co-directors, and has furnished approximately half the research assistance; and (5) statistical aid was supplied by the sponsoring Committee's office at the University of Rochester. Dr. H. S. Odbert of that office shared in much of the statistical planning, and Dr. Leon Festinger in supervision of computational work. The Editorial Staff of the Committee on Selection and Training of Aircraft Pilots, in particular, H. K. Brobst, cooperated in editing the report and in setting it up for publication.

Needs Underlying this Research. In classification centers, time pressures have made it impractical to use individual flight aptitude tests which take more than about 15 or 20 minutes of a candidate's time. There is a need for longer samples of test behavior, permitting thorough investigations of reliability. It is distinctly possible, moreover, that a few tests on which the candidate performs for a long period might have greater predictive value than a larger number of tests, each of which is given for only a brief interval.

A second need arose from the fact that most tests, as they are now employed, indicate a candidate's rank or status only in terms of his total score. Little effort has been made to analyze differences in learning rate, i.e., that portion of total score variance due to individual differences in learning rate compounded with variance due to ability, or transfer, or other "causes" of individual differences in initial test performance. On the other hand, the flight criterion is a criterion or trainability or teachability, i.e., it is an index based on the success with which a new skill is mastered during a particular training period. Logically, then, there is a need for investigation of measures of change in the candidate (learning rate) in the test situation as predictors of the training criterion commonly employed in aviation research.

A third need grows out of the fact that no published materials, available at the time this research was carried out presented a comprehensive comparison of pencil-and-paper tests and psychomotor tests, although comparisons made "piecemeal," or for relatively short behavior samples had been reported. Pencil-and-paper tests might perhaps be more desirable for selection purposes because of their economy, but only if their predictive efficiency is high enough to warrant their use in place of psychomotor tests, which, it might be claimed, have greater "face validity." Moreover, if test batteries are finally to contain both types of tests, intercorrelations within and among the two general types of tests must be studied.

A fourth need is for the development of new tests with high predictive efficiency but which do not correlate closely with existing tests. One such psychomotor test has been included, built in accordance with suggestions from men in military classification service. The pencil-

---

[1](Continued) The work of gathering and analyzing the data has at various times been done by: Drs. George Bach, Howard Kendler, Merril Thompson, Gregory Kimble, Ray Denny, and by Messrs. Leo Reyna, Dale Nance, Hugh Ross, and Maurice Smith. Technical shop assistance was given by Mr. J. G. Sentinella and Mr. D. E. Wheeler.

Dr. E. C. Lundquist of the College of Engineering of the State University of Iowa and Lt. G. R. Cole, Resident Naval Officer, cooperated in scheduling subjects for the research. Logbook data for criterion purposes were secured at the Iowa City Municipal Airport through Mr. Frank Garver, Resident Civil Aeronautics Administration Inspector, and Mr. Frank Shaw, the local War Training Service contractor. All other data were collected in the laboratories of the Department of Psychology at the State University of Iowa.

and-paper tests employed are from an experimental battery developed by the AAF.

The fifth need in this type of research continues to be the collection of data under standardized conditions from a representative sample of flying trainees. This need was kept in mind in this study with Naval War Training Service cadets. Later descriptions of procedures will show the extent to which this need was met.

<u>The General Research Plan</u>. The tests used and the methods of administration are summarized below. The pencil-and-paper tests (see Table 1) consisted of an experimental battery (Aptitude Test, QP 901A) developed by the Army Air Forces.

The planning circuits test was one in which the subject visually traced simple circuit diagrams, to determine which of the alternative battery leads would correctly operate a meter without shorting. The hidden figures tests are of the Gottschaldt type, i.e., a simple geometrical figure must be detected within a complex or different figure. In the fourth test, the subject must judge distances between points indicated on aerial photographs of varying types of terrain. A modified McQuarrie[2] visual pursuit test was fifth in the series. In it the subject must follow a line from a starting point through a tangled series of lines and identify its endpoint. The sixth test dealt with information about simple physical principles, as exemplified in pictured problems. All tests were multiple-choice and were scored with corrections for guessing. In general, it may be presumed that these tests measure aspects of perceptual skill, including recognition, and of mechanical comprehension.

TABLE 1

[illegible]

Test	Number of Items	Working Time Allowed (Min.)
Planning Circuits	[illegible]	20
Hidden Figures I	[illegible]	7
Hidden Figures II	[illegible]	10
Path Distance	[illegible]	20
Visual Pursuit	25	10
Mechanical Comprehension	[illegible]	15
Total	[illegible]	82

---

[2] This pencil-and-paper battery, designed to measure mechanical ability, is described in: McQuarrie, [illegible] Mechanical ability test. J. Personnel Research, [illegible]

The tests were given in exact accordance with AAF procedures, except for the fact that the second of two visual pursuit (McQuarrie) tests and a visual maze test were omitted to shorten administration time. Half or all of each cadet class took these tests in a single session of approximately two hours, which occurred during the first two weeks of the training program.

Table 2 shows the psychomotor tests and measures which were employed.[3] The rotary pursuit test of the School of Aviation Medicine, a modification of the Koerth pursuit rotor, is pictured in Figure 1. The task of the subject is to follow with a stylus an approximately dime-sized target imbedded flush with the surface of a hard-rubber disc. This target revolves in a circle at the rate of 1 r.p.s.; the stylus is so constructed that no pressure other than its own weight can be applied on the surface of the disc to assist in the pursuit. In spite of the apparent simplicity of this task, practically no one can perform it well initially, but practically every subject improves within a relatively brief practice period. The presentation of all trials and rest periods was automatically controlled by a special control unit (see Figure 1). The score (in units of .001 min.) was the total time of contact between stylus and target.

Shown in Figure 2 is the serial coordination test, an exact duplicate of the model described by McFarland and Channell.[4] The subject faces a panel which carried three paired rows of jewel signal lights. One row in each pair contains only red jewels and the other only green. Within each red row, a particular jewel is automatically lighted, so that the stimulus pattern for the subject is a pattern of three red lights. The

TABLE 2

PSYCHOMOTOR TESTS AND TEST CONDITIONS

Test	Type of Measure	N of Trials	Trial Length	Rest Length
Rotary Pursuit	Time on target	50	10"	20"
Serial Coordination	No. of matches	16	120"	15"
Two-Hand Coordination	Time on target	30	30"	30"
Division-of-Attention	Track time, reaction time, and net time	30	30"	30"

[3]Appendix B contains the operating procedure and instructions for each test. Administrative procedures for the serial coordination and two-hand coordination tests differed from those employed in other Committee projects, i.e., the Boston-Midwest Project and the studies at the Institute of Aviation Psychology, University of Tennessee, and the Ohio State University.

[4]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. 34, September 1944.

rudder and stick used by the subject control the position of the signal light on each of the green rows. When the stick and rudder are properly positioned, green and red lights are paired, i.e., they are side by side, and the test unit automatically flashes on three new red stimulus lights for the subject to match. In this investigation there was not a constant series of stimulus patterns for each trial. Rather, the first pattern for each trial was that one which happened to follow the last pattern of the preceding trial, without any re-setting of the instrument. To simplify the scoring, a special control unit was built, which has been described by Nance, Buxton, and Spence.[5] It incorporates a master switch for control of current to the test, and a telephonic message-register as a counter. The latter is activated by the main stepping relay of the test when it sets up a new stimulus pattern. The score on each trial is the number of 3-way matches made. Trials and rests were timed by the experimenter.

The two-hand coordination test[6] is shown in Figure 3. A small target button on the top disc of the instrument moves in an irregular pattern around the circumference of a circle, radial deviations from a true circle being controlled by the action of cams concealed beneath the discs. The subject pursues the target button with a pointer, controlling the position of the latter by means of two lathe-control handles. One control moves the pointer in the near-far dimension, and one in the left-right. The score is the total time (in units of .001 min.) during which contact is made. The cams controlling the position of the target button also automatically control the length of work period by activating microswitches. Rest periods were timed by the experimenter.

In Figure 4 is shown the division-of-attention test,[7] a new apparatus developed by Buxton and Spence. The task of the subject is two-fold. First, with the preferred hand he performs a pursuit task, consisting of controlling the position of a slider, shown in a front center position in Figure 4, so that the tip of a microswitch leaf attached to it is in contact with a raised ridge which moves toward and away from him horizontally in a slot through the false lid of the instrument. The second (and coincident) part

---

[5]Nance, R. D., Buxton, C. E., and Spence, K. W. The effect of distraction lights upon performance on the Washburn Serial Coordination Test. Washington, D. C.: CAA Division of Research, Report No. 29, April 1944.

[6]An earlier revision of the two-hand coordination test was used in selection studies conducted by the National Research Council Committee on Selection and Training of Aircraft Pilots. See: McFarland, R. A., and Channell, R. C. A revised two-hand coordination test. Washington, D. C.: Airman Development Division. Report No. 36, October 1944.

[7]Buxton, C. E., and Spence, K. W. A preliminary study of the division-of-attention test. (Progress report in the files of NRC Committee on Selection and Training of Aircraft Pilots.)

f the task is to respond to signal lights turned on automatically every 3 seconds in one of the four positions shown. To turn any given light off, the subject with his non-preferred hand operates and holds one of the four toggle break-switches in the half-arc diagonally opposite the slider hand. Since the subject must fixate directly on the pursuit task, his switch responses must be governed by peripheral vision of the signal lights. He must also learn which switch operates which light, so that one could, with some appropriateness, term this a "speeded multiple-choice learning task based on peripheral cues, plus pursuit."

To explore the possibilities of this new test, three measures were taken on all trials: reaction time (cumulative time, within a trial, between presentation of signal light and operation of proper breakswitch); track time (total time slider-pointer is in contact with moving ridge-target); net time (slider on target, signal lights off). Each measure was recorded in units of .01 second. Timing of trials and rests was completely automatic.

In general, it can be seen that the psychomotor tests emphasize to varying degrees such factors or abilities as steadiness, reaction time, smoothness of control, speed of decision, peripheral vision, etc., but that strength or sheer speed of repetitive action is not demanded to a large extent. Also, each test, although it presumably emphasizes some one or two aspects of ability, nevertheless, calls for a complex performance on the part of the subject, and the abilities presumed to be important are merged with many others in the actual performance. Thus, although measures of simple functions like reaction time or tapping speed have little predictive value for complex criteria,[8,9] it is important to note that such abilities are merely presumptive components in what is measured by the present complex tests which have been most successful to date in predicting a complex criterion such as flight proficiency.

Psychomotor tests were administered either individually or to pairs of subjects (see Table 2). Each man kept four one-hour appointments during his first two weeks in the flight training program, with occasional exceptions due to illness, etc. Transfer from flying to the tests was, thus, held at a lower level than would be the case if tests were given throughout the 8-week elementary flight stage.

<u>The Subjects</u>. The subjects were secured from the War Training Service program for naval cadets. All were in the elementary stage of flight training. Group A subjects came from WTS classes 44D, 44E, and 44F; Group B subjects from succeeding classes 44G, 44H, and 44I. Table 3 indicates

---

[8] Seashore, R. H. The aptitude hypothesis in motor skills. <u>J. exper. Psychol.</u>, 1931, <u>14</u>, 551-561.

[9] Walker, R. Y., and Adams, R. D. Motor skills: the validity of serial motor tests for predicting typewriting efficiency. <u>J. gen. Psychol.</u>, 1934, <u>11</u>, 173-186.

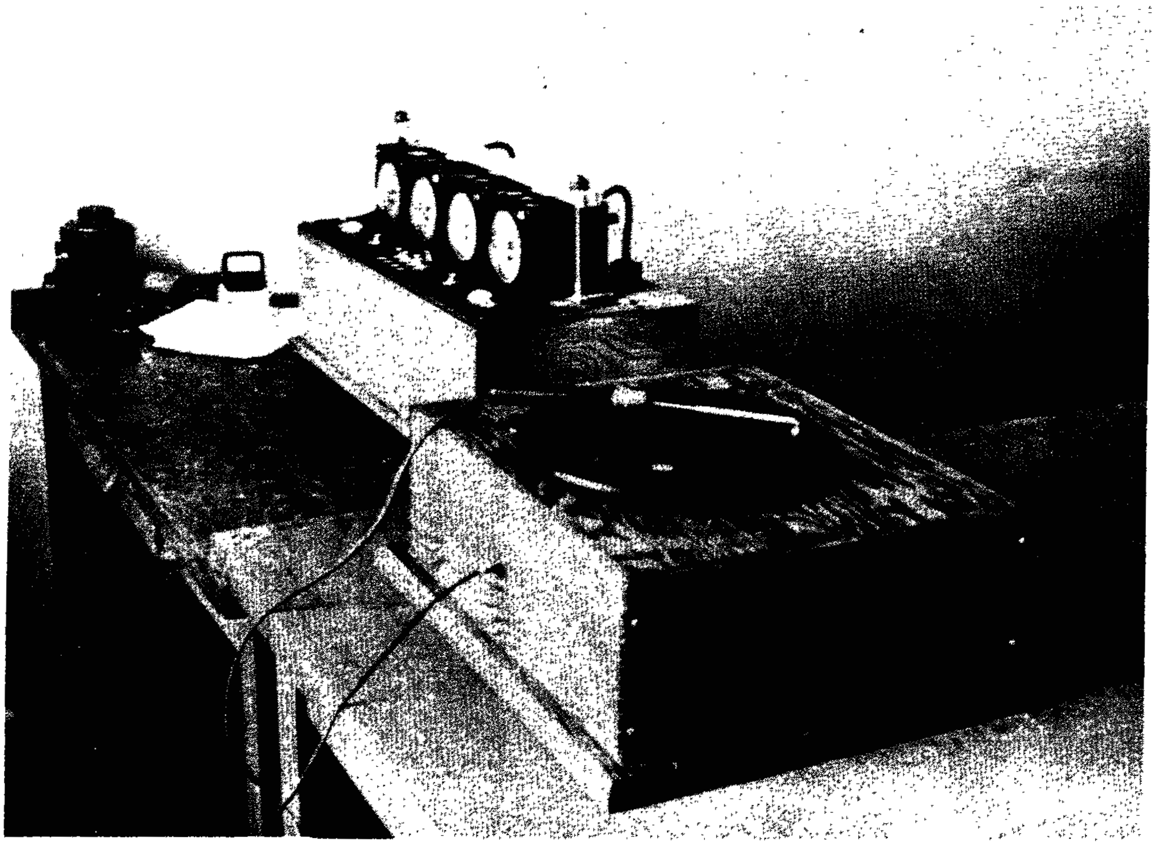

FIGURE 1

ROTARY PURSUIT TEST

FIGURE 2

SERIAL COORDINATION TEST

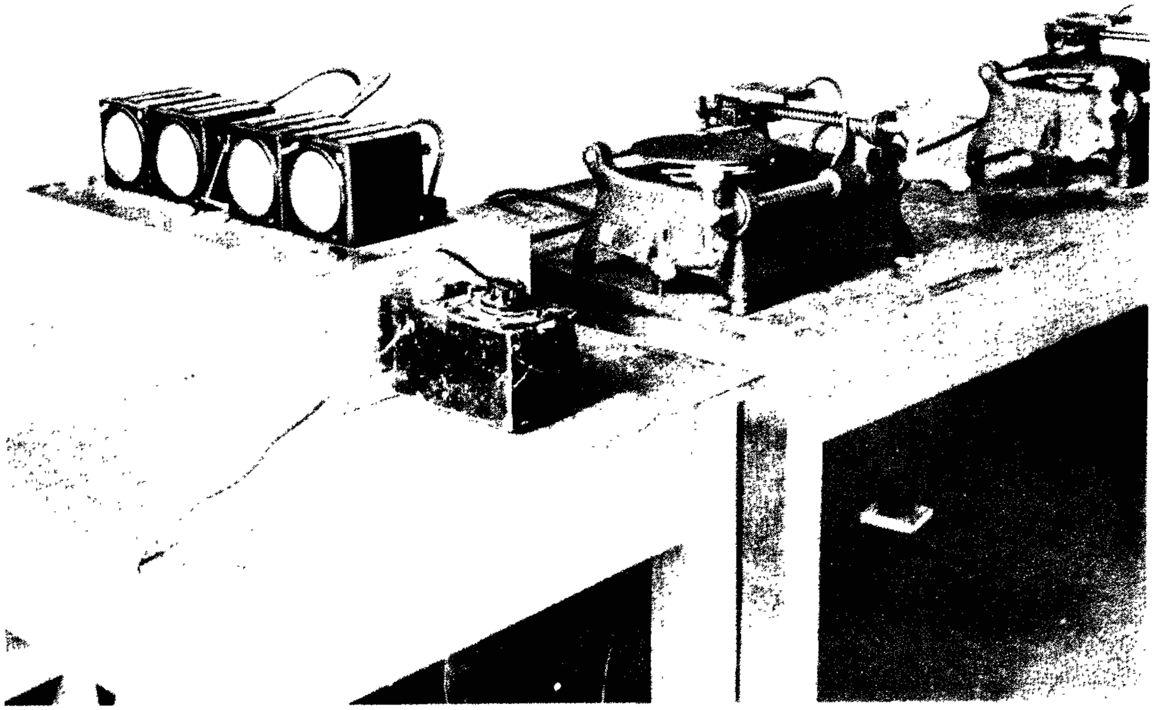

FIGURE 3

TWO-HAND COORDINATION TEST

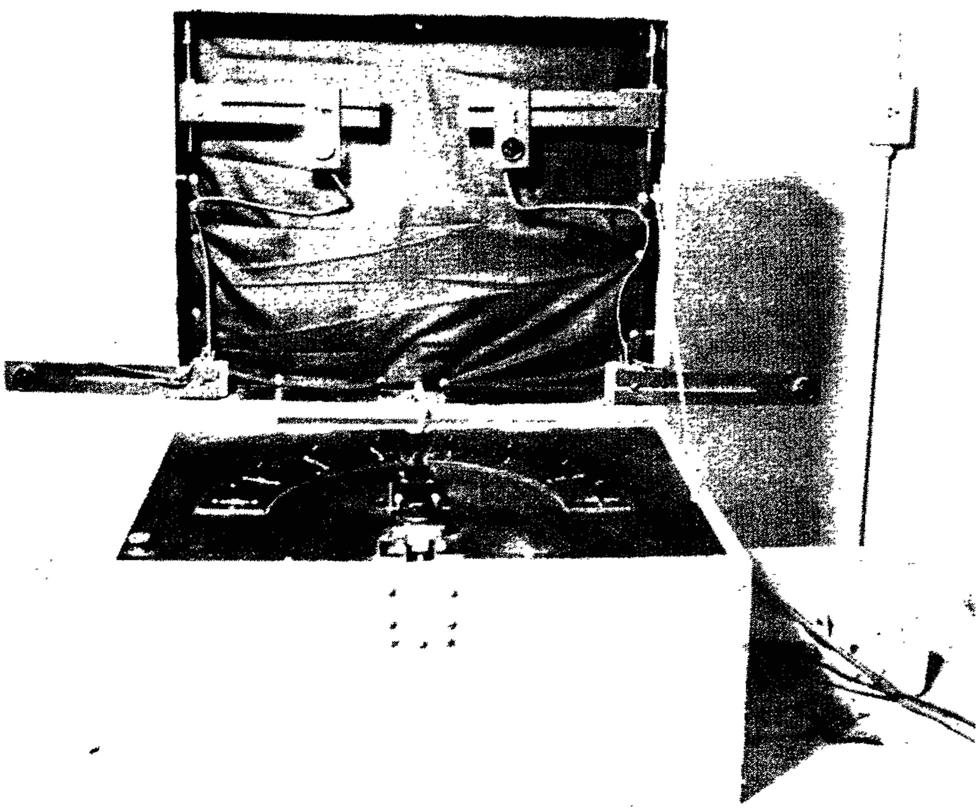

FIGURE 4

DIVISION-OF-ATTENTION TEST

the numbers of men ... in Groups A and B, and gives the reasons why certain subjects were not included in the final data.

Nineteen cases were lost during the testing program proper. Early in the research, certain recurrent flaws in the performance of the serial coordination test, and one breakdown of the division-of-attention test, caused the discarding of 12 records which otherwise were in perfect order. Three men had practiced the pursuit rotor at the University of Missouri before entering the service, and since this skill may be retained for long periods of time, this previous experience caused the discarding of these otherwise acceptable subjects. In four cases, an experimenter made some error in the operation of a test or employed an incorrect test order.

TABLE 3

NUMBERS OF SUBJECTS DISCARDED

Group	N Tested	Apparatus Breakdown	Error by E	Tested Elsewhere	Med. Washout	Re-classed	Faulty Criterion Data	Number of Valid Cases
A	112	12	3	3	3	1	-	90
B	72	-	1	-	1	1	6	63
Totals	184	12	4	3	4	2	6	153

Twelve cases were rejected because of faulty criterion data. Four exhibited airsickness and were transferred before they had many hours of flight time. Two were hospitalized for minor illnesses part way through the program, then set back a class when they recovered. Since, in this case, they began the flight training all over again, their flight records were ambiguous. The remaining six were eliminated from the study because they were the students, for a month, of a flight instructor who was thereafter dismissed. In the judgment of the resident CAA inspector and the chief instructor, his criterion ratings were not dependable.

After these eliminations, the data for 90 cases in Group A and 63 cases in Group B were found to be complete and without known flaw and were used as the basis of this report.

THE CRITERION

Source of Criterion Data. The standard CAA logbook was used by all flight instructors.[10] This logbook requires ratings, by the instructor, of each maneuver practiced during each flight lesson. Depending on the

---

[10]See Appendix C for sample page.

stage of training, the instructor, the length of the lesson, etc., from 4 or 5 up to as many as 15 or 18 ratings may be given on a single lesson. In addition, an over-all grade for the lesson is given on a percentage basis -- with 70 per cent the lowest passing mark.

On specially prepared data sheets, the experimenters tallied the numbers of ratings of 1, 2, 3, 4, or 5 (1 being the "good" end of the scale), the over-all percentage grades, and the name of the instructor. In this tally, the particular maneuver to which a rating was given was disregarded, i.e., it was assumed for convenience that all maneuvers were rated with equal adequacy, that they were equally difficult, and of equal significance as samples of criterion behavior.[11]

<u>Criterion Problems</u>. The instructors whose ratings we employed were not skilled in technical procedures of rating. Furthermore, before the study began it was quite clear that the CAA 5-point rating scale was not being fully utilized. The instructors almost never gave a rating of 1 and were but little more liberal in their use of rating 5. They also showed a tendency to stereotype their ratings on a particular cadet from day to day, even though it is probable that no trainee could be so perfectly consistent with himself. In addition, there were variations in the mean ratings from instructor to instructor, and variations in the degree to which the whole scale was employed, i.e., in the spread of the ratings.

A fair amount of improvement in the ratings was secured through group discussions led by the resident CAA inspector and the local flight contractor. They indicated to the instructors the desired use of the rating scale, and the need for making ratings immediately after a given maneuver was performed rather than waiting until the flight was finished. Differences among instructors in mean ratings and in the spread of ratings remained, but so far as could be determined, most of the instructors did their best to be cooperative and improve their ratings. One specific point which they seemed finally to accept, e.g., was that ratings should be made in comparison with trainees at the same stage of training rather than with skilled pilots.

<u>Criterion Measures Employed</u>. All ratings for flight lessons in Stages A and B of the elementary program were copied directly from the logbook. Criterion data were not taken from check flights, since these flights provided only a few ratings per man. The records of an instructor who gave less than 20 maneuver ratings on any particular cadet were also discarded.[12] Since the data could be used in an almost infinite variety of ways considerable effort was expended in trial and error computations before the following four criterion indices were selected. The reasons for choice are given with each measure.

---

[11]It would be the task of a special project to go beyond this assumption.

[12]Furthermore, there appeared to be a difference of attitude toward the purpose of the check flight among the raters. One of the raters in particular, as far as could be judged, tended to be not only independent but extreme in his ratings, and great concern was expressed by the students over a check by him.

CRITERION MEASURE A: Mean maneuver rating, with mean and variability of all instructors equalized. The mean and standard deviation of all ratings on all students by each instructor, regardless of stage of instruction, caliber of student, etc., were first determined. Then the mean maneuver rating of each cadet on all lessons by each of his instructors was secured. This mean maneuver rating of each cadet by each instructor was next converted into a z-score, as follows:

$$\text{z-score} = \frac{\text{Instructor's mean} - \text{Student mean by that instructor}}{\text{S. D., instructor}}$$

Then the instructor who gave the smallest number of ratings to each cadet was located, and the ratios of the number of ratings given by each other instructor to this smallest number determined. Treating the ratio for the smallest number of ratings as 1.00, the square root of each of the ratios was computed, and each of the instructor z-scores given to a cadet was multiplied by the appropriate root. The sum of these weighted z-scores for a cadet was divided by the sum of the square roots of the ratios for that cadet. Finally, the weighted mean z-score was multiplied by 10, increased by 50, and rounded to two places, for convenience of treatment.

The procedure for computing Criterion Measure A is based on several assumptions: First, the degree of leniency of individual instructors, and the range over which they spread their ratings, should not be allowed to influence the mean index assigned to their students. Second, the "ability" of the students, randomly assigned to each instructor, should on the average be equal. (Students were assigned at random to instructors by an academic administrator of the whole program.) Third, in determining an index for a particular cadet, allowance should be made (weight given) for the fact that some instructors had longer samples of his behavior than did others. Weighting in terms of raw number of maneuver ratings by an instructor seemed less desirable than a weight which takes into account the general fact that reliability of ratings goes up in a decreasing-returns fashion as the number of ratings increases. After various considerations it was somewhat arbitrarily decided to weight mean ratings by particular instructors on a given cadet in terms of the square root of the number of ratings.

CRITERION MEASURE B: Mean percentage grade, "last-10" lessons, adjusted for leniency and variability of instructors.[13] For a given cadet

---

[13]It should be realized that the "last-10" ratings on which this and the remaining two criteria are based include few direct ratings by an instructor, i.e., they are mainly for solo flights. The ratings are then the outcome of a post-flight interview. The fact that Criterion A and the other three criteria correlate fairly closely may be due to: (1) the analytical shrewdness with which an instructor rates after this interview; (2) a stereotyped opinion which the student and instructor form concerning the student's skill; (3) the fact that "last-10" ratings are included in over-all ratings, producing some spurious correlation.

the principal terminal instructor, i.e., the instructor who gave at least 10 lessons near the end of primary Stage B (or earlier, in the case of washouts), was determined, and the mean percentage grade on these lessons computed. For each instructor were computed the mean and standard deviation of all his student means. As in Criterion A, the the mean "last-10" grade of each cadet was converted into a z-score rating in terms of the mean and standard deviation of the particular instructor who gave the "last-10" lessons, except that, of course, there was no problem of weighting ratings by different instructors on the same cadet. Z-scores again were established for each cadet.

The reason for developing this criterion is that decisions about cadets rest primarily on performance during the latter part of the elementary training period. Furthermore, there seem to be some tendencies for instructors gradually to change the ratings on a particular cadet, i.e., ratings early in training do not give the same index as later ratings. This in itself would be a desirable flexibility in the ratings, were it not open to the charge that instructors gradually stereotype their judgments on a particular trainee. Finally, ratings may be more valid after thorough acquaintance of instructor and student.

CRITERION MEASURE C: Mean maneuver rating "last-10" lessons (analogous to Criterion B). The mean maneuver ratings by one "terminal instructor" were determined for each cadet. The mean and standard deviation of each instructor's distribution of student means were obtained and the "last-10" maneuver rating converted into a Z-score for each cadet, as for Criterion B.

CRITERION MEASURE D: Mean percentage grade, "last-10" lessons, uncorrected for leniency or variability of the individual instructors. For each cadet the mean of percentage grades under a "terminal instructor" was determined. Then the mean and standard deviation for all "last-10" percentage grades for all instructors combined were computed. The cadet's rating was converted into a z-score as before, employing his deviation from the mean of combined instructors and the standard deviation of that distribution. This measure likewise was transformed into a Z-score.

This index was included to determine whether the rather elaborate weighting procedures of the three preceding criterion indices actually produce a criterion significantly different from unadjusted indices.

Measures Discarded. At one time it was thought possible to score the criterion data in terms of the proportion of 5 or 4-and-5 ratings assigned to each student. This measure could not be employed because its distribution was badly skewed (many subjects tallying a zero frequency of these percentages). The over-all mean of percentage grades was not used because it correlated so closely with the over-all mean maneuver rating ($r = .86$).

TECHNICAL PROBLEMS AND PROCEDURES: PSYCHOMOTOR TESTS

The Buffer Test. In order to give the candidate a chance to become familiar with psychomotor testing and to recover from any initial apprehension or excitement aroused by the test situation, a buffer test was employed. This was simply a steadiness test of the Whipple variety, a stylus being held in a small hole in a metal plate, touching the sides of the hole as little as possible. The number of contacts in each of ten 30-second trials was electrically recorded, but the data were not treated.

Test Order. Two rotary pursuit tests and two two-hand coordination tests were available, so that it was desirable to employ two test orders in order to permit efficient scheduling of the subjects. The rotary pursuit test always occurred first (immediately after the buffer test) and the two-hand coordination test, third. Of a given pair of trainees, who always came together for test appointments, one took the serial coordination test and one took the division-of-attention test, second in the test series, and the remaining test as the fourth test in the series. The total group can therefore be subdivided into two groups, one having test order I (rotary pursuit test, serial coordination test, two-hand coordination test, division-of-attention test) and one having test order II (rotary pursuit test, division-of-attention test, two-hand coordination test, serial coordination test)

The Motivational Problem. At the request of the Resident Naval Officer, the cadets were all told that their test scores would have no bearing on their success or failure in the local training program. The result was an apparent lack of interest in taking the tests, and perhaps greater variability in motivation from test to test for a given cadet than would have been evident in, e.g., a classification center. On the other hand, it is the opinion of the experimenters that the apparent lack of interest was a "cultural" phenomenon, mainly at the verbal level ("I thought I joined the Navy to fly, not to be a guinea-pig"), and did not actually indicate the interest aroused by the tests in the concrete testing situation. Psychomotor tests, in particular, tend to be interesting because they contain flashing lights, moving parts; they are unfamiliar; and they show transfer of competition habits. Beginning with Group B, an effort was made to explain the project to the cadets, and thereby secure better rapport. The apparent increase in motivation as a result of this will become evident in comparing the results of Group B with Group A, as presented in a later section of this report. In general, however, the motivation may be considered adequate for the experimental situation.

Distribution of Effort. As shown in Table 2, the trials in all tests except the serial coordination test were relatively short, and the rest intervals as long as, or longer than, the work periods. The trials on the serial coordination test were longer because this was thought necessary to give a reliable trial score. The rest intervals in this test were shortened so that a relatively large number of the longer trials could be obtained.

Evidence from studies by Spence and Buxton[14, 15] makes it clear that the trial and rest lengths chosen for the rotary pursuit and two-hand coordination tests represented relatively distributed practice. Practice on the serial coordination test was probably relatively massed.[16] In the absence of experimental data, it is probable that the division-of-attention practice was relatively distributed.

Choice of Segments. Since one of the primary aims of the research was to investigate not only total or final scores on the psychomotor tests but also gains (learning), it was necessary to choose certain segments of practice to serve as a basis for gains computations. It was desired, first, to choose a segment which would give a clear indication of the subject's status at a unique position on the practice continuum and, second, to include enough trials in the segment to make the total score within it a reliable measure of initial status or final status, etc., and to permit some estimate of its reliability.

The choice of initial and final segments was relatively simple, since they could include the first or last trials. After these choices had been made it was realized that the practice period on each psychomotor test was sufficiently long that work decrement factors might begin to operate significantly. This raised the possibility of using a measure which might be largely influenced by work decrement, as a possible predictor of flight success. It was therefore decided at first to find that segment of practice which represented the highest level reached on a particular test by each cadet, to compute the gain to this point and to compute the gain or loss from this point to the end of the practice period. But, as indicated above, the practice conditions in all tests but one called for what probably was distributed effort, so that this optimum segment in each record many times coincided with that final segment, i.e., the subject never ceased improving. Therefore, what is here called a central segment, a series of trials arbitrarily located about two-thirds of the way through the practice period, was substituted. Gains to this point represent rather well the rate of change and the gross amount of change for each cadet, and gains (or slight losses) from this central segment to the final segment may give some indication of the influence of work decrement. Over-all gains can, of course, be based on the initial and final status measures. Table 4 summarizes the segments chosen for each test.

---

[14] Spence, K. W., Buxton, C. E., and Melton, A. W. The effect of massing and distribution of practice on rotary pursuit test scores. Washington, D. C.: CAA Division of Research, Report No. 44, March 1945.

[15] Spence, K. W., Buxton, C. E., and Melton, A. W. The effect of massing and distribution of practice on two-hand coordination test scores. Washington, D. C.: CAA Division of Research, Report No. 45, April 1945.

[16] Spence, K. W., Buxton, C. E., and Melton, A. W. The effect of massing and distribution of practice on the S.A.M. complex coordination test. Washington, D. C.: CAA Division of Research, Report No. 53, December 1945.

TABLE 4

ORDINAL NUMBERS OF TRIALS WITHIN SEGMENTS OF PSYCHOMOTOR PRACTICE RECORDS (BASES OF STATUS SCORES AND GAINS COMPUTATIONS)

Test	Initial Segment	Central Segment	Final Segment	Min. of Work Each Segment
Rotary Pursuit	3-8	29-34	45-50	1.0
Serial Coordination	1-2	11-12	15-16	2.0
Two-Hand Coordination	3-6	17-20	27-30	2.0
Division-of-Attention	3-6	17-20	27-30	2.0

Choice of Gains Scoring. Many different types of index could be used to show how much a cadet improved during a fixed amount of practice. For example, it was thought that the common-points-of-mastery technique might be employed. This, however, was not feasible because some of the poorest subjects never attained a performance level at or above the initial performance of the best subjects. It would be possible also to use measures which were still more elaborate statistically. It was finally decided to use the simple unmanipulated gain secured by subtracting the initial status score of a cadet (sum of trial scores in the initial segment) from the central status score, or central status from final status, or initial status from final status. Such measures should show whether or not gains (or losses) have any predictive value, and are simple to compute and understand.[17]

Test Conditions. Each psychomotor test was given in a room reserved for its use. For all but one test central overhead illumination was used and the window-shade pulled down. In the serial coordination test room, only a 60-watt bulb was used, so that the level of general illumination was not high. For the division-of-attention test the over-head light was turned out, leaving only a dim illumination because the window-shade was not opaque. The signal lights of the test thus stood out clearly in contrast with the dark surroundings.

The various test periods were scattered throughout the day and throughout the week to mitigate the differential influence of fatigue or time of day upon the data. The laboratory where the tests were given was relatively quiet and there was no major distracting influence.

---

[17]Editor's Note. The gain score has been criticized on the ground that its correlation with the criterion is not statistically independent of the correlations of the initial and terminal points with the criteria, where these initial and terminal points are used in calculating the gain score. As an alternative an analysis was made of the relationships between the slopes at particular points of fitted curves to learning data and various flight criteria. For a full discussion of this analysis see Appendix D.

Differences in test administration among the experimenters were minimized through the following precautions. Each experimenter gave all four of the tests at one time or another, thus distributing among the four tests any differences which might have occurred. In addition, rapport was established, as it were, for the whole project rather than for a particular experimenter. The presentation of all the stimuli, the time of practice and rest periods (except for the serial coordination test and rests during the two-hand coordination test) and the scoring were quite mechanical. Finally, standardized instructions were given to the experimenter and the subject in an attempt to keep the performance of the experimenter constant (see Appendix B).

Range of Scores. According to the Resident Naval Officer, cadets in the local War Training Service program had already "passed" a Navy flight aptitude test (weighted with elementary physics and mathematics), and an educational achievement or intelligence test, and had been screened on a personality inventory. The probable effect of this prior screening was to decrease the range on both the tests and the criterion, and lower their intercorrelations.

RESULTS AND DISCUSSION

Since the successive-sample procedure was employed, data are presented for Groups A and B separately, and for the two groups combined. Certain measures and procedures employed in analyzing the data of Group A were obviously not worth the labor of further exploitation. Attention will be called to them, and to their absence from the tables for Group B and for the combined groups, at the appropriate places in the test.

Practice Effects.[18] In Figure 5 is shown the mean per cent of time on the target, trial by trial, for the rotary pursuit test. The baseline of the graph shows the cumulative amounts of practice in seconds, but since each trial was 10 seconds long, the individual cross-bars on this axis indicate a single trial, the entire baseline covering the 50 trials. (Scaling in terms of cumulative amounts of practice was done in this particular instance to facilitate a comparison with a previous study which will be mentioned later.) Figure 5 shows that the last trials, on the average, produced scores four to five times larger than those made on the first trials. There is some evidence of negative acceleration and some evidence that a limit is being approached at the very end of the practice period. It is possible that the last five trials may show signs of work decrement rather than mere approach to a limit. It also could be argued that the entire curves do not depart very far from a straight line. When these curves are compared with analogous curves of Figures 1 and 2 in Spence, Buxton, and Melton,[19] it may be seen that the present

---

[18]Raw trial-by-trial data for all learning curves are recorded in Table E-1, Appendix E.

[19]Spence, K. W., Buxton, C. E., and Melton, A. W. Op. cit. (See Footnote 14.)

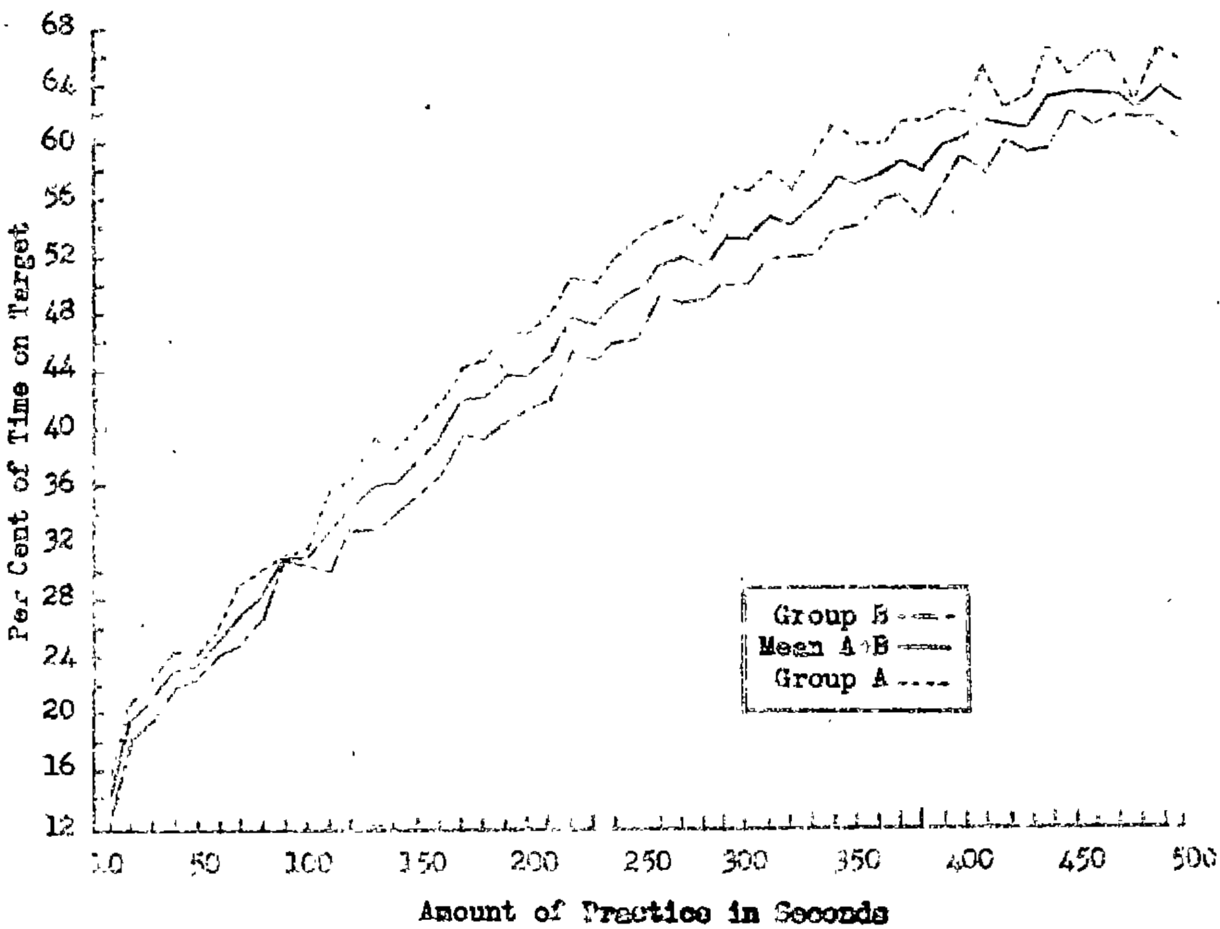

FIGURE 5

MEAN PER CENT OF TIME ON TARGET
ROTARY PURSUIT TEST

groups show very nearly the same learning curve as Group D in the earlier study, the group with the same work-rest conditions. The only differences are slight variations in the absolute height of the curves. The superior performance of Group B in the present study could be due to either stronger motivation or greater "ability."

Trial-by-trial means for the serial coordination test are shown in Figure 6. Each of the 16 trials was two minutes long and the score was the number of matches made during this interval. The mean score increased by one-half, approximately, between the first and last trials. The curves are notably uniform, and somewhat negatively accelerated, and may be approaching a limit. Group B is again consistently superior, and again there could be either a motivational or "ability" interpretation of this fact.

Learning on the two-hand coordination test is pictured in Figure 7. On the baseline are shown the 30 30-second trials and on the ordinate is shown the mean per cent of target time per trial. In over-all shape this curve is probably negatively accelerated, with some evidence that a limit is being approached, since the mean scores approximate 80 per cent of the possible score. Perhaps the most notable characteristic of this curve in comparison with others is its irregularity. The apparent irregularity of the curve was the outcome of mechanical factors in the test apparatus. The test was always begun with the target disc and cams in the same position, and the cams completed a cycle or pattern every four minutes. These four minutes provided eight of the 30-second trials, each based on a specific portion of the complete cycle, and each having its characteristic level of difficulty (because of the speed, complexity of movements, etc.). Examination of Figure 7 shows the effect of this on mean trial scores. "Lows" are present on Trials 2, 10, 18, and 26 (8 trials apart); "highs" are present on Trials 4, 12, 20, and 28, etc. There is again a tendency for Group B to be superior, but only slightly.

Figure 8a shows the learning curves for net time on the division-of-attention test. As for the two-hand coordination test, 30 30-second trials are shown on the baseline, but in this figure the vertical axis is the unmanipulated time score (to permit direct comparison with the reaction time curve of Figure 10). All three curves of Figure 8a are rather uniform, and there is evidence for both negative acceleration and approach to a limit. By the last trials, scores are approximately two and one-half times greater than for the first trial. Minor fluctuations in the curves for Groups A and B tend to be in the same direction, except at Trial 14 where there appears to be a slight difference between the two. This difference in direction is probably the result of chance factors.

Figure 8b presents learning curves based on track time scores in the division-of-attention test. The axes are exactly comparable to those of Figure 8a. Scores are nearly doubled during the practice period, and both negative acceleration and approach to a limit may be observed. Especially worthy of note here is the sharp increase in score

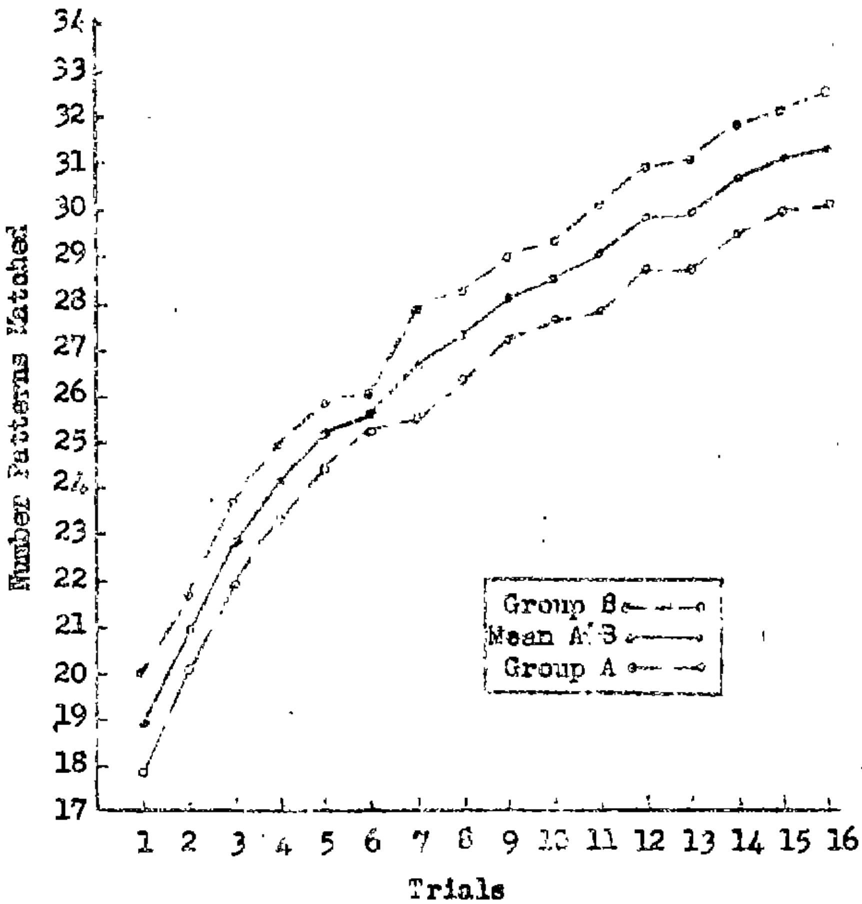

FIGURE 6

MEAN NUMBERS OF MATCHES
SERIAL COORDINATION TEST

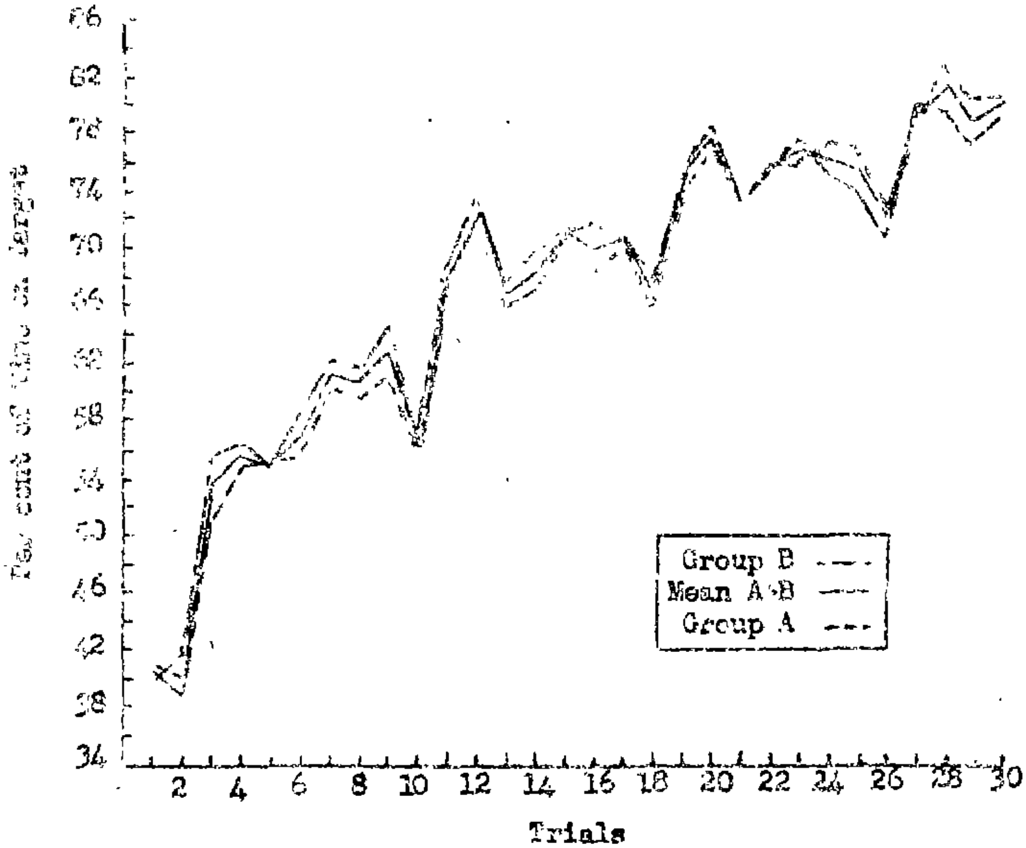

FIGURE 7

MEAN PER CENT OF TIME ON TARGET
TWO-HAND COORDINATION TEST

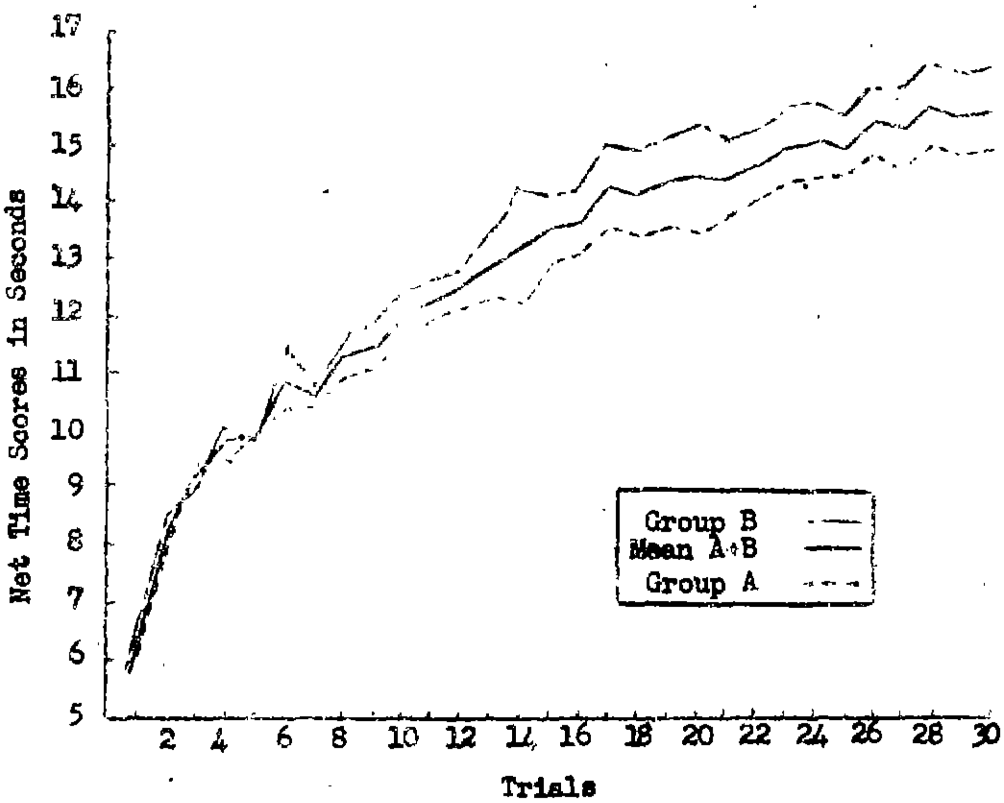

FIGURE 8a

MEAN SCORES, NET TIME
DIVISION-OF-ATTENTION TEST

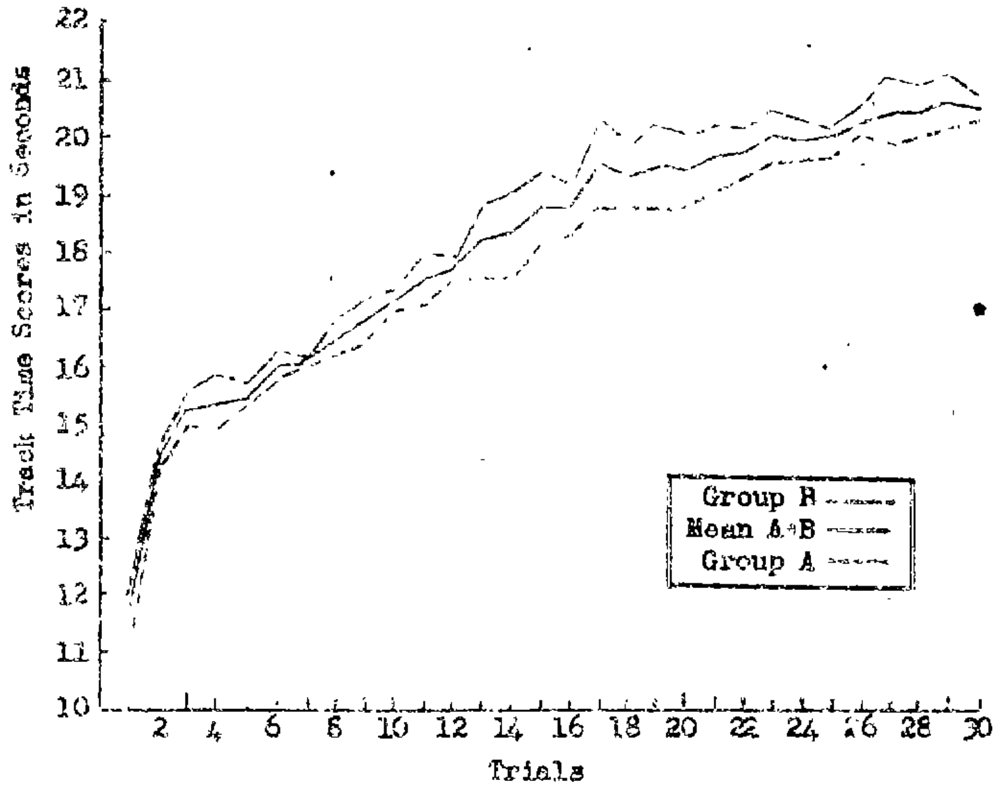

FIGURE 8b

MEAN SCORES, TRACK TIME
DIVISION-OF-ATTENTION TEST

at the second trial, the almost linear rise to Trial 17, and the gradual drift upward toward the end of the practice period.

The final set of learning curves, shown in Figure 8c, shows improvement in the reaction time scores on the division-of-attention test. Axes are identical with those of the preceding figures. As in the previous curves, there is sharp improvement on the second trial, and thereafter a negatively accelerated function. Final scores are approximately half those of the first trial, and improvement occurs at a steady pace, even in the last trials. It is probable that with only a few more trials this curve would level off, since a trial score is cumulated on eight "jump" choice-reaction times, and these would be unlikely to go much below one second each.

A comparison of the three sets of curves in Figures 8a, 8b, and 8c suggests that, in spite of the complexity of this task, the learning process was relatively "homogeneous," i.e., improvement as measured by track time and reaction time, presents about the same picture at any given stage in the process. Conversely, if the subjects had concentrated too much on the tracking task or the reaction time task this would have been reflected in the curves. Another point demonstrated by comparison of the three curves is that Group B tends always to be slightly superior to Group A.

In summary, it may be pointed out that in all of the tests there is a clear-cut improvement in average score as a result of practice and that all of the curves show signs of decreasing returns from practice, although evidence that a limit is being approached is not equally clear in all curves.[20]

Psychomotor Test Length and Reliability. Shown in Table 5 are split-half reliability coefficients for each of the psychomotor tests employed. (Definitions of trials on the psychomotor tests are given in Table 2, page 4.) The analyses were not repeated on Group B, since the results are clear-cut. Reliabilities of .90 or better are reached very early in each test period, except on the two-hand coordination test. The latter test did not reach a high level of reliability even when the coefficient is based on the complete practice period. It is not clear as to why this should be the case,[21] but there may be two possible reasons: on this test, if the subject moved off the target, it frequently was necessary for him to reverse the direction of rotation of the handles and then catch up with the target. This in fact meant "squaring" the signifi-

---

[20]Examination of the learning curves also shows that initial segments of the tests as presented in Table 4 exclude the portions of the two-hand coordination and division-of-attention curves (Trials 1 and 2) where the most rapid improvement occurs. Computation of gains from these measures of initial status does not reflect the wider range of gains which would have resulted if these early trials were utilized.

[21]This has not been the case in AAF research, although the practice conditions there did not exactly duplicate the present conditions.

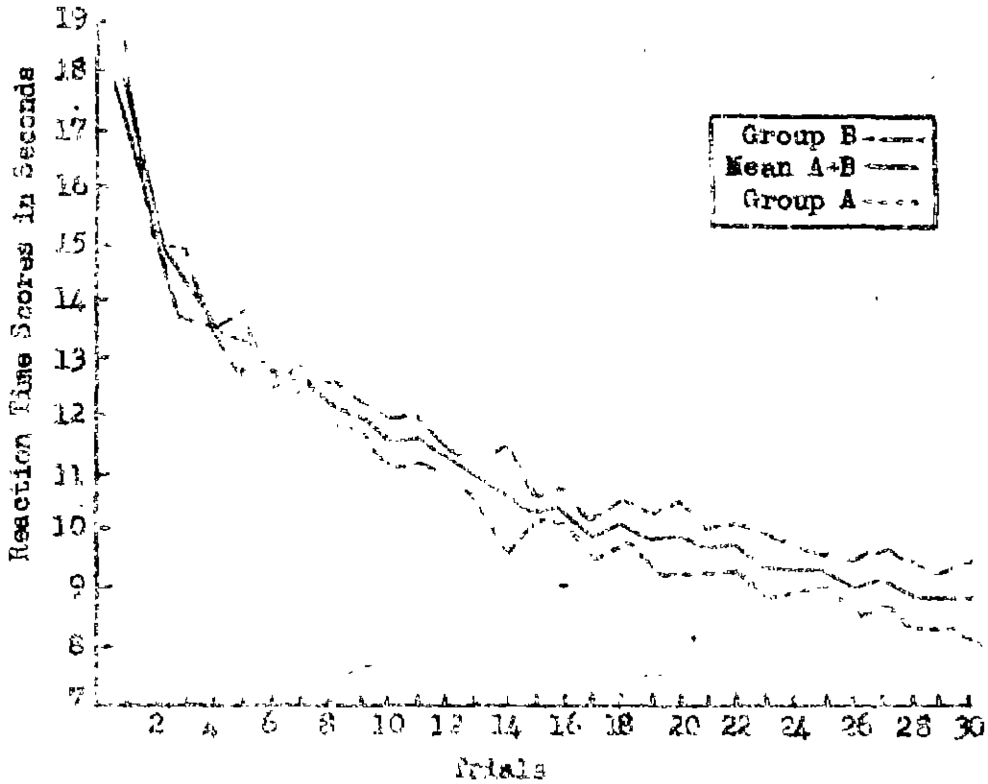

FIGURE 8c

MEAN SCORES, REACTION TIME
DIVISION-OF-ATTENTION TEST

TABLE 5

RELIABILITIES OF SUCCESSIVELY LONGER PORTIONS OF EACH TEST, GROUP A*
(N = 90)

	Rotary Pursuit		Serial Coordination		Two-Hand Coordination		Division-of-Attention			
Segments Cumulated	Trials	r	Trials	r	Trials	r	Trials	Reaction Time	Track Time	Net Time
1	3-8	.79 (.88)	1-2	.59 (.74)	3-6	.38 (.55)	3-6	.58 (.73)	.86 (.92)	.66 (.80)
1-2	3-14	.91 (.95)	1-4	.87 (.93)	3-10	.47 (.64)	3-10	.81 (.90)	.91 (.95)	.84 (.91)
1-3	3-20	.93 (.96)	1-6	.90 (.95)	3-14	.59 (.74)	3-14	.90 (.95)	.94 (.97)	.92 (.96)
1-4	3-26	.94 (.97)	1-8	.91 (.95)	3-18	.59 (.74)	3-18	.93 (.96)	.95 (.97)	.92 (.96)
1-5	3-32	.96 (.98)	1-10	.91 (.95)	3-22	.64 (.78)	3-22	.93 (.96)	.97 (.98)	.93 (.96)
1-6	3-38	.97 (.98)	1-12	.95 (.97)	3-26	.73 (.84)	3-26	.95 (.97)	.97 (.98)	.96 (.98)
1-7	3-44	.98 (.99)	1-14	.95 (.97)	3-30	.73 (.84)	3-30	.97 (.98)	.97 (.98)	.97 (.98)
1-8	3-50	.98 (.99)	1-16	.95 (.97)						

*Values in parentheses are measures of reliability estimated by the Spearman-Brown formula.

cance of any error. Such a factor might possibly lead to large departures from equality in the odd and even sums for a particular subject. In the opinion of the experimenters, another factor is probably more important, viz., that improvement was rapid and a high score was reached rather early in the practice period; thereafter rather wide and irregular variations in motivation occurred from trial to trial. More specifically, if a subject got off to a good start on a trial, he might continue trying hard to see if he could "break his record," but if he got off to a poor start, he might put out little effort for the remainder of that trial.[22]

It may be pointed out that a segment contains 1 minute of work on the rotary pursuit test; 4 minutes on the serial coordination test; 1 minute on the two-hand coordination test; and 1 minute on the division-of-attention test. In spite of the brevity of these samples, the reliability coefficients tend to be high, which is to some extent a reflection of the tendency of learning to inflate split-half reliability coefficients. In addition, these high reliabilities are characteristic of motor skills data.[23,24,25]

The primary reason for assembling these data was to determine the length of test necessary to give a score reliable enough for diagnostic purposes. This level of reliability may be arbitrarily set, for example, at .90 for split-half coefficients, or .95 for coefficients estimated for the full length of test. Examination of Table 5 shows that after about 14 trials (actually 16 since the first 2 were omitted from this computation) the rotary pursuit test reaches a satisfactorily high reliability. Under the present administration conditions this involves about 2½ minutes of work, with rests totaling about 5 minutes. The serial coordination test reaches a reliability of .90 after only 6 trials, which involve 12 minutes of work plus 1¼ minutes of rest. The division-of-attention test reaches the arbitrary level after 14 trials (again, actually 16), involving about 8 minutes of work and 7½ minutes of rest.

From the point of view of reliability alone, three of the tests would have been quite adequate had they been only about half or less than half as long as they actually were. Under the present conditions, the total test time (work plus rest) for the rotary pursuit could be as little as 7 or 8 minutes; for the serial coordination test, 13 or 14 minutes; and for the division-of-attention test, 15 or 16 minutes. On the basis of reliability data alone there is no clear indication as to how the two-hand coordination test might be shortened, because it was not of fully adequate split-half reliability even with 30 trials.

---

[22] Editor's Note. A third possible reason is that the odd and even trials which were correlated were not of equal difficulty. (See discussion on page 18, paragraph 3.)

[23] Buxton, C. E. The application of multiple factorial methods to the study of motor abilities. Psychometrika. 1938, 3, 85-93.

[24] Buxton, C. E., and Humphreys, L. G. The effect of practice upon intercorrelations in motor skills. Science. 1935, 81, 441-442.

[25] Seashore, R. H. Individual differences in motor skills. J. Gen. Psychol. 1930, 3, 38-66.

Another approach to the problem of determining a desirable test length is illustrated in Table 6, which presents correlations of successively longer segments of each test with total test score. The results of Group A are used; and only one division-of-attention measure is employed. For each test, only enough coefficients are computed to ascertain when (i.e., at what length) the sample is long enough to correlate closely with total score. The desired level of correlation can arbitrarily be set again at .90. The data of Table 6 support what was said after examination of Table 5, i.e., that not only would tests half as long as the present ones be of satisfactory reliability, but the total score on the shorter test would place the subject in almost exactly the same rank in his group as does his total score on the longer test. Of special significance is the fact that on the two-hand coordination test, despite its lower split-half reliability, the first three segments combined ranked the subjects in very much the same order as the total score.

Reliability of the Various Types of Psychomotor Scores. In Table 7 are shown split-half and estimated total test reliability coefficients for each score computed. Since the results for Groups A and B turned out to be quite similar, data for the two groups combined are presented.[26] The coefficients for gain scores were obtained by correlating the gain from the sum of odd trials of a specified segment to the sum of odd trials on the second specified segment with the gain between sums of even trials of the same two segments. Such a score for a given person may be subject to factors making for unreliability in either segment.

This is the first point to be observed when Table 7 is examined with attention to comparisons of types of score. All measures of gain tend to be unreliable; this is accentuated for the central to final gain (C-F), where the possible range of scores was restricted (see Table E-2, Appendix E) and where motivation was probably more variable. Further compar-

TABLE 6

CORRELATIONS OF SUCCESSIVELY LONGER TEST SEGMENTS WITH TOTAL SCORE, GROUP A
(N = 90)

Test	Trials per Segment	Min. Work per Segment	Segments Cumulated 1	1, 2	1, 2, 3
Rotary Pursuit	6	1	.74	.86	.93
Serial Coordination	2	4	.74	.89	.93
Two-Hand Coordination	4	2	.71	.85	.92
Division-of-Attention Net Time	4	2	.76	.86	.90

[26] Reaction time scores were discarded for Group B when this analysis was performed, and are omitted from Table 7.

TABLE 7

SPLIT-HALF RELIABILITY COEFFICIENTS FOR SEVEN TYPES OF SCORE ON EACH OF FIVE MEASURES, FOR GROUPS A AND B COMBINED*
(N = 153)

Test	Initial Score (I)**	Central Score (C)	Final Score (F)	Gain I-C	Gain C-F	Gain I-F	Total Score (T)
Rotary Pursuit	.89 (.91)	.87 (.93)	.86 (.93)	.56 (.72)	-.25	.63 (.77)	.98 (.99)
Serial Coordination	.72 (.84)	.76 (.86)	.71 (.83)	.32 (.49)	-.01	.38 (.55)	.95 (.97)
Two-Hand Coordination	.42 (.59)	.49 (.66)	.52 (.68)	.09 (.17)	.03 (.06)	.17 (.29)	.78 (.88)
Division-of-Attention Net Time	.53 (.69)	.82 (.90)	.82 (.90)	.33 (.50)	.22 (.36)	.56 (.72)	.96 (.98)
Division-of-Attention Track Time	.81 (.90)	.89 (.94)	.90 (.95)	.58 (.73)	.50 (.67)	.74 (.85)	.97 (.99)

*Reliabilities for the total score, estimated by the Spearman-Brown formula, are shown in parentheses.

**Definitions of segment scores are presented in Table 4, page 15.

isons of types of score suggest that initial segment scores are in general less reliable than central or final segment scores, with the total score being the most reliable of all. It is not surprising that the latter is the case, since a relatively long sample of performance is the computational base for this reliability coefficient. Further, it is perhaps noteworthy that the reliability of central and final segment scores should be so high when the "length of test" on which they are based corresponds to only a brief work period. The two measures on which gain scores are relatively reliable are the rotary pursuit test and division-of-attention test (track time), the latter being the highest in the table. In contrast, the coefficients for C-F gains on the serial coordination and two-hand coordination tests are in effect zero.[27]

In comparing tests, it may be noted that the two-hand coordination test gives scores which tend to be relatively unreliable, no matter what work period is used. The serial coordination test, in which a segment contains four times the work period contained by a segment of any other test or measure, is nevertheless not as reliable as either the rotary pursuit test or the two division-of-attention test measures. This probably means that not the trial length, but the "amount" of behavior actually recorded is the primary factor in reliability, and the two are not perfectly related. That is, only a limited number of matches is recorded as the serial coordination score, while on the rotary pursuit almost every instant in a trial affects the score earned. Furthermore, the subject is "paced" by the rotary pursuit test, and must pace himself on the serial coordination test, resulting in a less consistent performance on the latter test.

In general, it may be said that the initial scores are less reliable than central or final scores, presumably because in the former the various factors of set, work method, etc., are still variable. Coefficients for gains are lower than other coefficients. Initial performance and performance at later stages of practice, if the sample of performance is relatively brief (segment score), are less reliable than the longer sample, i.e., total score.

<u>Intra-test Relationships (Psychomotor)</u>. In a further attempt to analyze the psychomotor test data, and especially to evaluate the several types of measures employed, correlations among all the types of score for each test and measure were computed. This type of analysis was directed not so much at specific tests as at the discovery of relationships of some generality among all tests. A graphic means of presentation of the relationships was selected.[28]

---

[27]Some further light is thrown on the data of Table 7 by examination of Table E-2, Appendix E. C-F gains, notably, show less variability than other gains, and initial scores, in general, show less variability than other status scores. These facts are reflected, presumably, in the reliability coefficients of Table 8.

[28]Table E-3, Appendix E, gives the exact values upon which all graphs in this section are based. Because Groups A and B did not differ markedly, and because successive samples are not of direct interest at this point, only data of the combined groups will be discussed. The reaction time score on the division-of-attention test is omitted from all analyses reported here.

In Figure 9 the central and final status scores are more closely related than the initial and central status scores. The former are closer together in the test period and are in the more reliable portion of the learning curve (see Tables 6 and 7). Initial and central status scores are more closely related than initial and final scores (the former being closer together). Another interpretation of these two sets of facts is that there are differentials in rate of learning, e.g., initial status is only a moderately good indication of status after any (progressively larger) amount of practice. These observations are in accord with the common finding that adjacent segments of any performance are likely to correlate more closely than non-adjacent or distant segments. (Note that the two-hand coordination test, the least reliable of the psychomotor tests, tends to show the lowest intra-test coefficients, and the rotary pursuit test, the highest.)

A further observation may be made with respect to correlations of test segments with total score (see Figure 10). Both central and final status scores correlate closely with total score, the former showing slightly higher coefficients, and these correlations are slightly higher than those between initial and total scores. These coefficients appear to be in line with the interpretation that total score on these tests is a function of both initial or transferred skill, and the "distance" moved from that score (learning). Both factors determine central and final scores, but only one is important in the initial status score, which reflects improvement over only a very limited series of trials. That central status scores should correlate more closely than final scores with total score is perhaps to be expected in this study, since central scores are the indices which best indicate how quickly a subject reaches a high score level. The height of this general score level thus relates most closely with the total score of a long series of trials.

A third systematic distribution of coefficients may be observed in Figure 11. These indicate interrelationships among gain scores. The gain in the first part of the test (I-C) correlates with over-all gain (I-F) more closely than the gain from central status to final (C-F) with over-all gain. This is to be expected, since the range of gains is greater early in the practice period, and the I-F gain is affected somewhat by the fatigue and motivational changes during the last part of each test. In other words, there is more overlapping or duplication between gains I-C and I-F than there is between C-F and I-F. It may also be noted that the I-C gain is even less closely related to the C-F gain, and that the coefficients are negative in sign. This suggests that the more rapidly the subject improves during the early trials the less rapidly can he improve during the later trials.

In Figure 12 are shown the relationships of gain measures to total score. As mentioned above, the total score depends, in part, on how far a subject moves away from his initial scores. Therefore, I-C and I-F gains are positively correlated with total score. Also, the subject with a high score in the central segment continues to pile up a large total score but shows little or no C-F gain, while the subject with a low central score continues to gain, but his total score cannot be so great.

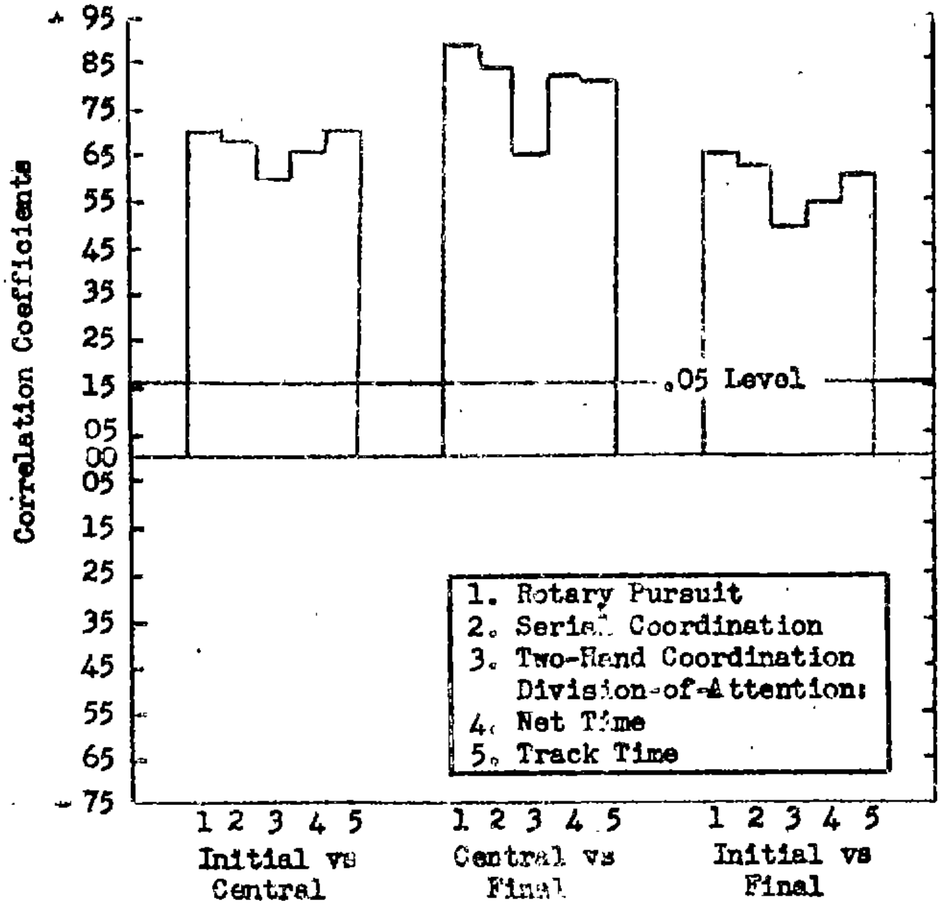

FIGURE 9

INTERCORRELATIONS OF STATUS SCORES

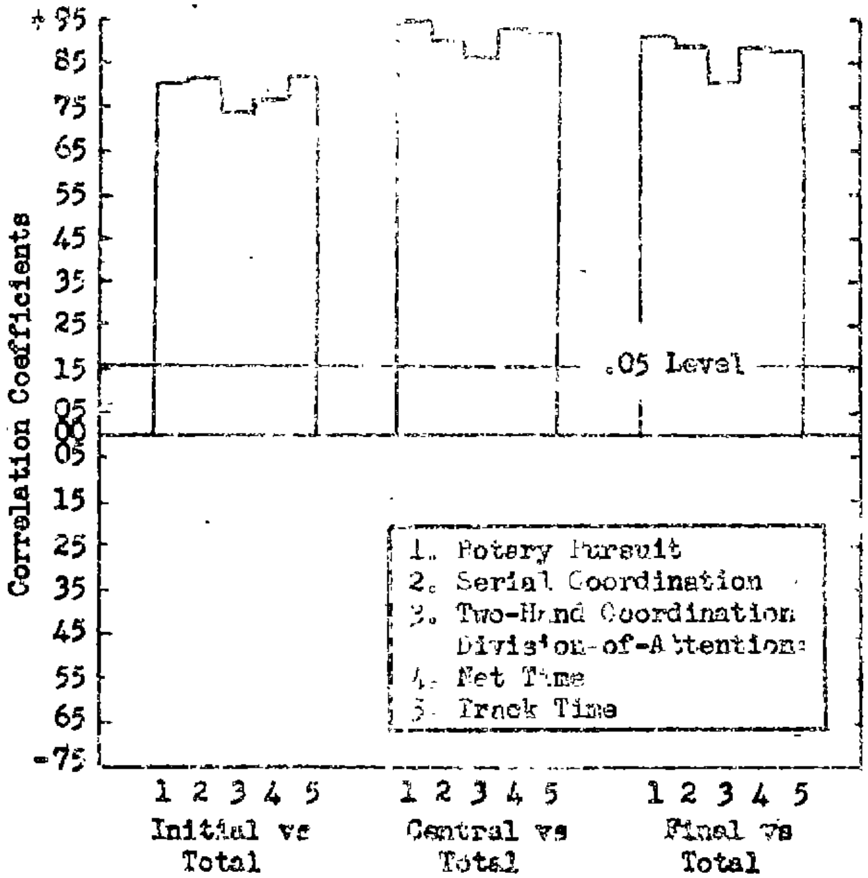

FIGURE 10

CORRELATIONS OF STATUS SCORES WITH TOTAL SCORE

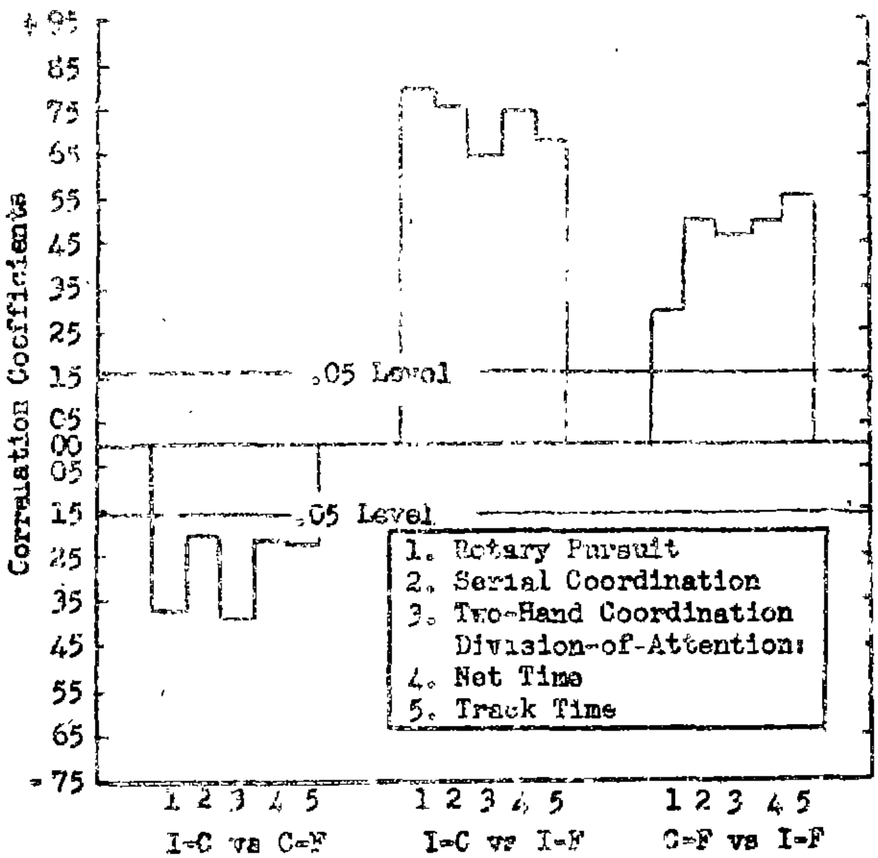

FIGURE 11

INTERCORRELATIONS OF GAIN SCORES

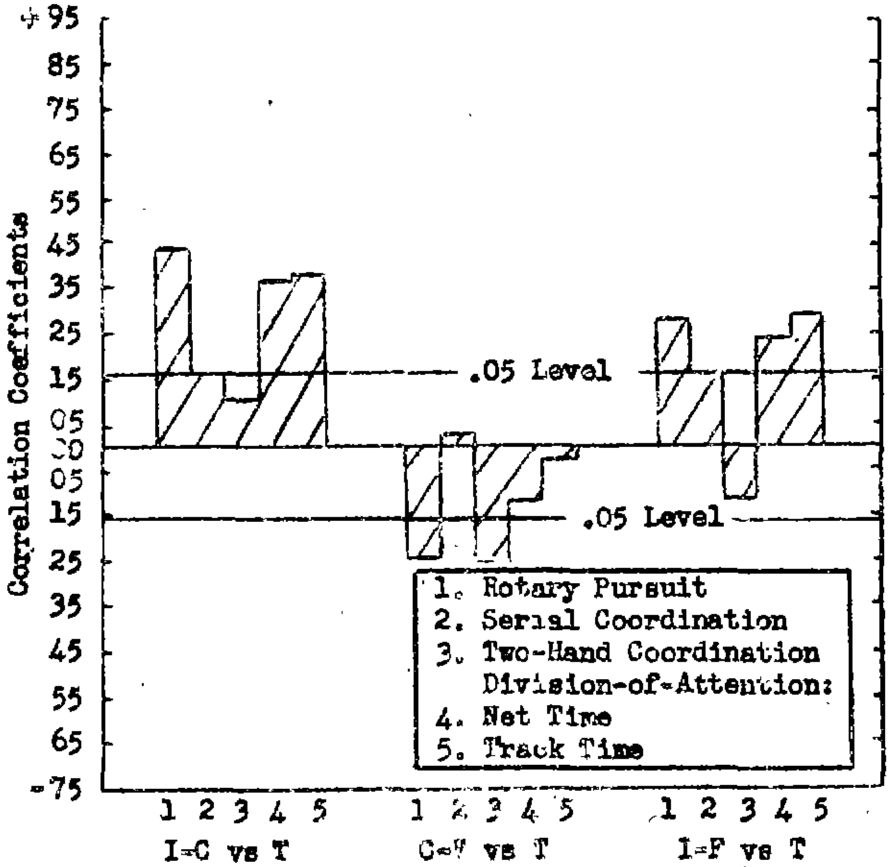

FIGURE 12

CORRELATIONS OF GAIN SCORES WITH TOTAL SCORE

Hence, the negative correlations between C-F gain and total score. There are two rather clear exceptions to these generalizations: (1) C-F gains on track time and on the serial coordination test correlate at almost the zero level with total score, probably because the range of these gains was very small (the mean gain in each case was about two points); and (2) I-F gain on the two-hand coordination test is negatively, although unreliably related to total score. The latter finding may result from the fact that improvement early in the practice period was quite rapid -- in fact, by the twelfth to the fourteenth trial, a score about seven-eighths as high as the final level had been reached. Therefore, a subject with a large I-C gain would have a small C-F gain, large I-F gain, and a large total score. However, a subject with a small I-C gain would have opportunity, under the test ceiling, to continue improving and make a clear C-F gain. He would thus have a moderately large I-F gain, but still not so large a total. The correlation between I-F gain and total score could well be zero, or possibly negative.

The final set of intra-test analyses concerns relationships between gain scores and segment scores. These correlations are represented in Figure 13, and are tabulated according to type of score. Although the absolute size of the coefficients for the several tests and measures vary, their relative positions form a consistent pattern: (1) initial status tends to correlate negatively with any gain score, i.e., the higher a subject ranks initially, the less room has he for later improvement;[29] (2) central status is positively related to I-C and I-F gains, since highest central status scores are achieved by those who have moved furthest from their initial score, but negatively related to C-F gain, since those who are highest midway in practice have the least room for improvement thereafter; (3) final score correlates positively with gain, since it is the gain which, in addition to initial skill, determines the final status score

The correlations tabulated according to type of score in Figure 13 are re-plotted according to test or measure in Figure 14, which shows that the patterns of intercorrelations, among the nine coefficients for each test or measure, are consistent. The correlations vary in their location on the ordinate (i.e., both degree of correlation and algebraic sign vary), but the positions of the points with respect to one another are uniform.

In summary, the status measures are more closely interrelated (and, as previously indicated, more reliable) than gain measures. Without repeating the detailed findings, it may be said that the degree of correlation among various scores depends upon (1) the communality of the data on which they are based, and (2) the type of representation (position on, or segment of, the learning curve). The coefficients in general are under-

---

[29]The tendency is most marked in the data for the two-hand coordination test, a fact which supports the interpretation of the relationship between I-F gains and total score, as previously suggested.

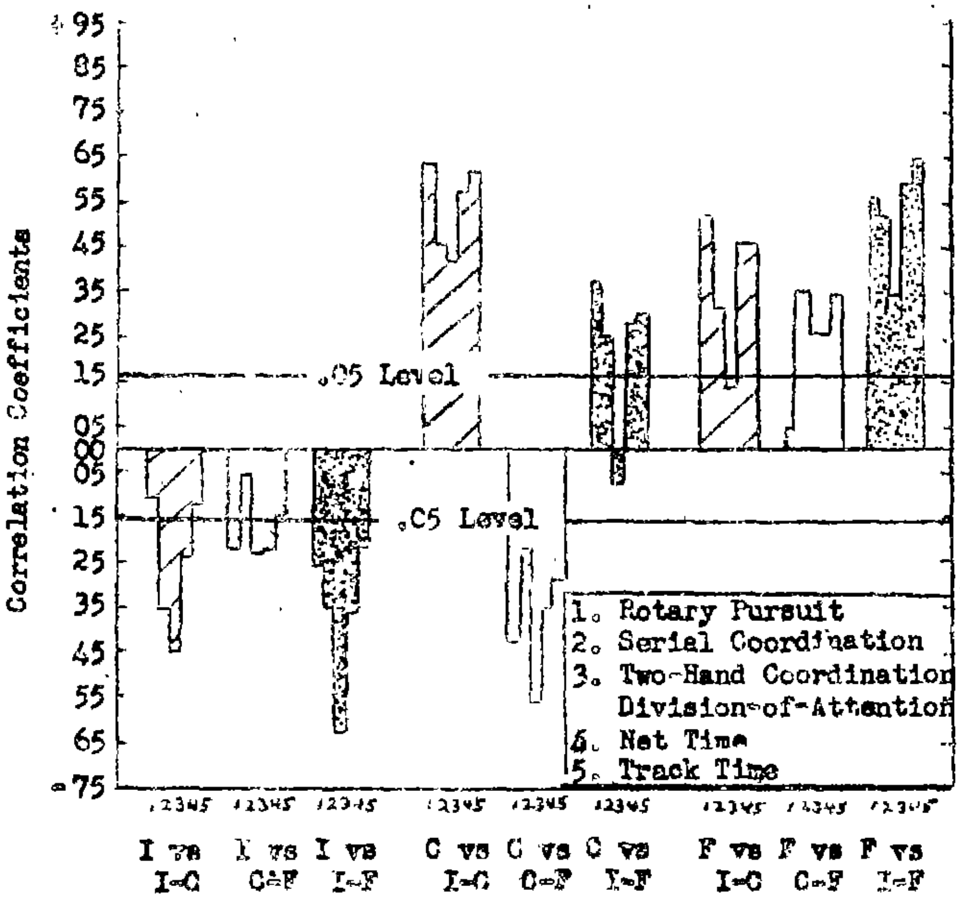

FIGURE 13

INTERCORRELATIONS OF GAIN SCORES AND STATUS SCORES

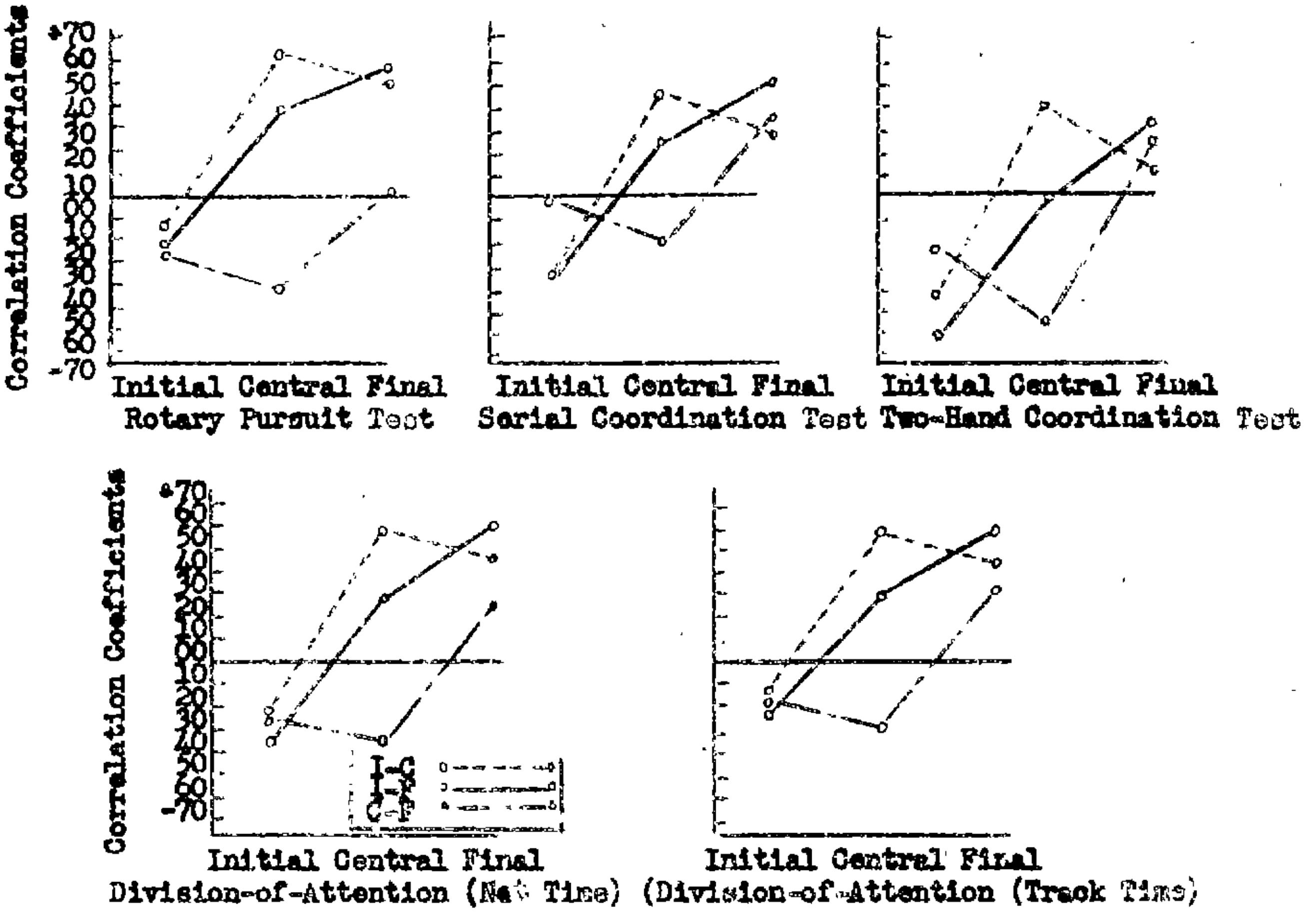

FIGURE 14

PATTERNS OF INTERCORRELATION OF VARIOUS SCORES
FOR EACH TEST OR MEASURE

s...dable in terms of knowledge about these two factors although deviations from the typical negatively accelerated curve (as in the two-hand coordination test) complicate the problem of interpretation.

Inter-test Correlations (Pencil-and-Paper). The six sub-parts of the pencil-and-paper test are here treated as separate tests, and the intercorrelations are presented in Table 8. Attention is called to the following facts: (1) the highest correlations are between the two hidden figures tests, and between planning circuits and mechanical comprehension; (2) the lowest correlations are between the visual pursuit test and others; (3) there is a rather large change in degree of intercorrelation of certain tests, from Group A to Group B, but the directions of these changes exhibit no systematic pattern; (4) in general, the intercorrelations are not very high, although planning circuits, path distance, and mechanical comprehension tend to form a cluster.

Some of these facts may be interpreted with the aid of Table 9, which shows the means, variabilities, and reliabilities of the pencil-and-paper tests. It may be pointed out that (1) the mean score on the first hidden figures test was nearly a perfect score, resulting in small standard deviations and low split-half coefficients; (2) on the planning circuits and the second hidden figures tests the mean score was approximately in the center of the chance-to-perfect range, with an appropriate spread and a higher reliability; (3) the visual pursuit test shows a mean which is somewhat low, and less variability, while the path distance and mechanical comprehension tests were too difficult. It is probable that the latter facts are a reflection of lowered motivation. This could well occur, since the path distance test was rather poorly printed, and the visual pursuit test just following it was also a contributor to eye-strain and lessened interest. The mechanical comprehension test, coming last, would suffer from these influences, despite the fact that its intrinsic interest value is probably higher than that of most of the other sub-tests. In general, only the planning circuits test appears highly reliable, although the overall reliability for combined tests is adequate.

Table 9 indicates that Group B, which scored higher on the psychomotor tests than Group A, also tended to score higher on the pencil-and-paper tests, and to show a more reliable performance. The most marked increase was in the mechanical comprehension test, where the standard deviation increased quite markedly as well. These differences and the increased reliabilities make it seem likely that Group B differed from Group A in strength and consistency of motivation. As confirming evidence of this interpretation, the subjective and casual judgments of the experimenters indicated that the subjects of Group B did less complaining and seemed to cooperate more whole-heartedly than subjects in Group A.

Inter-test Correlations (Psychomotor). Out of a large quantity of data, only general findings will be selected for attention. This may be done by grouping the inter-test coefficients according to certain classification systems.[30]

---

[30] Table E-4, Appendix E, contains the data upon which this section is based.

TABLE 8

INTERCORRELATIONS OF PENCIL-AND-PAPER SUB-TESTS FOR GROUPS A AND B AND GROUPS A AND B COMBINED

Test	Hidden Figures I			Hidden Figures II			Path Distance			Visual Pursuit			Mechanical Comprehension		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
Planning Circuits	.21	.18	.20	.40	.33	.37	.33	.34	.34	.23	.14	.20	.44	.41	.43
Hidden Figures I				.58	.37	.49	.18	.28	.21	.23	.05	.16	.21	.24	.21
Hidden Figures II							.24	.33	.28	.32	.18	.26	.37	.05	.20
Path Distance										.06	.16	.11	.26	.46	.37
Visual Pursuit													.06	.23	.15

Group A (N = 90)
Group B (N = 63)
Groups A and B combined (N = 153)

TABLE 9

SPLIT-HALF RELIABILITY COEFFICIENTS, MEANS, AND VARIABILITIES OF PENCIL-AND-PAPER TESTS FOR GROUPS A AND B AND GROUPS A AND B COMBINED

Test	Reliability			Mean			Possible	Standard Deviation		
	A	B	A+B	A	B	A+B	Maximum	A	B	A+B
Planning Circuits	.86	.94	.90	28.7	30.1	29.2	45	8.1	10.3	9.1
Hidden Figures I	.41	.55	.49	23.0	22.9	23.0	25	2.5	2.5	2.5
Hidden Figures II	.73	.83	.77	13.6	14.2	13.9	20	4.6	4.8	4.7
Path Distance	.55	.67	.61	11.7	13.0	12.2	44	3.9	3.9	3.9
Visual Pursuit	.60	.71	.62	11.7	12.3	12.0	25	2.9	2.8	2.9
Mechanical Comprehension	.59	.78	.71	10.5	11.8	11.0	30	4.9	7.0	5.9
Total Score	.85	.93	.88							

Group A (N = 90)
Group B (N = 63)
Groups A and B combined (N = 153)

[illegible] question [illegible] degree to which the [illegible] retained. This question may be answered by inspection of Figure [illegible], where the [illegible] possible inter-test coefficients are grouped according to the type of score on which they are based. Except for the rather high intercorrelations of different aspects of performance on the division-of-attention test, it is clear that the separate tests are not closely related. Many of the coefficients reach the [illegible] per cent level of significance, but few go very far above it.[31]

Another question of significance is the dependence of the test intercorrelationships upon the stage of practice. Historically, the question arose in an effort to explain the specificity of individual differences in motor skills. Briefly, the hypothesis has been that motor skills do not correlate closely with each other because, in an unsystematic fashion, each subject will vary in amount of pre-experimental experience with a particular task, or tasks closely resembling it.[32] If this hypothesis is correct, scores for initial trials on motor skills tasks should show lower intercorrelations than scores for later trials, because by the latter time (especially very late in the practice curve) transfer or specific practice should have lost their initial significance. Buxton and Humphreys[33] found that the inter-test coefficients remained about the same for early and late trials in a group of pursuit, tapping, and packing tests. Figure 35 of this study shows, on the contrary, a tendency for the coefficients based on central scores (and possibly final scores) to be higher than those based on initial scores.

Although the explaration of the differences between the two studies is outside the province of the present report, it probably is due to a difference in test conditions. The Buxton and Humphreys intercorrelations are on performances 24 hours apart, whereas the initial and central scores of the present study were obtained from trials only a few minutes apart. It is possible, therefore, that the greater the degree of massing in each test, the more likely is the correlation between tests to rise with continued practice. In the present research, the serial coordination test was the only one given under massed conditions (and even it was not paced, but depended rather upon the subject's own urge to force the pace); the division-of-attention test was not markedly massed but is inherently a "pressure" test because of its complexity. The greatest increase in degree of intercorrelation with continued practice is observed with these two tests.

---

[31]The total score intercorrelations are the ones most directly comparable to previous findings of several other investigators. See: Seashore, R. H. Op. cit. (See Footnote 25 of this report.) Also, Seashore, R. H., Buxton, C. E., and McCollom, I. N. Multiple factorial analysis of fine motor skills. Amer. J. Psychol., 1940, 53, 251-259.

[32]This is a combination of Seashore's transfer and specific practice hypotheses as described in: Seashore, R. H. Experimental and theoretical analysis of fine motor skills. Amer. J. Psychol., 1940, 53, 86-98.

[33]Buxton, C. E., and Humphreys, L. G. Op. cit. (See Footnote 24 of this report.)

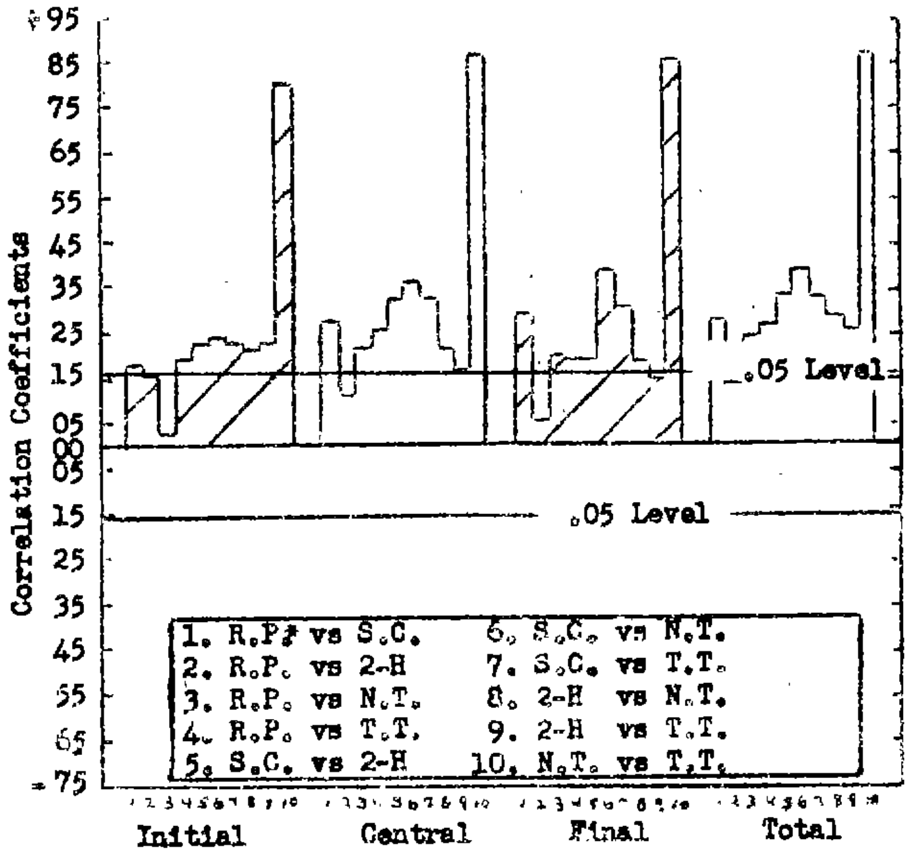

FIGURE 15.

TEST AND MEASURE INTERCORRELATIONS
BASED ON STATUS SCORES

*R.P. (Rotary Pursuit Test)
S.C. (Serial Coordination Test)
2-H (Two-Hand Coordination Test)
N.T. (Net Time Division-of-Attention Test)
T.T. (Track Time Division-of-Attention Test)

providing tentative support for the hypothesis that inter-test correlations tend to rise if the tests are both given under massed conditions.

The interrelationships of gains on the several psychomotor tests are shown in Figure 16. With the exception of the two measures on the division-of-attention test, none of the gains correlates with any other gain to a very significant degree. As in the studies by Hall,[34] Husband,[35] and Buxton,[36] there is little evidence that a general (statistical) factor of learning rate exists. It may be mentioned again, in passing, that gain scores as here computed are relatively unreliable, and also that these particular intercorrelations differed almost at random, as between Groups A and B.

The final and most directly practical information about the inter-test correlations can be ascertained from Figure 17. (Figure 15 may be examined again for the same purpose.) In this figure, the correlations of each test with every other test (excluding gain scores) are represented. That the test intercorrelations are generally low is desirable, since component parts of a test battery should not correlate closely.

Apart from this general point, four others may be made: (1) The rotary pursuit test is not closely related to the two-hand coordination test, with which it has previously correlated at least moderately. An interesting and testable interpretation at once presents itself. Especially in testing in the Army Air Forces, all tests have been given under massed, in fact continuous, practice conditions. Therefore, in addition to whatever specific factors might otherwise produce low to moderate correlation (.15 to .25) between these two tests, there is the common factor of "fatigability" of the subject.[37] This might raise the intercorrelations to .25 or .35. In the present study, however, distributed practice -- at least, more distributed than that employed by the Army Air Forces -- was used, so that the common factor of fatigability is less significant. Hence, the greater specificity of individual differences, i.e., lower intercorrelation. (2) The two-hand corrdination test tends to correlate with others at a rather low level, except when total score is the basis for the coefficient. Both facts are in line with the reliability data on this test

---

[34]Hall, C. S. Intercorrelations of measures of human learning. Psychol. Rev., 1936, 43, 179-196.

[35]Husband, R. W. Intercorrelations among learning abilities: I. J. Genet. Psychol., 1939, 55, 353-364.

[36]Buxton, C. E. Op. cit. (See Footnote 23 of this report.)

[37]It is recognized that classical studies of fatigue tend to discourage the suggestion of a general (statistical) factor, but it is possible that the reactive inhibition concept, proposed by Hull, may apply in this instance. See: Hull, C. L. Principles of behavior. New York: Appleton-Century Co., 1943.

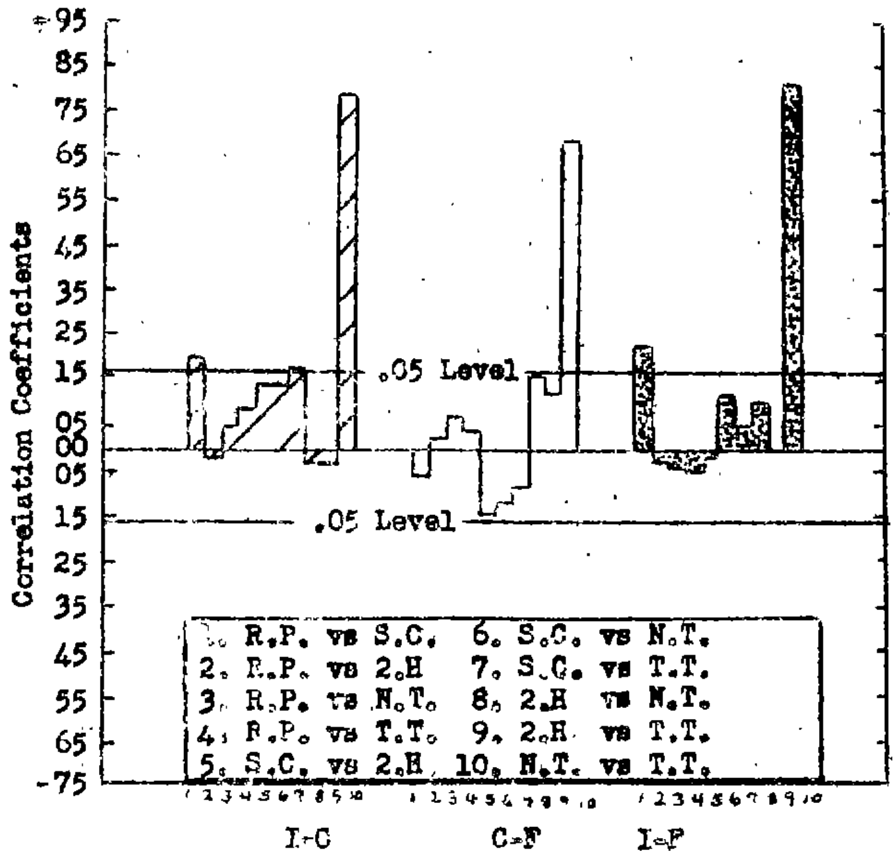

FIGURE 16

TEST AND MEASURE INTERCORRELATIONS
BASED ON GAIN SCORES

*I-C (Initial - Central)
C-F (Central - Final)
I-F (Initial - Final)

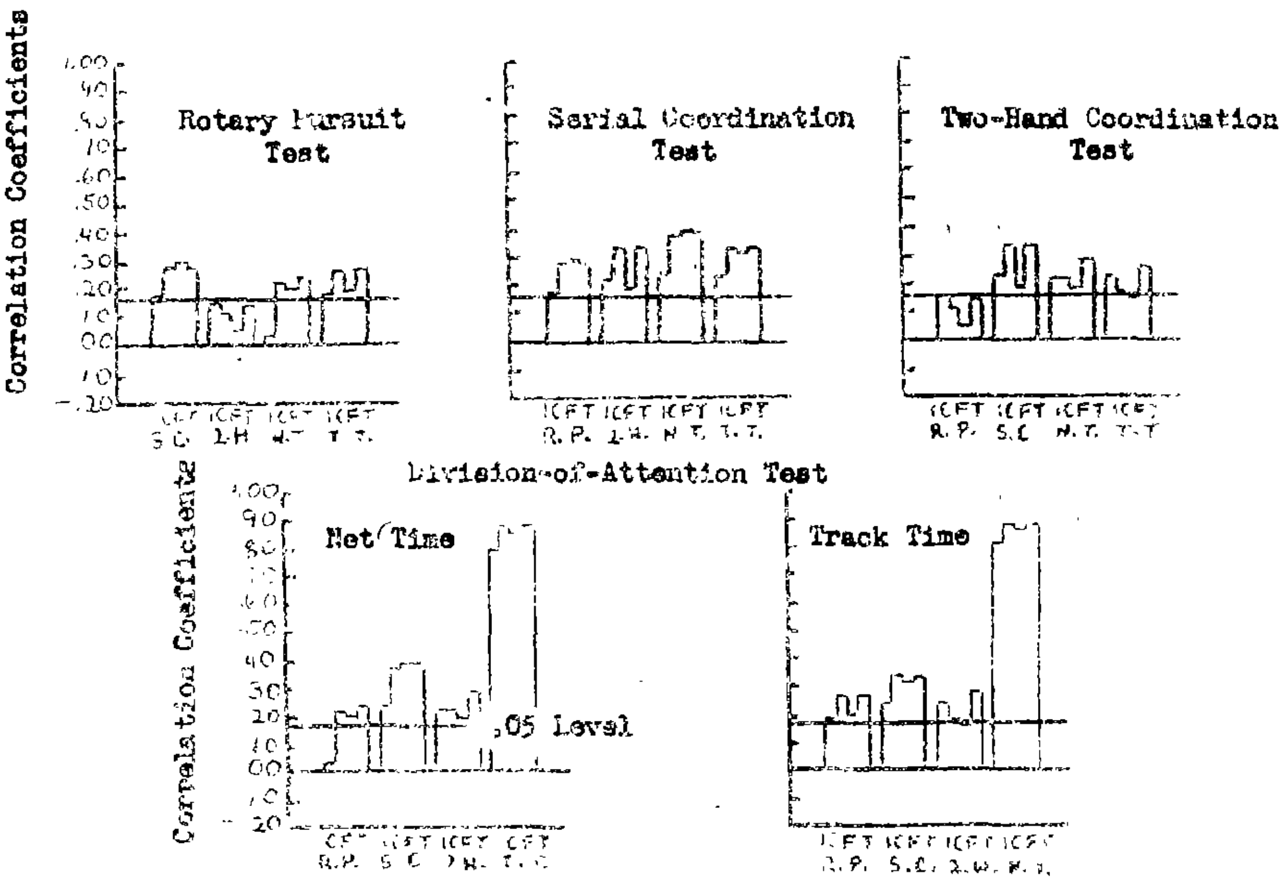

FIGURE 17

INTERCORRELATIONS OF EACH TEST WITH EVERY OTHER TEST BASED ON STATUS SCORES see Table 5). (3) The division-of-attention test is most closely related to the serial coordination test. This at first may seem appropriate, since common elements of pursuit, peripheral visual stimulation, and reaction time seem to exist. It must be remembered, however, that motor skills tests constructed previously on the common-element principle have not necessarily shown even moderate correlation, as pointed out by Seashore.[38] (4) Since net time and track time are so closely related in the division-of-attention test, it is probably safe and desirable to eliminate one of them because of its lower reliability. (Reaction time on the same test was eliminated from further consideration after the analysis of Group A.) Of the two measures, track time is probably the one to be discarded since, if it were retained in a practical selection situation, and word got around that "staying on the track" was all that counted, the test would very quickly lose its complication aspects, the new or significant feature of this test.

Relationships Between Pencil-and-Paper and Psychomotor Tests. Table [illegible] presents intercorrelations between the pencil-and-paper tests and total scores on the psychomotor tests. Comparison of columns shows that only three pencil-and-paper tests (Planning Circuits, Hidden Figures II, and Mechanical Comprehension) bear an appreciable relationship to the psychomotor tests, the highest coefficients resembling the highest coefficients found in the analyses of inter-test relationships of psychomotor tests. In general, the present correlations show a tendency to be slightly higher for Group B than for Group A, as were the test reliabilities (see Table 9).

Comparison of rows in the present table shows that no psychomotor test correlates more than moderately with pencil-and-paper tests, and, as might be expected, no psychomotor test shows correlations of consistent size with all pencil-and-paper tests. The rotary pursuit test tends to show the lowest correlations, and two-hand coordination, the highest. There is little further information in this table to indicate what tests might best be used in an inclusive battery.[39]

Intra-criterion Correlations and Criterion Analysis. As indicated in Table 11, all four criterion scores are quite closely interrelated. It has been mentioned previously that these correlations may be somewhat inflated. A and C are both based on maneuver ratings, A on all of them, C on those of the "last-10" lessons. Since C is contained in A, their intercorrelation is to a certain extent spuriously high. B and C are both based on the "last-10" lessons, C being maneuver ratings and B the average over-all percentage grade. The instructor supposedly derived B from C by a computation specified in the logbook; the failure of the two

---

[38] Seashore, R. H. Op. cit. (See Footnote 31 of this report.)

[39] The slight tendency of these complex psychomotor tests to relate to one of the hidden figures tests supports the suggestion that a space perception factor is to be found in motor skills performances. See: Seashore, R. H., Buxton, C. E., and McCollom, I. N. Op. cit. (See Footnote 31 of this report.)

TABLE 10

CORRELATIONS BETWEEN PENCIL-AND-PAPER TESTS AND TOTAL SCORES OF PSYCHOMOTOR TESTS

	Planning Circuits			Hidden Figures I			Hidden Figures II*			Path Distance			Visual Pursuit			Mechanical Comprehension		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
Rotary Pursuit	.21	.00	.12	.04	-.06	-.01	.12	-.01	.07	.07	.17	.14	.33	.13	.25	.12	.18	.16
Serial Coordination	.13	.26	.21	.15	.18	.15	.17	.22	.20	.04	.22	.15	.16	.13	.17	.27	.34	.32
Two-Hand Coordination	.27	.45	.36	.07	.01	.04	.27	.20	.24	.10	.25	.17	.24	.34	.28	.32	.47	.42
Division-of-Attention:																		
Set Time	.14	.41	.26	.13	.20	.15	.28	.33	.31	.19	.26	.24	.00	.12	.06	.29	.30	.29
Track Time	.07	.35	.20	.11	.11	.11	.32	.27	.30	.20	.26	.24	.11	.09	.12	.26	.31	.28
Reaction Time*	-.08			.07			-.13			-.07			.12			-.18		

*Reaction time coefficients should be interpreted with sign reversed.

TABLE 11

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS
OF CRITERIA*

	Criterion B			Criterion C			Criterion D
	A	B	A+B	A	B	A+B	A
Criterion A	.79	.80	.80	.83	.84	.83	.80
Criterion B				.89	.91	.90	.85
Criterion C							.78

	Criterion A			Criterion B			Criterion C			Criterion D
	A	B	A+B	A	B	A+B	A	B	A+B	A
Mean	49.4	49.7	49.5	50.0	50.0	50.0	50.2	50.0	50.2	50.0
Standard Deviation	3.9	4.0	3.9	8.0	7.8	7.9	4.4	4.7	4.5	9.8

*Criteria are defined on page 11 ff.

to correlate perfectly may signify either a slightly different meaning of the over-all grade than this, or failure to compute accurately. Criterion D, computed in a relatively simple manner, nevertheless correlates quite closely with the others. The relationships of Criterion D to other criteria were determined only for Group A, after which it was discarded.

Table 11 shows that the variabilities of the various criteria are by no means equal. For Criteria A and C especially, the variability was small (about two-thirds of the measures lying within four and five points of the mean, respectively). In terms of the original rating scale, the standard deviations would be approximately .4 and .5 points, along the 5-step scale applied to each maneuver rating.

Various efforts were made to determine the reliability of these criteria. If reliability is interpreted as consistency from rating to rating, the criterion data here used probably all rank high. Only one such index was computed -- a split-half coefficient for the mean of odd lessons vs. the mean of even lessons on Criterion C ("last-10" lesson maneuver ratings). The coefficient was .90. It may merely indicate, however, stereotyping of instructor judgment.

A more significant estimate of criterion reliability would be based upon inter-instructor correlations. The present data did not permit the straightforward correlation of two instructors rating the same men, due to insufficient number of cases. The closest approximation was the correlation between a given instructor's ratings of a group of cadets and ratings on the same cadets by several different instructors, each of whom rated but a few of the men. This was done for four "key" instructors on Criterion C, and the "coefficients of agreement" were: .72, .50, .19, and .67. Apparently some instructors agree fairly closely in their ratings, while others do not.

Correlations of Test Scores with Criteria. The large number of correlations between all tests and measures and the four criteria are summarized for comparative purposes in three figures.[40] In Figure 18 all types of score and all criteria but D are combined, so that the numbers of coefficients above the 5 per cent level of significance are an indication of the relative values of the several tests for predicting flight success. In terms of the combined groups, track time is first in the number of significant coefficients. Allowing for the negative coefficients for the two-hand coordination test, it leads in the tally for the combined groups. Taking into account the stability of the number of significant coefficients from group to group, the tests do not differ markedly. Reaction time on the division-of-attention test clearly is least closely related to the criteria. For this reason, analyses of the reaction time measure, already shown to be the least reliable of the three division-of-attention measures, were not obtained for Group B or the combined group. In general, comparison of the results on Groups A and B suggests that the prediction coefficients of Group B tend to be slightly higher than for Group A.

The task of evaluating Figure 18 is made easier by study of Figure 19. In this figure, all tests and measures, and all criteria, have been combined so that the numbers of coefficients above the 5 per cent level of significance are now indications of the relative predictive value of the types of scores. In terms of the number of such coefficients for the combined groups, all status scores are of about equal merit, while, in terms of the absolute size of coefficients, it is fairly clear that total score and central status score are the best predictors. Final status and initial scores are somewhat less valuable, more often showing values below the 5 per cent confidence level.

There is only a small amount of support for the hypothesis stated in the introduction to this report, viz., that improvement measures of test performance should be worthwhile predictors of a training criterion. The majority of the correlations between the gains and the criteria is insignificant. Obscured by the method of tallying used in Figure 19, however, are several points which may be discerned by re-examination of Figure 18 or by study of the individual gain coefficients (see Table E-5, Appendix E). First, most of the negative coefficients for C-F and I-F gains are based on the two-hand coordination test. This test seems to have had a relatively low "ceiling"; men who showed the larger gains were ones low on initial status and on flight criteria. Inspection of the raw data also suggested that men who exhibited a high level of skill by the time the central segment was reached sometimes had a loss, i.e., a negative gain, thereafter, and thus contributed to, or exaggerated, the negative relationship for the entire group. However, some of the significant negative C-F correlations occurred with the rotary pursuit and division-of-attention tests,

---

[40]The exact correlation coefficients are to be found in Table E-5, Appendix E. Reaction time was not employed in these tabulations, except in Figure 18. Criterion D was omitted, except in Figure 20, in order to make the summaries in each column refer to the same number of coefficients.

Test	Rotary Pursuit			Serial Coordination			Two Hand			Rat Time			Track Time			Reaction Time
Groups	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A
71-75																
66-70																
61-65											1					
56-60		1												1		
51-55																
46-50																
41-45					1											
36-40		1	1		1			2								
31-35		2		1	3	1	2	7	2	1	1		1	2		
26-30	3	5	2	4	3	3	2	2	6	2	7	4	3	5	2	
21-25	1	3	6	8	3	6	4	1	1	5	1	5	3	1	7	
16-20	7		3		1	2			2	4	1	3	7	6	5	
11-15	6		2	3		2	1		1	6	1	2	2	1	1	3
06-10	1	1	1	4		2	5			1	2	3	2	1	4	3
01-05		2	3			2		1	3	2	1	2	3		1	2
-04-00	1			1	1	3	1	2			1	1			1	3
-09-05	2	3			4			2				1		1		3
-14-10			3		2		1	3	1		3					3
-19-15		2			1			1	2					1		1
-24-20					1		1		3		1					2
-29-25							2				1			1		1
-34-30		1					1							1		
-39-35																
-44-40							1									
+rs Above 5 Per cent	4	9	12	13	8	12	8	11	11	8	9	12	7	8	14	0
-rs Above 5 Per cent	0	1	0	0	0	0	5	0	5	0	1	0	0	2	0	3
Insignificant rs	17	11	9	8	13	9	8	10	5	13	11	9	14	11	7	18

(Vertical axis label: Correlation Coefficients; boxed legend: -.05 LEVEL)

FIGURE 18

FREQUENCY DISTRIBUTIONS OF CORRELATION COEFFICIENTS OF PSYCHOMOTOR TESTS AND MEASURES WITH CRITERIA

Types of Scores	Initial			Central			Final			I-C			C-F			I-F			T		
Groups	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
61-65		1																			
56-60		1									1										
51-55																					
46-50																					
41-45		1																			
36-40			1																	4	
31-35		5		2	4	1	3	2									1			3	2
26-30	1	4	4	4	9	5	4	1	2										5	8	6
21-25	2	1	4	4	2	8	3	6	6	2		1				3			7		6
16-20	4	2	6	4		1	2	5	6	1	1		3			1		1	3		1
11-15	6			1					1	6	1	5				5		2			
06-10	2						2	1		3	3	5	1		1	3	1	4			
01-05										3	3	4	2	1	3			4			
-04-00							1				[illegible]		2	2	4		1	1			
-09-05											3		3	3	1		4				
-14-10													1	1	3	1	7	1			
-19-15											2			3	1			1			
-24-20														1	2	1	1	1			
-29-25													1	2		1					
-34-30													1	2							
-39-35																					
-44-40													1								
+rs Above .05 Level	3	12	15	10	13	15	10	3	14	2	1	1	0	0	0	3	1	1	12	15	15
-rs Above .05 Level	0	0	0	0	0	0	0	0	0	0	0	0	3	4	3	2	0	2	0	0	0
Insignificant rs	12	3	0	5	2	0	5	12	1	13	14	14	12	11	12	10	14	12	3	0	0

Correlation Coefficients

- .05 LEVEL

FIGURE 19

FREQUENCY DISTRIBUTIONS OF CORRELATION COEFFICIENTS OF THE VARIOUS PSYCHOMOTOR MEASURES WITH THE CRITERIA suggesting that some factor other than mere ceiling on performance is involved. Second, the significant positive correlations between gains and the criteria are mainly found in the serial coordination test. These coefficients, for the only test given under massed practice conditions, are the only evidence for the original hypothesis as mentioned above, but as a final point it should be remembered that the coefficients here discussed are based on relatively unreliable scores, i.e., in general, the gain scores exhibit low split-half reliability coefficients.

A further question is: Which criterion is best predicted by the various tests and measures? To answer this, Figure 20 was constructed, in which all tests and all types of score are combined, so that the numbers of coefficients above the 5 per cent level of significance are an indication of the relative efficiencies with which the various types of criteria are predicted. Criterion D, computed without correction for the mean and spread of the instructors' ratings, was least satisfactorily predicted in Group A. It was therefore omitted from further consideration. There is reason to believe that the weighting procedures employed in Criteria A, B, and C may have made it possible for the tests to show more predictive value; correlations with Criterion C (mean maneuver rating, "last-10" lessons) are the highest obtained. On the other hand, the fact that results for Groups A and B differ so much on this criterion measure suggests the desirability of the more laboriously obtained Criterion A, in view of its probable stability. Since Criterion B is predicted slightly less well than A, and also since it is a "last-10" measure and seemingly overlaps Criterion C, most of the following discussions place the greatest emphasis on Criteria A and C.

<u>Predictive Efficiency of Individual Psychomotor Tests</u>. This section presents bar-graphs indicating the extent to which each psychomotor test predicts the criterion measures. For this purpose the cadets of Groups A and B combined were classified into quintiles on test performance, total scores only being used, and the percentage of cadets in each quintile (N = 30 or 31) below either the lowest quartile or the median on the criterion was computed. Only two criteria, A and C, were used for this purpose.

In Figure 21 are shown the results of such an analysis for the rotary pursuit test. For example, of the 31 cadets falling in the lowest quintile (Quintile V) on total rotary pursuit score, 32 per cent fall below the lowest quartile score, and 65 per cent below the median on Criterion A. If scores on the rotary pursuit test were completely unrelated to standing on the criterion of the cadets in this quintile, about 25 per cent would fall below the quartile score on the criterion and 50 per cent below the median. Figure 21 suggests that: (1) rotary pursuit total scores discriminate on Criterion A only very slightly for the two cut-off points, the differentiation occurring in the lower quintiles on the test score; (2) on Criterion C, the discrimination for the quartile cut-off point is a little better than for the median, with the greatest effectiveness again exhibited in the lower test score ranges; (3) over the entire range of scores the two criteria and the two cut-off points are predicted with about equal accuracy, but in

Criterion	A			B			C			D		
Groups	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
61-65								1				
56-60								2				
51-55												
46-50												
41-45		1										
36-40		3			1				1			
31-35	2	6	3	1	3		2	6				
26-30	6	5	8	3	11	2	5	6	7	5		
21-25	4	3	7	10	3	11	7	3	7	2		
16-20	6	3	2	5	2	9	6	3	6	5		
11-15	7		3	7	1	3	4	1	2	11		
06-10	4	3	3	4		4	5	1	3	6		
01-05	2		3	1	2	5	2	2	3	3		
-04-00	2	2	1		2		1		2	2		
-09-05		3		1	4		[illegible]	3	1			
-14-10		2	1	1	3	2		3	1			
-19-15		2	1		1	1		2				
-24-20	1	1	[illegible]		1				2	1		
-29-25				1	1		1	1				
-34-30	1	1						1				
-39-35												
-44-40							1					
+rs Above .05 Level	12	15	20	14	15	20	14	15	21	7		
-rs Above .05 Level	2	1	2	1	1	1	2	2	2	1		
Insignificant rs	21	19	13	20	19	14	19	18	12	25		

Correlation Coefficients

[.05 LEVEL]

FIGURE 20

FREQUENCY DISTRIBUTIONS OF THE VARIOUS TYPES OF CRITERIA WITH THE PSYCHOMOTOR TESTS

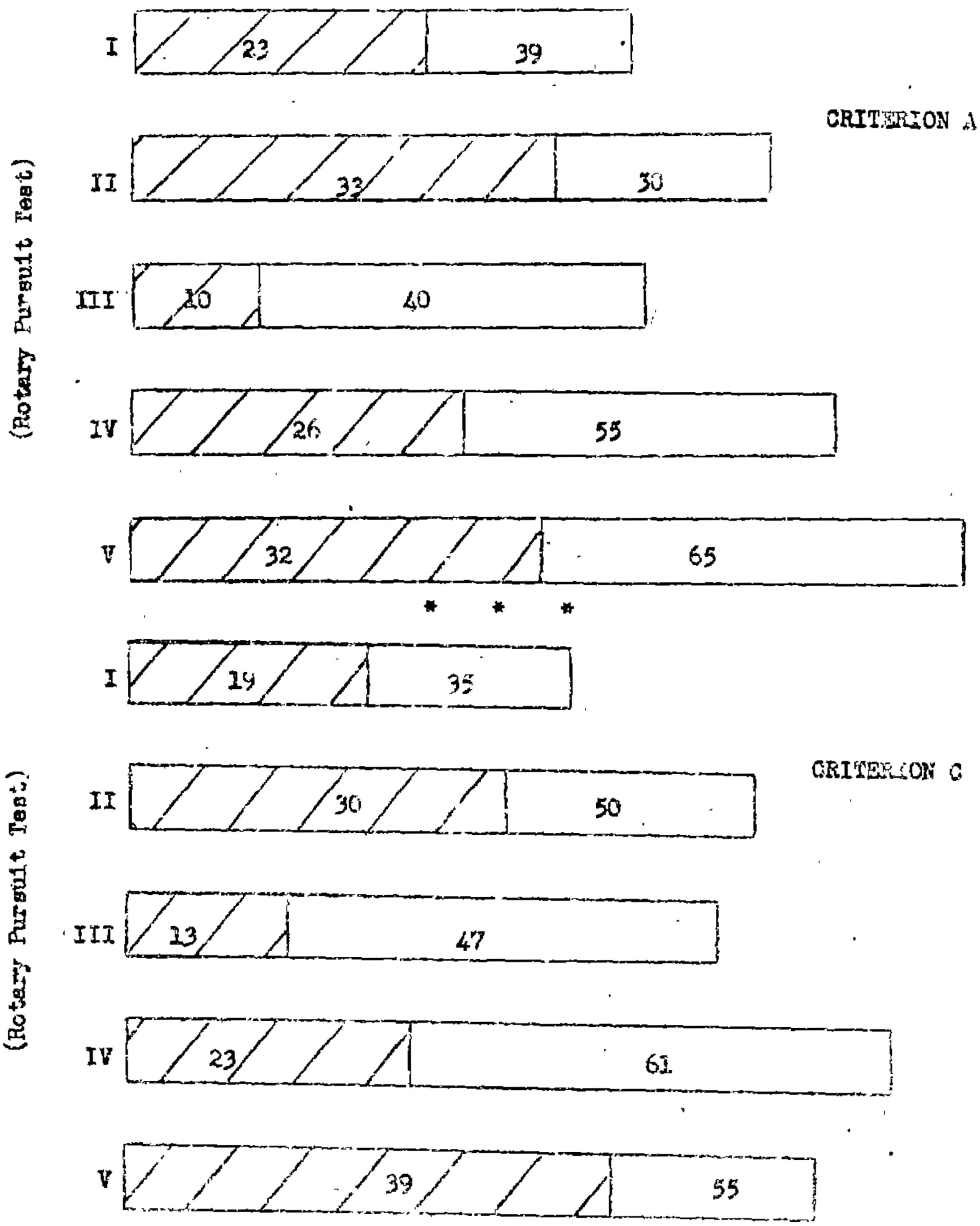

FIGURE 21

PERCENTAGES OF CADETS, IN EACH QUINTILE ON ROTARY PURSUIT TEST TOTAL SCORE, FALLING BELOW THE LOWEST QUARTILE SCORE ON THE CRITERION (CROSS-HATCHED BARS) OR THE MEDIAN SCORE ON THE CRITERION (OPEN BARS) neither case is the prediction very good. The probable difference in predictive efficiency of upper and lower ranges of scores is worth mention, however, since it denotes a curvilinear relation between test and criterion. This is confirmed in the etas between rotary pursuit total score and Criteria A and C, which are .51 and .60, as compared with r's of .25 and .26, respectively.

The predictive efficiency of the serial coordination test is shown in Figure 22. It may be noted that: (1) both criteria are discriminated more sharply than was the case for the rotary pursuit tests; (2) for the quartile cut-off point the discrimination is sharpest in the lower ranges of test scores, on Criterion A, but (possibly) in the upper score ranges on Criterion C; (3) for the median cut-off point, the upper and lower ranges of scores seem almost equally discriminating; (4) a curvilinear relationship may exist for Criterion A (eta = .47, r = .31), but probably not for Criterion C (eta = .52, r = .49).

Inspection of Figure 23 on the two-hand coordination test shows that: (1) the lower ranges of test scores are the most discriminatory for either criterion or for either cut-off point; (2) over the whole range of test scores, neither criterion or neither cut-off point can be said with any confidence to be predicted more efficiently. If there is any difference the graph shows it to be in favor of Criterion A. There is more of a curvilinear relation between the two-hand coordination test and Criterion C (eta = .48, r = .26) than between the test and Criterion A (eta = .45, r = .33).

Figure 24 presents the results of a similar analysis of net time scores on the division-of-attention test. It is probable that: (1) this test discriminates both criteria better in the upper ranges of test score; (2) Criterion A is discriminated better than Criterion C; (3) in general, the test predicts the median cut-off point slightly better than the quartile cut-off point; (4) there is some curvilinearity in the relation between the division-of-attention test and Criterion A (eta = .44, r = .27) and rather marked curvilinearity in the relation between the test and Criterion C (eta = .60, r = .19).

In summary, it can be stated that: (1) on the assumption that a rectilinear relationship between test and criterion is simplest to deal with, i.e., that r is the desired index of test-criterion relationship, Criterion A (mean over-all flight maneuver rating) is predicted more efficiently than Criterion C (mean maneuver rating, "last-10" lessons) in three of the four cases;[41] (2) on the assumption that the shape of regression lines is secondary in importance, i.e., that eta is the desired index of relationship,

[41]In other words, when the most reliable test score is employed, the criterion which in general has been declared previously to be most stable, etc., is the one predicted best.

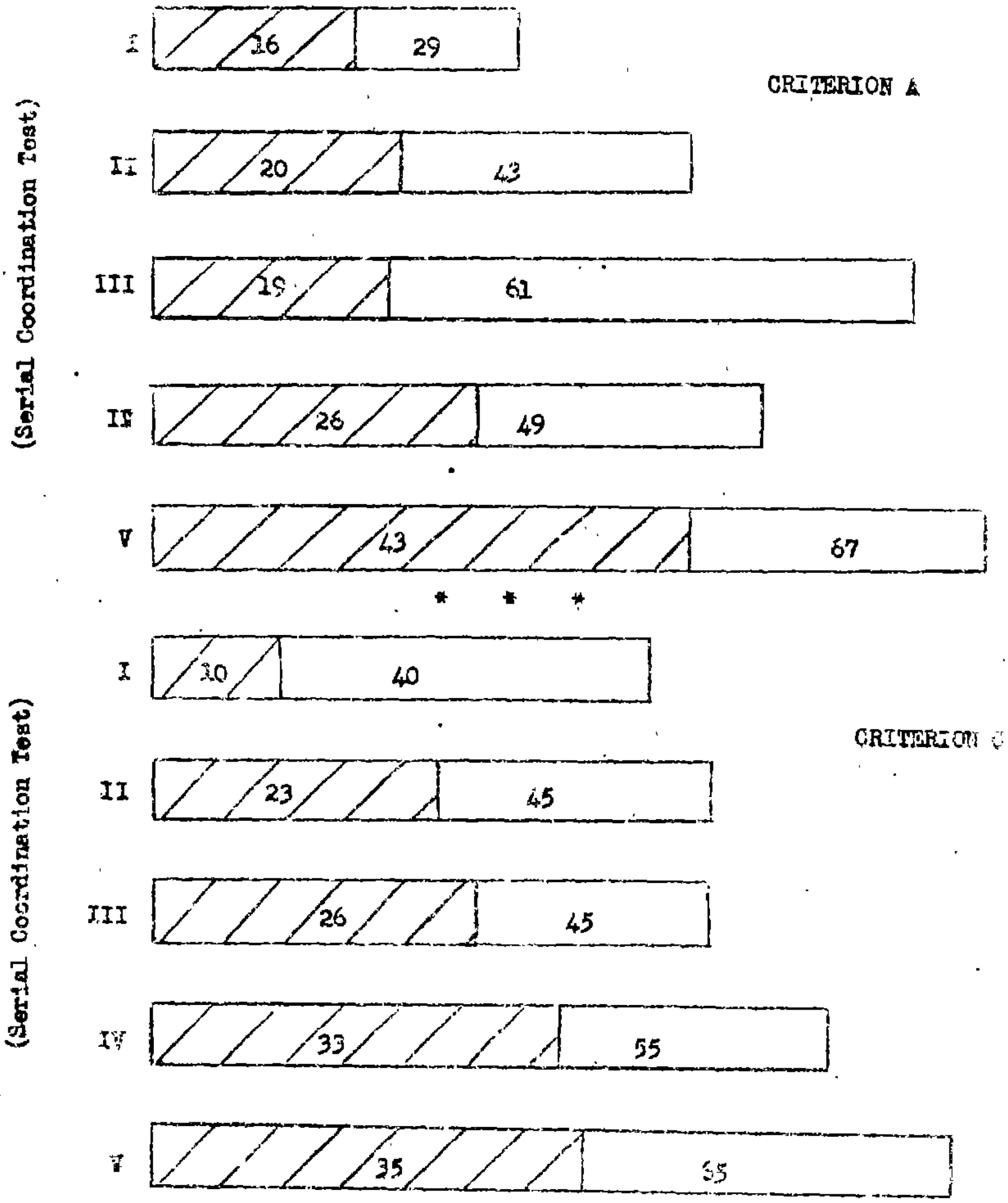

FIGURE 22

PERCENTAGES OF CADETS, IN EACH QUINTILE ON SERIAL COORDINATION TEST TOTAL SCORE, FALLING BELOW THE LOWEST QUARTILE SCORE ON THE CRITERION (CROSS-HATCHED BARS) OR THE MEDIAN SCORE ON THE CRITERION (OPEN BARS)

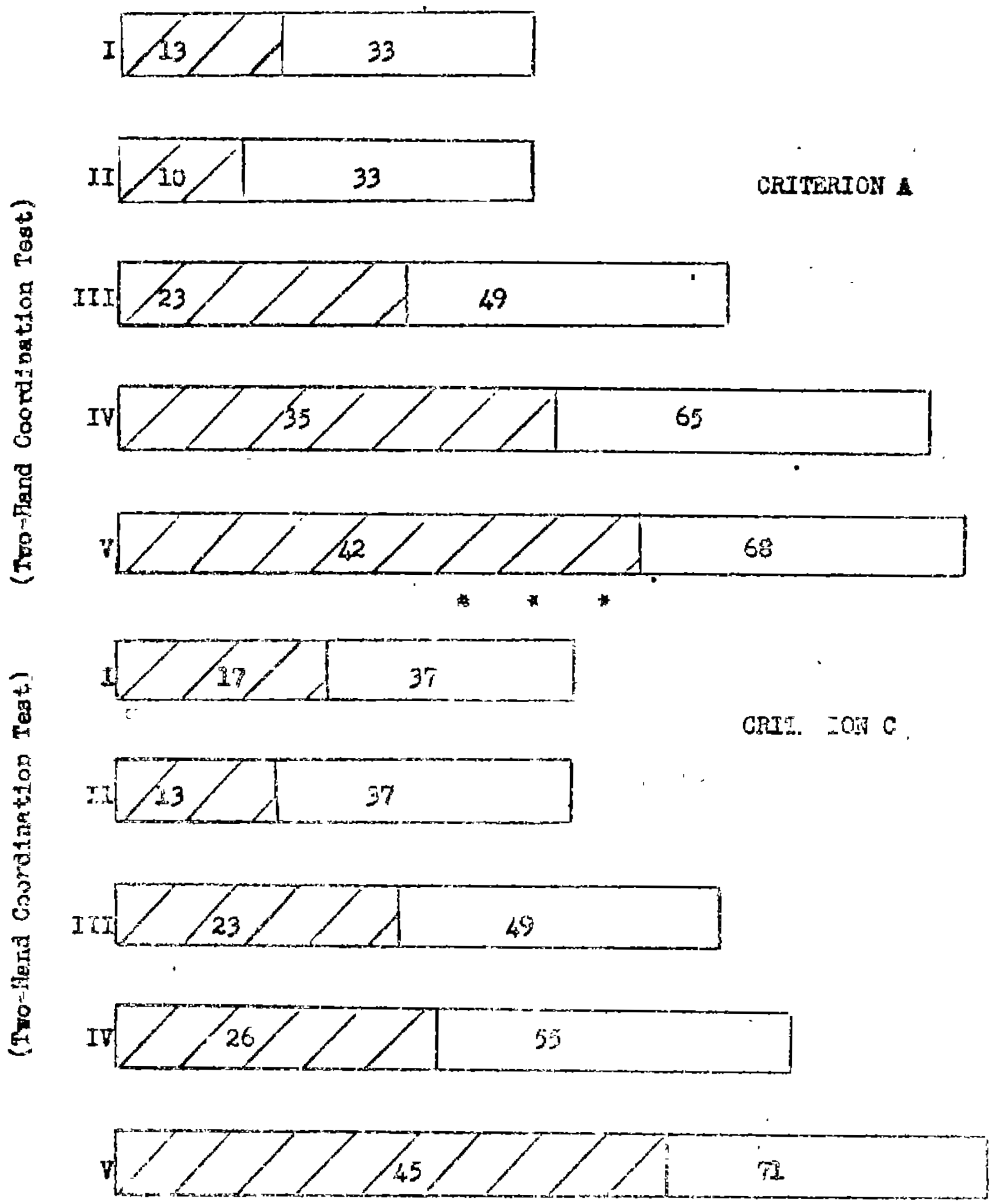

FIGURE 23

PERCENTAGES OF CADETS IN EACH QUINTILE ON TWO-HAND COORDINATION TEST TOTAL SCORE, FALLING BELOW THE LOWEST QUARTILE SCORE ON THE CRITERION (CROSS-HATCHED BARS) OR THE MEDIAN SCORE ON THE CRITERION (OPEN BARS)

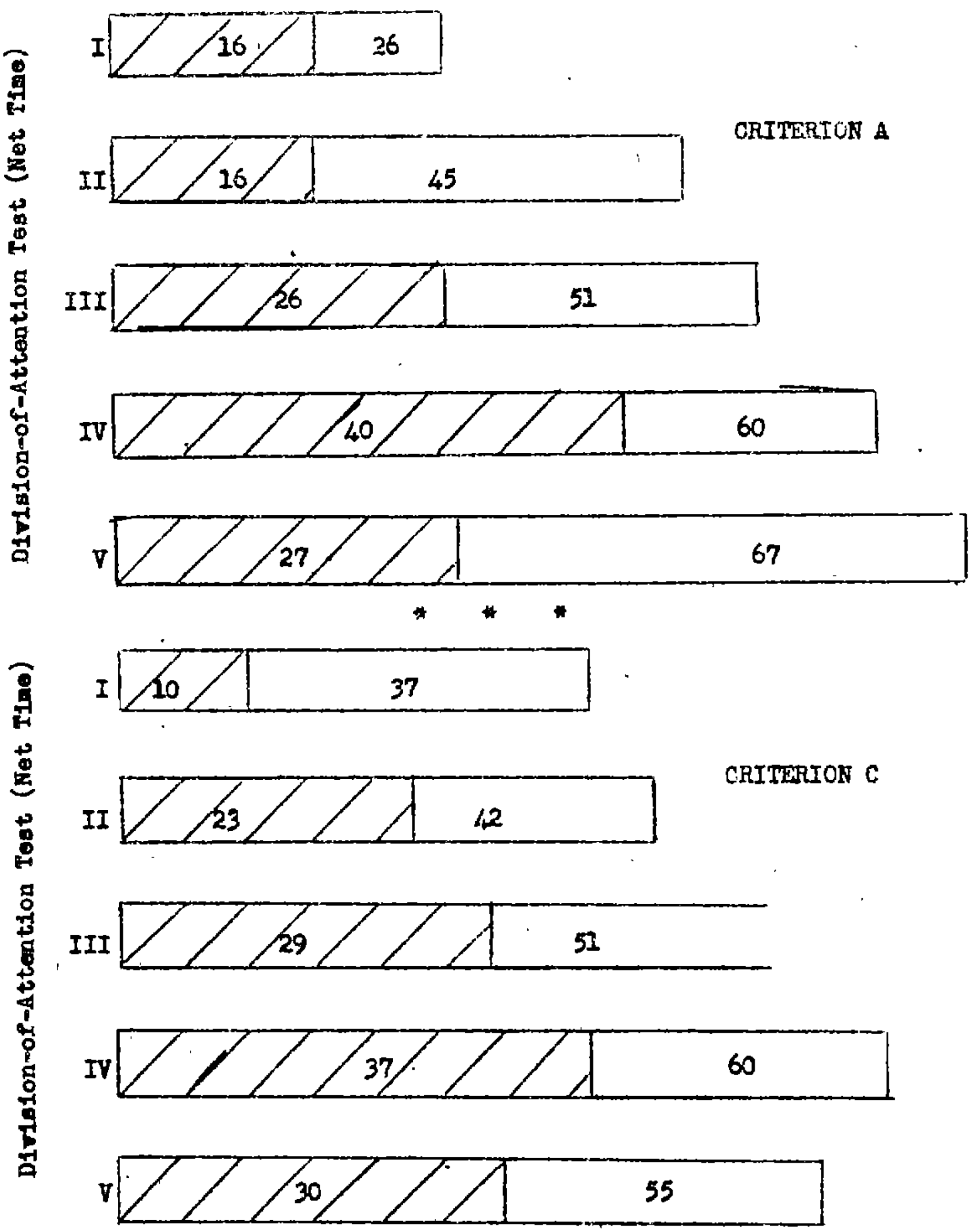

FIGURE 24

PERCENTAGES OF CADETS IN EACH QUINTILE ON NET TIME TOTAL SCORE, FALLING BELOW THE LOWEST QUARTILE SCORE ON THE CRITERION (CROSS-HATCHED BARS) OR THE MEDIAN SCORE ON THE CRITERION (OPEN BARS)

Criterion C is in every case predicted better than A;[42] (3) it is difficult to decide whether the quartile or median cut-off points on the criteria are discriminated more sharply; (4) the tests vary in predictive efficiency according to the range of test scores employed, e.g., scores on the rotary pursuit test and two-hand coordination test are more discriminating at the lower limits, net time scores are probably more discriminatory in their upper ranges, depending upon the criterion to be predicted, and serial coordination test scores are about equally effective over their entire range.

<u>Correlations Between Pencil-and-Paper Tests and the Criteria</u>. In Table 12 are recorded the predictive coefficients for each sub-test of the pencil-and-paper group, for the total score obtained by simple addition of the raw scores of the sub-tests, and for certain pairs of sub-tests combined by simple addition.

Four observations may be made with respect to this table: (1) The mechanical comprehension test is the most effective predictor, with hidden figures I and path distance the only other promising tests. That these two latter tests especially are worthy of further attention is emphasized if it is recalled that they are the least reliable tests in the pencil-and-paper battery, and the range of scores on hidden figures I is fairly restricted. The visual pursuit and hidden figures II are clearly almost valueless. Possibly the planning circuits test merits further trial, but this is not at all clear. (2) Criterion D probably is predicted with the least efficiency, as in the case of the psychomotor tests, and was therefore not used for Group B. It is likely that Criterion A is predicted best, but Criterion C is a close second. (3) The total score obtained by addition of sub-test scores is less efficient as a predictor than simple sums for either of two combinations of the best sub-tests, and less efficient than the mechanical comprehension test taken singly. (4) A comparison of single pencil-and paper tests with single psychomotor tests indicates that the best measures in each group are about on a par as predictors, with the psychomotor tests having a slight but consistent margin of superiority.

<u>Revised Scoring and Further Validity Coefficients</u>. After all the previous analyses had been made for Group A, an attempt was made to improve the predictive efficiency of the Group A scores on psychomotor tests by (1) "shortening" each test to include no trials beyond the most promising segment, the initial status score; (2) including the previously omitted initial trials on the rotary pursuit, two-hand coordination, and division of attention tests; and (3) computing a new total and a new improvement score for each test, where the index was per cent gain (to reduce the influence of differences in initial status).

---

[42] To what extent the curvilinearity is a function of the relatively small sample employed (N = 153) cannot be determined, but it has been suggested that curvilinearity is found with somewhat similar test conditions and very large samples. See: Britton, A. F. The selection of pilots by means of psychomotor tests. *J. Aviation Med.*, 1944, 15, 115-123.

TABLE 12

CORRELATIONS BETWEEN PENCIL-AND-PAPER TESTS AND CRITERIA

Criterion	Planning Circuits			Hidden Figures I			Hidden Figures II			Path Distance			Visual Pursuit		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
A	.13	.24	.19	.20	.09	.16	.03	.01	.03	.14	.37	.24	.04	.00	.03
B	.15	.08	.12	.22	-.01	.12	.11	.09	.10	.14	.30	.21	.08	-.02	.04
C	.15	.05	.10	.32	.08	.22	.13	.01	.08	.18	.25	.20	.14	.04	.09
D	.18			.16			.05			.11			.14		

Criterion	Mechanical Comprehension			Hidden Figures I + Mechanical Comprehension			Path Distance + Mechanical Comprehension			Total Score		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
A	.27	.54	.40	.32	.49	.41	.49	.55	.41	.23	.39	.31
B	.28	.44	.35	.35	.41	.37	.41	.34	.36	.29	.24	.26
C	.25	.44	.34	.34	.38	.35	.28	.41	.34	.28	.26	.27
D	.17			.21						.22		

The results of these analyses are shown in Table 13.[43] With few exceptions, the revised gain scores -- raw gain during the "shorter" test, or per cent gain -- are of lower predictive efficiency than the original gain scores. The revised total score likewise is in the main inferior to other status scores. The only probable improvement in the revised scoring seems to be in the two-hand coordination test, where inclusion of all early trials in the total score very slightly improves the prediction over the original central status and total score predictions. The revision was not attempted for Group B or for combined groups.

Test Order and Predictive Efficiency. It will be recalled that two orders were employed in testing Groups A and B. In order I, the rotary pursuit test came first, then the serial coordination, two-hand coordination, and division-of-attention tests. In order II, the rotary pursuit test was followed by the division-of-attention, two-hand coordination, and serial coordination tests. In Table 14 are shown the predictive coefficients for the two test-order groups.

In only one instance (the two-hand coordination test) does the difference approach significance (2 per cent level of confidence). Considering the difference on the (initial) rotary pursuit test, it seems unwise to attribute what is observed in Table 14 to anything other than sample differences, i.e., there is no evidence here to show that predictive efficiency of a psychomotor test is a function of which tests precede it in a battery.

Multiple Correlation Analyses. To determine the relative value of various combinations of tests as predictors of flight success, certain tests and scores were combined in multiple correlations against the criterion. Only the most promising single indices and only Criteria A and C were used for this purpose, since test battery construction was not a primary objective of this project.

The results of these analyses are presented in Table 15.[44] Some general observations may be made first. (1) In Group A, the correlations with Criterion C tend in general to be higher than those with Criterion A, but the trend is reversed in Group B. For the combined groups, Criterion A is probably predicted more adequately. (2) Within Criterion A, the trend is toward higher coefficients for Group B than for Group A. Within Criterion C the reverse tends to be true. It is important that there can be such a systematic change in the prediction coefficients from group to group.

Other interpretations of Table 15, drawn from comparisons of selected rows may be noted. (1) As shown in rows 1, 2, 3, psychomotor total scores

---

[43] Revised scoring of reaction time on the division-of-attention test was not carried out.

[44] The additional zero-order r's necessary for computing these multiple correlation type coefficients are recorded in Table E-6, Appendix E.

TABLE 13

COMPARISONS OF PREDICTION COEFFICIENTS BASED ON
REVISED AND ORIGINAL SCORINGS, GROUP A
(N = 90)

Test	Revised Scoring	Criterion A Rev.	Criterion A Orig.*	Criterion C Rev.	Criterion C Orig.*	Criterion B Rev.	Criterion B Orig.*	Criterion D Rev.	Criterion D Orig.*
Rotary Pursuit	Σ, 3-14	.09	.16	.15	.20	.21	.26	.12	.13
	Seg. 2 Seg. 1	-.06	.12	.05	.15	.11	.17	-.02	.07
	% Gain	-.12		.02		.03		-.10	
Serial Coordination	Σ, 1-6	.24	.27	.17	.24	.22	.28	.17	.26
	Seg. 3 Seg. 1	.10	.12	.12	.21	.13	.22	.11	.22
	% Gain	.02		.07		.07		.06	
Two-Hand Coordination	Σ, 3-22	.32	.29	.22	.22	.27	.23	.28	.28
	Σ, Trials 1-22	.33		.23		.28		.29	
	Seg. 3 Seg. 1	.06	.07	.09	.09	.06	.10	.11	.11
	% Gain	.06		-.01		-.05		.03	
Division-of-Attention: Net Time	Σ, 3-14	.19	.25	.09	.24	.08	.16	.05	.18
	Σ, Trials 1-14	.15		.03		.05		.00	
	Seg. 3 Seg. 1	.11	.14	.23	.14	.14	.15	.19	.15
	% Gain	.02		.14		.11		.12	
Track Time	Σ, 3-14	.25	.30	.17	.23	.17	.21	.14	.16
	Σ, Trials 1-14	.24		.17		.17		.11	
	Seg. 3 Seg. 1	.05	.01	.10	.04	.03	.02	.03	-.01
	% Gain	.00		.03		-.01		-.03	

*Original values listed on the same line with a new total score are original total-score correlations; original values listed on the same line with a new gain score, raw or per cent, are the original I-C gain correlations.

TABLE 14

CORRELATIONS OF TOTAL SCORES ON PSYCHOMOTOR TESTS WITH CRITERION A, IN TWO TEST-ORDER GROUPS

Order	N*	Rotary Pursuit	Serial Coordination	Two-Hand Coordination	Division-of-Attention
I	72	.18	.29	.22	.18
II	79	.27	.31	.44	.27

*The test-order notation was missing on two records.

are somewhat better predictors of the criteria than are central or final status scores. The generalization is limited, however, by the fact that the superiority occurs mainly in the Group B results. (2) As seen in rows 4 and 5, the three-variable multiple including track time (final status score) tends to differ little in predictive significance for all groups on both criteria from the three-variable multiple including net time (final status score). (3) One combination of gain scores -- the most promising of those obtained for Group A -- does not stand up in successive samplings (see row 6). There is a suggestion that a status score and a gain score combined, on the serial coordination test (see row 7) has some stability. However, it is doubtful whether the finding will stand up in later investigations, since the C-F gain involved has a reliability of approximately zero. (4) The combination of the three best predictors among the pencil-and-paper tests (see row 8) is about equal to the best psychomotor measures (see rows 1, 2, 3). (5) A combination of two of the best pencil-and-paper tests with status scores from psychomotor tests (see rows 9, 10, 11) is about as efficient in predicting the criteria as is a combination of two psychomotor scores and one pencil-and-paper measure (see rows 12, 13, 14). (6) A five-variable multiple has little more predictive value than certain three-variable combinations.

It is of interest to note that a number of the multiple correlation coefficients with Criterion A in Table 15 are in the .40's, for the combined groups, and many of the coefficients with Criterion C are in the .30's. Furthermore, for group B on both criteria certain coefficients reach the level of .50. In a recent study by the AAF it has been indicated that the value of .50 is rather typical for combinations of 12 pencil-and-paper tests or six psychomotor tests in AAF work. Coefficients for smaller numbers of tests or for specific tests have not been published, but it is quite possible that the multiple correlation coefficients found in this study approximate to some extent those obtained by the AAF.[45]

One further evaluation of the multiple correlation coefficients is shown in Table 16, in which are presented multiple correlations for Groups A and B, obtained by the application of the beta weights from each sample to the other. The maximized multiples of Table 15 are repeated to make it

[45]See: Staff. Psychological activities in the training command, Army Air Forces. Psychol. Bull., 1945, 42, 37-54.

TABLE 15

MULTIPL 'RRELATIONS BETWEEN CRITERIA A AND C AND 'TED TEST OR SCORE COMBINATIONS

	Test	Criterion A Group A	Criterion A Group B	Criterion A Group A+B	Criterion C Group A	Criterion C Group B	Criterion C Group A+B
1.	Rot. Pur. (C)*, Ser. Coord. (C), Two-Hand (C)	.37	.44	.39	.45	.36	.39
2.	Rot. Pur. (F), Ser. Coord. (F), Two-Hand (F)	.37	.38	.37	.45	.31	.37
3.	Rot. Pur. (T), Ser. Coord. (T), Two-Hand (T)	.37	.51	.42	.39	.41	.38
4.	Ser. Coord. (F), Two-Hand (C), Net Time (F)	.40	.36	.38	.43	.27	.34
5.	Ser. Coord. (F), Two-Hand (C), Track Time (F)	.44	.34	.40	.42	.27	.35
6.	Ser. Coord. (I-C), Two-Hand (C-F)	.34	.16	.21	.47	.09	.26
7.	Ser. Coord. (T), Ser. Coord. (C-F)	.29	.39	.31	.33	.34	.26
8.	Path Dist., Mech. Comp., Hid. Fig. I	.31	.56	.42	.39	.45	.38
9.	Hid. Fig. I, Mech. Comp., Ser. Coord. (F)	.35	.55	.43	.45	.45	.39
10.	Hid. Fig. I, Mech. Comp., Two-Hand (C)	.39	.56	.46	.46	.45	.42
11.	Hid. Fig. I, Mech. Comp., Net Time (F)	.36	.55	.43	.39	.44	.38
12.	Ser. Coord. (F), Two-Hand (C), Hid. Fig. I	.38	.35	.37	.49	.28	.39
13.	Ser. Coord. (F), Two-Hand (C), Mech. Comp.	.39	.56	.46	.43	.45	.41
14.	Ser. Coord. (F), Two-Hand (C), Path Dist.	.37	.46	.40	.44	.34	.37
15.	Net Time (F), Mech. Comp., Path Dist.	.33	.56	.43	.30	.44	.39
16.	Ser. Coord. (T), Hid. Fig. I, Mech. Comp.	.36	.58	.45	.42	.48	.41
17.	Rot. Pur. (T), Ser. Coord. (T), Two-Hand (T), Net Time (T), Mech. Comp.	.41	.63	.49	.42	.51	.43

*C (central status score); F (final status score); T (total score).

TABLE 16

MULTIPLE CORRELATIONS WITH THE CRITERIA OBTAINED BY APPLYING BETA WEIGHTS FOR EACH SAMPLE TO THE OTHER SAMPLE

		Criterion A				Criterion C			
		Group A		Group B		Group A		Group B	
		Beta Weights		Beta Weights		Beta Weights		Beta Weights	
	Test	Gr. A	Gr. B	Gr. B	Gr. A	Gr. A	Gr. B	Gr. B	Gr. A
1.	Rot. Pur. (C), Ser. Coord. (C), Two-Hand (C)	.37	.36	.44	.43				
2.	Rot. Pur. (F), Ser. Coord. (F), Two-Hand (F)	.37	.36	.38	.37				
3.	Rot. Pur. (T), Ser. Coord. (T), Two-Hand (T)	.37	.34	.51	.49	.39	.38	.41	.42
4.	Ser. Coord. (F), Two-Hand (C), Net Time (F)	.40	.40	.36	.35				
5.	Ser. Coord. (F), Two-Hand (C), Track Time (F)	.44	.40	.34	.33				
6.	Ser. Coord. (I-C), Two-Hand (C-F)	.34	.01	.16	.04				
7.	Ser. Coord. (T), Ser. Coord. (C-F)	.29	.24	.39	.32	.33	.23	.34	.24
8.	Path Dist., Mech. Comp., Hid. Fig. I	.31	.31	.56	.51	.38	.24	.45	.30
9.	Hid. Fig. I, Mech. Comp., Ser. Coord. (F)	.35	.28	.55	.44				
10.	Hid. Fig. I, Mech. Comp., Two-Hand (C)	.39	.32	.56	.48				
11.	Hid. Fig. I, Mech. Comp., Net Time (F)	.36	.29	.55	.45				
12.	Ser. Coord. (F), Two-Hand (C), Hid. Fig. I	.38	.38	.35	.34				
13.	Ser. Coord. (F), Two-Hand (C), Mech. Comp.	.39	.33	.56	.49				
14.	Ser. Coord. (F), Two-Hand (C), Path Dist.	.37	.32	.46	.42	.44	.41	.34	.32
15.	Net Time (F), Mech. Comp., Path Dist.	.33	.31	.56	.51	.30	.26	.44	.40
16.	Ser. Coord. (T), Two-Hand (C), Path Dist.	.36	.31	.58	.50	.42	.29	.48	.34
17.	Rot. Pur. (T), Ser. Coord. (T), Two-Hand (T), Net Time (T), Mech. Comp.	.41	.36	.63	.57	.41	.36	.51	.46

easier to observe the shrinkage. Only a few of the test combinations are included as a check on correlations with Criterion C.

Three points can be made concerning the correlations[46] with Criterion A in Table 16. First, the single instance of greatest shrinkage (see row 6) is in the multiple correlation of the two gain scores, serial coordination test (I-C) and two-hand coordination test (C-F), with this criterion. Second, particularly in regard to Criterion A the shrinkage appears somewhat less when all the tests in a combination are either pencil-and-paper (see row 8) or psychomotor (see rows 1 to 5, inclusive) than when the two types of test are represented in a given combination (see rows 9 to 14, inclusive). Third, the correlations based on the five-test battery did not show less shrinkage than the correlations based on the three-test battery.

In summary of these analyses it may be said that, just as the zero-order correlations with the criteria did not readily permit choices among the predictors, the multiple correlations suggest that several different combinations of measures are of about equal value. From the point of view of test economy, it is probable that some pencil-and-paper measures should be combined with a few psychomotor measures. Apparently the exact choice of measures and scores will depend in part on the type of scoring of the flight criterion. Certain of the multiple correlations show some shrinkage when beta weights are applied to another sample; the amount of shrinkage is apparently in part a function of the reliability of the zero-order r's.

SUMMARY

Four complex psychomotor tests and a pencil-and-paper test containing six sub-tests were administered to 153 War Training Service cadets. Data

---

[46] Dr. Leon Festinger secured the multiples with the new weightings according to the formula derived below. Since the samples were relatively small, weightings in z-score form were applied, to reduce the variability due to fluctuation of sample sigmas.

$$\tilde{z}_c = b_1 z_1 + b_2 z_2 + \ldots\ldots\ldots\ldots b_n z_n$$

where $\tilde{z}_c$ = predicted criterion in z-score form, and b = arbitrary weightings

$$R_{0.1,2,\ldots n} = r_{z_o \tilde{z}_c} = \frac{b_1 \Sigma z_o z_1 + b_2 \Sigma z_o z_2 + \ldots b_n \Sigma z_o z_n}{N \sigma_{\tilde{z}_c} \sigma_{z_o}}$$

$$= \frac{b_o r_{o1} + b_2 r_{o2} + \ldots\ldots b_n r_{on}}{\sigma_{\tilde{z}_c}}$$

$$= \frac{\Sigma b_i r_{oi}}{\sqrt{\Sigma b_i^2 + 2\Sigma b_i b_j r_{ij}}}$$

on elementary flight training were also secured. An attempt was made to evaluate and compare the several tests, including various methods of scoring performances on the psychomotor tests, as instruments for predicting flight training success. The pencil-and-paper tests were presumed to measure aspects of perceptual skill and of mechanical comprehension, and the psychomotor tests were presumed to measure such factors as steadiness, reaction time, smoothness of control, and speed of decision, as components in complex performance.

Psychomotor tests employed, and lengths of work and rest periods on each, were as follows: rotary pursuit test, 10 seconds work, 20 seconds rest; serial coordination test, 120 seconds work, 15 seconds rest; two-hand coordination test, 30 seconds work, 30 seconds rest; division-of-attention test, 30 seconds work, 30 seconds rest (net time, track time, and reaction time measures were taken simultaneously on the same performance). Each test was scored in seven ways: initial, central, and final status scores, total score, early gain, later gain, and over-all gain. The serial coordination test practice was relatively massed, all others relatively distributed.

Included in the pencil-and-paper test were the sub-tests: planning circuits; hidden figures I, II; path distance; visual pursuit; mechanical comprehension.

Criterion data consisted of ratings (on a 5-point scale) for every maneuver during Stages A and B of elementary flight training, and an overall percentage grade for the 5 to 20 maneuvers of each lesson. Four scorings of these data were developed: mean flight maneuver rating, mean percentage grade on "last-10" lessons, mean maneuver rating on "last-10" lessons (all three being corrected for instructor's mean and variability), and uncorrected mean percentage grade on "last-10" lessons.

A general summary of the findings of this study is as follows:

1. A study of psychomotor performance records over 25 to 40 minutes of practice and rest on various tests showed that (a) there was clear-cut improvement in average score; (b) all curves were negatively accelerated; (c) a limit apparently was being approached in some cases.

2. Split-half reliability coefficients for successively longer segments of the psychomotor tests showed that (a) in three cases adequate reliability (.90 or better) was reached halfway through the test period or earlier, viz., after 16 trials on the rotary pursuit test, 6 on the serial coordination test, and 16 on the division-of-attention test; (b) the two-hand coordination test, full length (30 trials), had a split-half coefficient of only .73.

3. Correlations of successively longer segments with total score showed, nevertheless, that the relative unreliability of the two-hand coordination test did not prevent its correlations from reaching the level of .90 as early as for all other tests, i.e., at about half-length, or less.

4. Split-half reliability coefficients for the various types of psychomotor scores indicated that (a) gain scores, especially for late gains, were in general not very reliable; (b) initial status was less reliable than the central or final status score, but all three were reasonably stable scores; (c) total score was quite reliable except for the two-hand coordination test.

5. Intra-test correlations for the various psychomotor tests indicated that (a) status scores are more closely intercorrelated than gain scores; (b) the degree and algebraic sign of the correlation between any two types of score is apparently dependent upon both the communality or overlapping of the data on which they are based, and their comparative positions on the learning curve.

6. Analyses of sub-tests in the pencil-and-paper group showed that (a) there was a range from low to moderately high reliabilities and intercorrelations; (b) the highest correlation was between the two hidden figures tests and lowest between the visual pursuit test and the others; (c) there was a tendency for path distance, planning circuits, and the mechanical comprehension test to form a cluster.

7. Findings with respect to inter-test correlations among the psychomotor tests showed that (a) none of the tests intercorrelated at more than a moderately low level -- possibly as a function of the use of distributed practice on all but the serial coordination test; (b) the level of intercorrelation tended to rise somewhat for the serial coordination test on which practice was massed; (c) inter-test correlations based on gain scores tend to be lower than coefficients based on status scores.

8. Studies of the relationships between pencil-and-paper tests and psychomotor tests showed that (a) the sub-tests tended in most instances to correlate positively, but not highly, with total scores on the psychomotor tests; (b) no psychomotor total score correlated higher than others with the pencil-and-paper tests in general; (c) the "inter-battery" coefficients, rather low as they are, tend to resemble the other inter-test correlations here reported.

9. Criterion scores based on all flight maneuver ratings, on the maneuver ratings for the "last-10" lessons, and the over-all percentage grades given by instructors on the second of these, intercorrelated at the level of about .80 and above, i.e., all the methods of scoring flight data which were retained throughout the study resulted in somewhat the same ranking for the cadets. However, the correlations between criteria were low enough for differences in predictability of the several criterion scores to appear. Checks on the quality of criterion data indicated that (a) instructors show variance in amount of agreement on a given cadet; (b) a given instructor tends repeatedly to give similar judgments on a cadet; (c) instructors differ considerably in the mean and spread of the ratings made on their respective students; (d) the widest possible range of ratings is not utilized.

10. An analysis of psychomotor test criterion correlations showed that (a) when all criterion scores excepting D are combined and various psychomotor measures correlated with these criteria, track time for groups A and B combined is first in the number of significant coefficients, while the two-hand coordination test (allowing for negative coefficients) leads in the tally for the combined groups, the reaction time measure is least closely related to the criteria; (b) the most promising predictor among the types of score was the total score, with central status scores ranking very close to it; (c) gain scores have little promise as predictors of the training criterion, perhaps because gain scores are less reliable than other types of score, or because learning rate is so specific to any given test that test and criterion could not have a learning factor in common; (d) Criterion D -- "uncorrected" percentage grades on the "last-10" lessons -- is clearly less readily predicted than other criterion measures, while Criterion A (mean flight maneuver rating) is probably predicted best with C close behind (mean maneuver rating "last-10" lessons); (e) none of the tests shows a clearly rectilinear relationship with the two Criteria A and C, examined for this purpose; and (f) serial position of a test in the battery had no demonstrable effect on its predictive efficiency.

11. Among the pencil-and-paper devices, the mechanical comprehension test, the hidden figures I and mechanical comprehension tests, and the path distance and mechanical comprehension tests offered the most promise as predictors of flight success, and were almost on a par with the best psychomotor predictors.

12. Attempts were made, with data from Group A only, to increase predictive coefficients by shortening psychomotor tests to a more nearly optimum length (by using the earlier portions of the obtained practice records), and by computing relative (percentage) gains. All attempts were unsuccessful, except possibly for re-scorings on the two-hand coordination test.

13. Multiple correlations were computed between selected combinations of measures and/or scores, on the one hand, and Criteria A and C on the other. These coefficients showed that (a) Criterion A was probably the more accurately predicted; (b) total scores, among those explored for the psychomotor tests, were the best predictors; (c) combinations of the most promising psychomotor test scores were about equal in predictive value to combinations of the best individual pencil-and-paper sub-tests; (d) to a large extent, one could probably interchange certain psychomotor and certain pencil-and-paper tests, in an effort to construct a test battery, and achieve about the same efficiency of prediction.

14. The successive sample technique was of value in showing that (a) among intra-test and inter-test correlations, those based on the least reliable measures fluctuated rather widely; (b) from one sample to the next there could be a consistent difference in mean test scores, accompanied by consistent differences in test reliability and fairly consistent differences in validity coefficients, probably as a function of motivation, among other variables; (c) although the over-all ranking of various tests as predictive

instruments may remain about the same, in successive groups specific types of scores on specific tests may interchange in their (adjacent) ranks, the same may be said of criterion measures; (d) certain tests and certain scores are quite consistent under successive sampling, even when the samples are relatively small, i.e., multiple correlations based on weights for one sample produce little shrinkage in r when applied to another sample.

APPENDIX A

SUPPLEMENTARY DATA, GROUP C, AND COMPARISONS WITH DATA OF GROUPS A AND B

SUPPLEMENTARY DATA, GROUP C, AND COMPARISONS WITH DATA OF GROUPS A AND B

INTRODUCTION

It will be recalled that a new logbook and flight rating procedure were introduced just at the time the first cadet class of Group C (N = 56) began its flight training. It was felt that it was not justifiable to use the data of Group C in analyses where they were exactly comparable to the earlier groups, but to omit them where they differed (any analysis involving the flight criterion). This appendix therefore consists of graphs and summary tables for Group C separately, presented in the same order as for Groups A and B, with a running commentary on how these final data compare with the earlier ones. (Several of the detailed analyses have not been done again because it was felt they could add little to what had been presented previously.) All score definitions and tests are exactly like those for the earlier data, except in instances where criteria are analyzed.

ANALYSIS OF THE GROUP C DATA AND COMPARISONS OF THESE FINDINGS WITH DATA FROM GROUPS A AND B

The trial-by-trial means for the rotary pursuit test are presented in Figure 25. The raw scores (in .001 min. units) have been translated into per cent time-on-target values. Comparison with Figure 5 of the main study shows that Group C is very slightly superior to Groups A and B combined, although their initial scores are an almost exact duplicate of those for the earlier combined groups.

The learning curve for the serial coordination test is shown in Figure 26, and is to be compared with Figure 6 of the main study. Group C has slightly higher initial status than Groups A and B combined, and ends its practice period slightly inferior to the earlier groups.

On the two-hand coordination test (compare Figure 27 with Figure 7 in the main study) Group C initially is about equal in skill to Groups A and B combined, but is slightly inferior from about trial 19 to the end of practice.

Figure 28 to be compared with Figures 8a and 8b of the main study, shows that on both the track time and net time measures for the division-of-attention test, Group C started at about the same level as Groups A and B combined, but throughout most of the practice thereafter was somewhat inferior to the combined groups.

In summary, Group C started each of the five tests or measures at almost the same level of skill as did the earlier combined groups. In the rotary pursuit test, Group C became somewhat superior, in the remainder

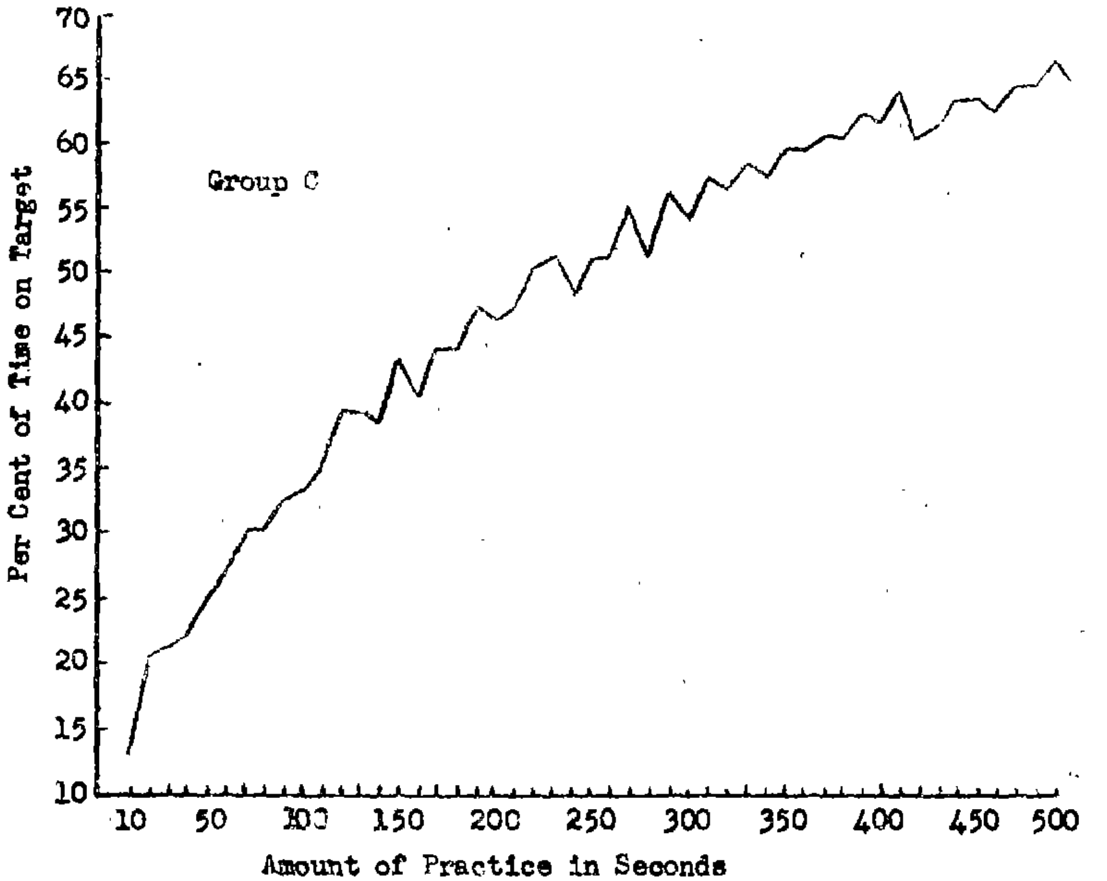

FIGURE 25

MEAN PER CENT OF TIME ON TARGET
ROTARY PURSUIT TEST

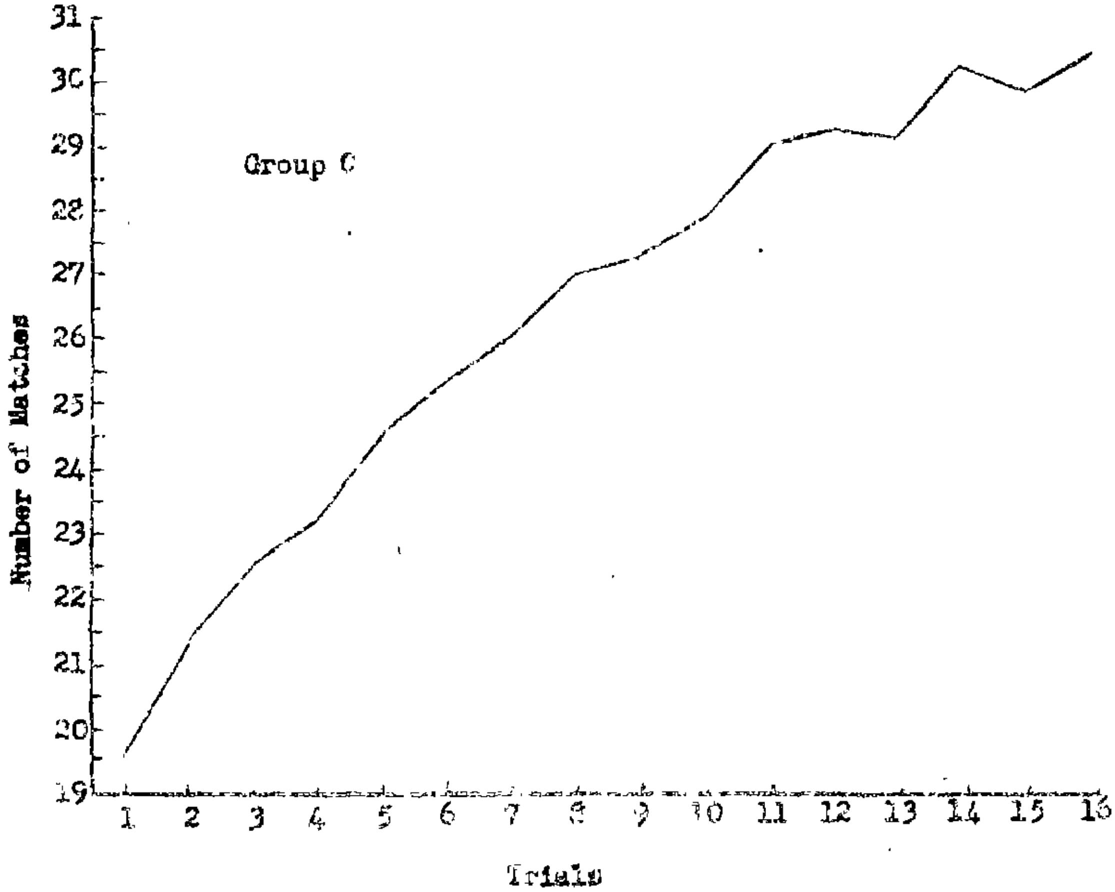

FIGURE 26

MEAN NUMBER OF MATCHES
SERIAL COORDINATION TEST

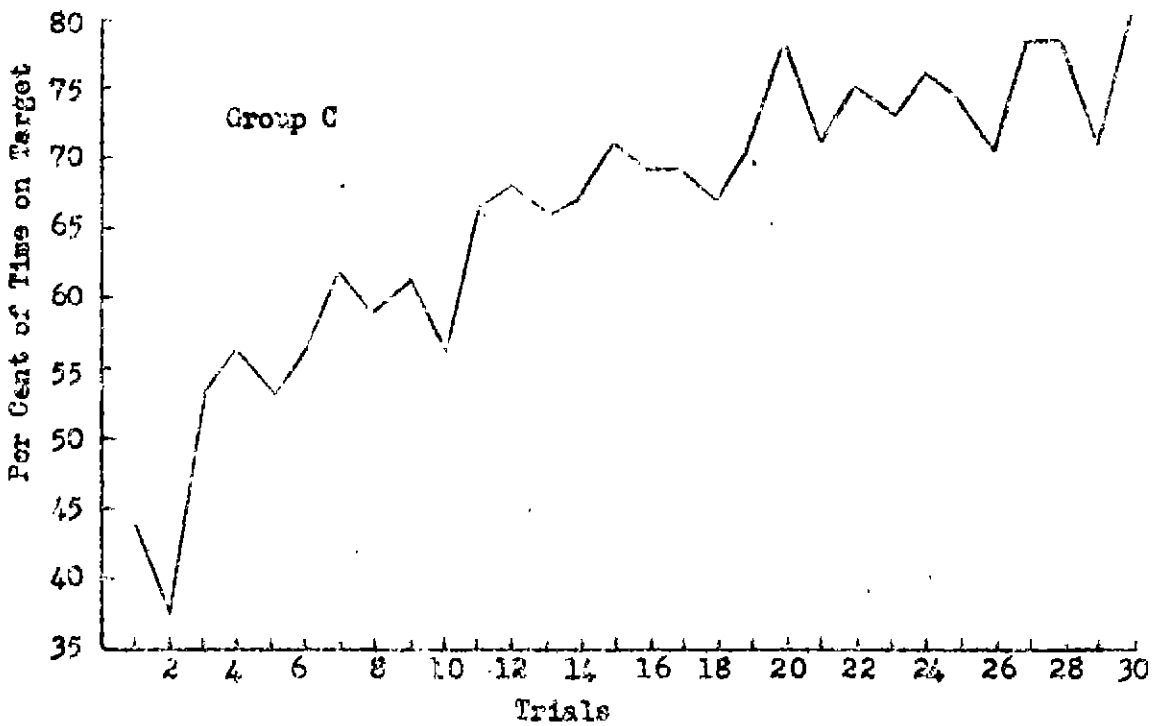

FIGURE 27

MEAN PER CENT OF TIME ON TARGET
TWO-HAND COORDINATION TEST

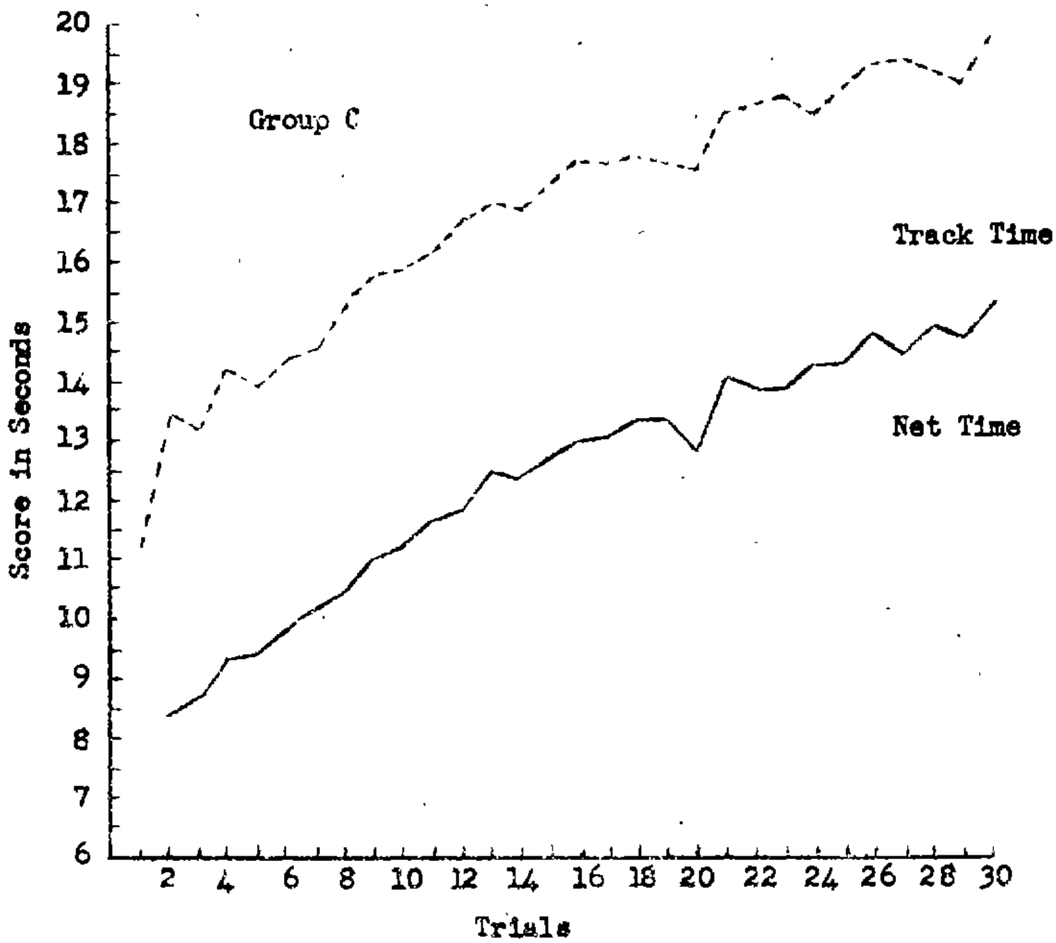

FIGURE 28

MEAN SCORES FOR TRACK TIME AND NET TIME
DIVISION-OF-ATTENTION TEST

of the measures somewhat inferior, to the combined groups. It is to be expected that such differences in learning, although not large, will have an effect on the several types of score applied to each test or measure.

Table A-1 is to be compared with Table E-2 of Appendix E. Since most of the differences between Group C and the combined earlier groups are so small that evaluations of them are not warranted, only a few of the larger differences may be pointed out.

On the rotary pursuit test, I-C gain is slightly greater in Group C than in Groups A and B combined; C-F and I-F gains are somewhat smaller. Standard deviations are slightly greater in Group C for all scores.

Since the serial coordination test curve for Group C starts slightly higher and levels off earlier than in the combined groups, the gain scores on this test are smaller for this group than for the combined groups. Standard deviations are nevertheless slightly greater throughout.

Examination of scores on the two-hand coordination test makes it seem probable that the gradually increasing inferiority of Group C, compared to the combined groups, was due to a motivational change, i.e., not only are mean scores relatively lower, but variabilities increase, as though erratic performance was occurring. It will be recalled that this was first noticed in Group A.

There is little to say about the two division-of-attention test measures except that Group C had lower scores throughout, with somewhat greater variabilities.

TABLE A-1

MEANS AND STANDARD DEVIATIONS FOR ALL TYPES OF SCORES ON ALL PSYCHOMOTOR TESTS AND MEASURES, GROUP C
(N = 56)

Test		Initial Score (I)	Central Score (C)	Final Score (F)	Total Score (T)	Gain I-C	Gain C-F	Gain I-F
Rotary	M	258.0	561.8	637.0	3937.0	303.4	75.9	379.3
Pursuit	S.D.	117.8	151.1	129.8	1017.0	95.1	70.0	89.8
Serial	M	40.8	57.6	60.1	421.1	16.8	2.2	19.3
Coordination	S.D.	6.7	8.6	9.3	58.8	6.3	4.4	7.4
Two-Hand	M	1091.0	1420.8	1543.0	9470.0	331.8	120.4	452.2
Coordination	S.D.	227.4	230.2	207.0	1427.0	155.5	143.2	175.0
Division-of-Attention:								
Net Time	M	3718.0	5241.0	5946.0	35000.0	1512.0	716.0	2228.0
	S.D.	924.0	1106.0	1259.0	6770.0	860.0	876.0	1205.0
Track Time	M	5596.0	7073.0	7714.0	47890.0	1470.0	648.0	2118.0
	S.D.	1077.0	1275.0	1442.0	8200.0	972.0	791.0	1135.0

As in the earlier groups central and final status scores are the most closely related (see Table A-2). Initial and final scores are in each test slightly more closely related than are initial and central scores -- a reversal of the tendency in Groups A and B. This reversal comes about primarily through an increase in the size of the initial vs. final coefficients. No ready explanation for this increase is apparent. Table A-1 shows that range of talent may be slightly greater in Group C, but the significance of this for the correlation pattern of Table A-2 is not clear.

TABLE A-2

INTERCORRELATIONS OF STATUS SCORES, GROUP C

Test	I vs. C	C vs. F	I vs. F
Rotary Pursuit	.74*	.89	.77
Serial Coordination	.62	.88	.69
Two-Hand Coordination	.66	.79	.76
Division-of-Attention:			
Net Time	.54	.82	.66
Track Time	.63	.83	.68

*In this and in all succeeding tables, the 5 per cent level of confidence for r is .26 (N = 56).

As in the earlier data, initial scores correlate at a high level with total scores, but nevertheless not quite as high as to central and final scores (see Table A-3). In each test, final scores show slightly higher correlations with total score than do central scores, a reversal of the slight trend in the earlier data. The reversal probably comes about through increases in size of the final total r's.

Early gain relates more closely to over-all gain (see Table A-4) than does late gain, as in Groups A and B. As before, the correlations between early and late gains are low and negative. There seems to be a great similarity in the pattern of r's as related to specific tests, when the earlier data are compared with these.

TABLE A-3

CORRELATIONS OF STATUS SCORES WITH TOTAL SCORE, GROUP C

Test	I vs. T	C vs. T	F vs. T
Rotary Pursuit	.87	.92	.96
Serial Coordination	.82	.92	.94
Two-Hand Coordination	.84	.85	.93
Division-of-Attention:			
Net Time	.77	.90	.94
Track Time	.82	.85	.90

TABLE A-4

INTERCORRELATIONS OF GAIN SCORES, GROUP C

Test	Gain I-C vs. C-F	Gain I-C vs. I-F	Gain C-F vs. I-F
Rotary Pursuit	-.44	.72	.31
Serial Coordination	-.08	.81	.53
Two-Hand Coordination	-.32	.63	.54
Division-of-Attention:			
Net Time	-.04	.69	.70
Track Time	-.18	.73	.54

As previously, early and over-all gains tend to be positively related to total score (see Table A-5). Late gains tend to be negatively correlated with total score, or else correlated positively at a low level. The patterns of r's as related to specific tests, are much the same in Group C as in Groups A and B.

TABLE A-5

CORRELATIONS OF GAIN SCORES WITH TOTAL SCORE, GROUP C

Test	Gain I-C vs. T	Gain C-F vs. T	Gain I-F vs. T
Rotary Pursuit	.42	-.37	.16
Serial Coordination	[illegible]	.11	.42
Two-Hand Coordination	.17	-.26	-.06
Division-of-Attention:			
Net Time	.38	.13	.36
Track Time	.27	.11	.31

In Table A-6 initial status score, as earlier, tends to correlate negatively with gain score. Central status score again is positively related to early or over-all gain. In this group it tends to be related positively to late gain, a reversal in comparison with results for Groups A and B. Final status score again correlates positively with early and over-all gains, but in contrast with the earlier data, correlates negatively with late gains. The reversals mentioned above are related, i.e., where the C vs. C-F r is nearest zero (rotary pursuit test, two-hand coordination test), the F vs. C-F r is most definitely negative. The reverse is true of the remaining three measures, for the C vs. C-F r is definitely positive and the F vs. C-F r is near zero or slightly negative. As in the main study, these relationships are understandable in view of the shapes of the learning curves involved, i.e., in Group C, there was relatively less improvement during the latter part of the practice period on the rotary pursuit and two-hand

TABLE A-6

INTERCORRELATIONS OF GAIN SCORES AND STATUS SCORES, GROUP C

Scores	Rotary Pursuit	Serial Coordination	Two-Hand Coordination	Division-of-Attention Net Time	Division-of-Attention Track Time
I vs. I-C	-.03	-.13	-.31	-.23	-.23
I vs. C-F	-.30	-.03	-.26	-.07	.07
I vs. I-F	-.27	-.13	-.49	-.21	-.15
C vs. I-C	.48	.55	.22	.48	.40
C vs. C-F	-.06	.40	.17	.45	.48
C vs. I-F	.45	.70	.33	.66	.67
F vs. I-C	.61	.63	.39	.58	.56
F vs. C-F	-.51	-.08	-.46	-.09	-.09
F vs. I-F	.25	.49	-.03	.35	.42

coordination tests than there was in the remaining measures, as is seen by comparison of mean learning curves and mean status or gain scores for this group and the combined groups.

In Table A-6, for example, the subjects having higher central scores on the rotary pursuit test had obtained them by having the larger I-C gains ($r = .48$); there was not much improvement in the group as a whole, after this central point, and the C-F gains were distributed in a somewhat inverse fashion among those having either high or low central status ($r = -.06$). The I-C gain is the main or overlapping component in I-F gains ($r = .72$, see Table A-4), so the latter are also positively related to central status ($r = .45$). Those of high status at the central segment maintained it in the final score ($r = .89$, see Table A-2), and I-C gain as a large component of I-F gain determines final status, so that in the present table final status and I-C gain correlate .61. Further, subjects who had high central scores and therefore high final scores are the ones who show least C-F gain; and it is possible for final status and C-F gains to correlate -.51. Lastly, final status and over-all gain correlate positively ($r = .25$). This lengthy analysis could be applied to each of the remaining measures in Table A-6 to offer some support for the argument that intra-test correlations are a function of degree of overlap between measures and of the shape of the learning curve on which the measures are based.

As in the previous data, two of the highest correlations are between the two hidden figures tests, and between planning circuits and mechanical comprehension (see Table A-7). In Group C, however, the r between path distance and mechanical comprehension is higher than for Groups A and B. In general, the visual pursuit test appears again to be the least closely related to the other sub-tests. The fluctuations from previous correla-

TABLE A-7

INTERCORRELATIONS OF PENCIL-AND-PAPER SUB-TESTS, GROUP C

	Planning Circuits	Hidden Figures I	Hidden Figures II	Path Distance	Visual Pursuit	Mechanical Comprehension
Planning Circuits		.17	.13	.30	.36	.53
Hidden Figures I			.57	.18	.28	.26
Hidden Figures II				.49	.14	.32
Path Distance					.12	.57
Visual Pursuit						.38
Mean	26.7	22.7	14.6	12.6	12.2	12.4
Standard Deviation	9.3	3.3	4.3	3.3	2.0	6.3

tions to the correlations reported in Table A-7 are about of the same magnitude as changes from Group A to Group B. Such fluctuations are not great enough to conceal the general trends mentioned above.

As in the earlier study, psychomotor measures intercorrelate at a relatively low level (see Table A-8). However, the correlation coefficients based on total score (i.e., the most reliable score) tend to be the highest for the various tests or measures correlated. Further, there is no evidence for a systematic increase in size of intercorrelations from early to later stages of practice.

Again, with the exception of the r between two scores on the division-of-attention test, no gain correlation is very large (see Table A-9). There is no ready explanation, except in terms of chance fluctuations, for the predominance of negative coefficients in the C-F gain column. It may be worth recalling that the rotary pursuit test and the two-hand coordination test learning curves showed little improvement for Group C beyond the central score. When such C-F gains are correlated with those on tests showing greater improvement, rather low intercorrelations are to be expected, but not systematically negative coefficients.

Comparison of columns in Table A-10 shows that each pencil-and-paper test correlates approximately at the same general level with the psychomotor total scores. Comparison of rows indicates that possibly the rotary pursuit test shows lower correlations than other tests with the pencil-and-paper battery, and the two-hand coordination test the highest coefficients. The correlation coefficients again are similar in size to those among psychomotor tests.

CRITERION ANALYSES

Two different ways of scoring the criterion data seemed feasible. The first of these is called Criterion A', because it is exactly analogous to Criterion A of the main study in all respects but one, i.e., the

TABLE A-8

TEST AND MEASURE INTERCORRELATIONS BASED OF STATUS SCORES, GROUP C

Test	Initial	Central	Final	Total
Rotary Pursuit vs. Serial Coordination	.39	.41	.38	.48
Rotary Pursuit vs. Two-Hand Coordination	.24	.19	.25	.28
Rotary Pursuit vs. Net Time	.33	.24	.31	.31
Rotary Pursuit vs. Track Time	.21	.23	.31	.25
Serial Coordination vs. Two-Hand Coordination	.41	.39	.24	.52
Serial Coordination vs. Net Time	.22	.35	.31	.29
Serial Coordination vs. Track Time	.25	.30	.31	.28
Two-Hand Coordination vs. Net Time	.33	.20	.28	.45
Two Hand Coordination vs. Track Time	.27	.18	.15	.43
Net Time vs. Track Time	.88	.89	.90	.90

TABLE A-9

TEST AND MEASURE INTERCORRELATIONS BASED ON GAIN SCORES, GROUP C

Test	I-C	C-F	I-F
Rotary Pursuit vs. Serial Coordination	.06	.00	-.03
Rotary Pursuit vs. Two-Hand Coordination	.21	-.06	.20
Rotary Pursuit vs. Net Time	.06	-.14	.18
Rotary Pursuit vs. Track Time	.15	-.02	.22
Serial Coordination vs. Two-Hand Coordination	-.03	-.04	.07
Serial Coordination vs. Net Time	.27	-.11	.15
Serial Coordination vs. Track Time	.21	-.06	.24
Two-Hand Coordination vs. Net Time	-.24	-.12	-.20
Two-Hand Coordination vs. Track Time	-.19	-.22	-.16
Net Time vs. Track Time	.24	.77	.86

TABLE A-10

CORRELATIONS BETWEEN PENCIL AND PAPER TESTS AND TOTAL SCORES OF PSYCHOMOTOR TESTS, GROUP C

Test	Planning Circuits	Hidden Figures I	Hidden Figures II	Path Distance	Visual Pursuit	Mechanical Comprehension
Rotary Pursuit	-.04	-.02	.14	.10	.24	.13
Serial Coordination	.08	.25	.15	.11	.28	.29
Two-Hand Coordination	.18	.25	.21	.31	.15	.30
Division-of-Attention:						
Net Time	.27	.14	.12	.31	.21	.25
Track Time	.27	.16	.12	.29	.19	.13

ratings made on flight maneuvers in the new logbook (see Figure 29) offered only a 4-point range instead of the original 5-point range.

It was apparent, during the Criterion A' analysis, that not many cadets were rated by more than one instructor. This came about, in part, because instructor manpower had become great enough that it was unnecessary to shift cadets every time a vacancy occurred. Therefore Criterion E was secured. It was simply an uncorrected mean maneuver rating for the "last-10" lessons given by the principal instructor of each cadet. The original criterion most resembling E is D, which was an uncorrected mean of over-all percentage grades on the "last-10" lessons. Since such percentage grades were not available for Group C, the 4-point ratings had to be utilized. (The mean rating for a given cadet was multiplied by 10 and subtracted from 40, a large number representing good flight performance.)

Criterion A' and Criterion E correlated .67. Investigations of their reliabilities did not appear worthwhile. One point concerning them which is worthy of note is that Criterion A" had a standard deviation of 3.0. The latter value shows a greater restriction of ratings (possible range of ratings = 3 x 10) on Criterion A' than on Criterion A (standard deviation = 3.9, possible range of ratings = 4 x 10). Criterion E showed a mean of 27.[illegible] (best possible score, 30; poorest score, 0) and a standard deviation of 4.[illegible].

The new criteria (see Table A-11) are evidently predicted with less efficiency than the criteria examined in the main study. In the case of Criterion A', the restriction of range probably has something to do with this finding -- as may also the (presumptively) lower reliability of score correction procedures based on smaller samples per instructor.

Among the several types of scores, initial status when correlated against Criterion A' appears to be relatively better as a predictor than final status; this is the reverse of previous findings. Gain scores have the lowest predictive coefficients. In terms of total score on Criterion A', the serial coordination test and two-hand coordination test are the best predictors in Group C, with track time close behind. In contrast

Form [illegible]
(Revised January 194[illegible])

Notes should preferably be made in the air. Mark only those maneuvers performed. Marks are based on <u>what is expected at this stage of training</u>. The chief purpose of this slip is to note specific errors in order to help the student improve. Errors must be noted for all Borderline and Unsatisfactory marks.

Mark Only Maneuvers Performed	Unsat.	Border.	Sat.	Good	INDICATE ERRORS for type of error: for too fast, too much, too high over, late / for too slow, too little, too low, under, early / for variable or erratic
Taxiing					control ☐ not alrt ☐ brakes ☐ rdr ☐ tl lk ☐
Take-off					thrtle ☐ hdng ☐ nose ☐ pull off ☐
S turns					spd ☐ planning ☐ alt ☐
Orientation problem					planning ☐ result ☐
Stalls					entry ☐ reaction ☐ recovery ☐
Normal spins					fear ☐ Entry improper ☐ During aileron ☐ recovery ☐
Slips					spd ☐ hdng ☐ entry ☐ recovery ☐
Spirals					spd ☐ thrtle ☐ slips ☐ skids ☐
Landings					patrn ☐ spd ☐ hits / landing sense ☐
Slips to					patrn ☐ spd ☐ slps ☐ hits / lndg snse ☐
Dragging areas					spd ☐ thrtle ☐ alt ☐ planning ☐
Power wheel landings					spd ☐ pin ☐ accuracy ☐ thrtle ☐
Landings					apprch ☐ spd ☐ level off ☐ drift ☐
ALWAYS MARK:					
Basic airwk					climbs ☐ glides ☐ turns ☐ tab ☐ spd ☐ alt ☐ thrtle ☐ slp ☐ skd ☐ rough ☐
Headwork and air discipline					Doesn't know ☐ care ☐ think ☐ error due to look around ☐ remember ☐ lack of skill ☐
Emergencies					nose down ☐ field choice ☐ planning ☐ spd ☐
Total UBSG marks					REMARKS (use back if needed):
Cumulative stage total					

Instructor's Signature____________________

Student's signature____________________

Date________________ Session________

Flight No.__________ Plane No.________

Ground instruction________________

	Dual	Solo
Time this flight		
Previous stage time		
Total stage time		
Total course time		

STAGE B AND INTERMEDIATE (DUAL)

FIGURE 29

TABLE A-11

CORRELATIONS OF VARIOUS PSYCHOMOTOR TESTS AND MEASURES WITH THE CRITERIA, GROUP C

With Criterion A'

Test	Initial	Central	Final	Gain I-C	Gain C-F	Gain I-F	Total
Rotary Pursuit	.17	.23	.19	.15	-.14	.05	.21
Serial Coordination	.47	.35	.26	-.02	-.13	-.09	.36
Two-Hand Coordination	.28	.42	.24	.22	-.34	-.08	.41
Division-of-Attention:							
Net Time	.21	.09	.10	-.10	-.05	-.11	.18
Track Time	.29	.16	.18	-.12	.07	-.05	.31

With Criterion E

Test	Initial	Central	Final	Gain I-C	Gain C-F	Gain I-F	Total
Rotary Pursuit	.02	.14	.14	.19	-.03	.18	.09
Serial Coordination	.41	.28	.15	-.05	-.24	-.19	.25
Two-Hand Coordination	.07	.17	.02	.16	-.26	-.06	.15
Division-of-Attention:							
Net Time	.09	.10	.09	.04	-.03	.01	.13
Track Time	.16	.15	.14	.02	.01	.03	.19

with data from Group B, no gain score on the serial coordination test seems to hold much predictive promise.

The findings of Table A-12 contrast more with previous data than is the case for any other comparison. The mechanical comprehension test is a fairly good predictor, and hidden figures II a poor predictor. The cor-

TABLE A-12

CORRELATIONS BETWEEN PENCIL-AND-PAPER TESTS AND CRITERIA, GROUP C

Test	Criterion A'	Criterion E
Planning Circuits	.35	.38
Hidden Figures I	.07	.03
Hidden Figures II	.08	.10
Path Distance	.15	.19
Visual Pursuit	.29	.26
Mechanical Comprehension	.22	.28
Mechanical Comprehension + Path Distance	.22	.28
Mechanical Comprehension + Hidden Figures I	.21	.24
Total Score	.32	.35

relations for the other four tests are now high where they were low previously and vice versa. It is probable that only the total pencil-and-paper score is reliable enough to stand up well under repeated sampling. As was not the case with the psychomotor tests, Criterion E is predicted as well as or better than Criterion A'. The fact that pencil-and-paper measures show even more fluctuation in predictive efficiency than do psychomotor measures is perhaps a suggestion for the greater utility of the latter.

APPENDIX B

PROCEDURES FOR ADMINISTRATION OF PSYCHOMOTOR TESTS

1. General Procedures (Experimenter's Instructions)
2. Rotary Pursuit Test Instructions
3. Serial Coordination Test Instructions
4. Two-Hand Coordination Test Instructions
5. Division-of-Attention Test Instructions

I. GENERAL PROCEDURES (EXPERIMENTER'S INSTRUCTIONS)

A. Each man comes twice per week for two weeks, spending an hour at each of the four test sessions. In addition, he takes two hours of pencil-and-paper tests.

B. At the first session, the two men per hour are given the steadiness test and the rotary pursuit test. The steadiness test is a buffer -- the men should assume it counts, but it actually will not. About 10-15 minutes should be consumed in giving the instructions for this test <u>and</u> the test (10 30-sec. trials, merely holding the stylus in the smallest hole -- see instruction sheet). Then to the rotary pursuit test, for 50 trials on Condition D (10 sec. work, 20 sec. rest); clear the clocks after each trial. Be sure to fill out the first two lines on the rotary pursuit data sheet -- the third is not necessary, because we can get it from logbooks. Assign the <u>S</u>s at random to order I or order II, so that any subsequent <u>E</u> will know what tests to give a man at any appointment. An order-list will be kept in the waiting room.

C. If a man has test-order I, his next test is the division-of-attention test. After instructions are given, the test is self-administering. Thirty 30-sec. trials with 30-sec. rest. Clear clocks between trials.

D. If a man has test-order II, his second test is the Mashburn (called "serial coordination"). Use a 2-min. pre-trial (trial 0 on the data sheet), then <u>16</u> 2-min. trials with 30-sec. rests intervening. Be especially careful to learn all operational tricks with this test.

E. The third test for all men is the lathe; 30 30-sec. trials, with 30-sec. rest between trials. This one is tricky, too.

F. The fourth test is the remaining one (division-of-attention or serial coordination).

G. Notes:

1. On all tests after the first only the man's name needs to be entered on the data sheets.

2. No man is to be told any score; if it is requested, make a non-commital answer, indicating that you are not permitted to give the information. Make no comparisons of men, and use only such encouragement as is necessary to keep each man trying to do his best.

3. Every test has a Variac; keep the voltage at 115 or slightly better.

2. ROTARY PURSUIT TEST INSTRUCTIONS

To E: CLEAN STYLUS AND TARGET; SHADE DOWN; OVERHEAD LIGHT ON; ASK FOR PREVIOUS EXPERIENCE WITH THIS TEST.

To S:

A. "This is a test of your ability to follow a moving target. This black disc (point) will rotate in a clockwise direction." (Examiner traces motion with finger.) "Your task is to keep the point of this stylus (point) in contact with the round brass target (point) while the disc is turning. Your score is the amount of time the stylus is on the target."

B. Start unit motors, Condition D. Unit 1 off; unit 2 on.

C. "The stylus is to be held lightly in your preferred hand in this manner." (Examiner demonstrates correct manner of holding stylus.) "Hold the back of the hand up with all four fingers and the thumb grasping the handle. Hold the cord in your other hand." (Examiner follows moving target with stylus.) "Follow the target like this, trying to stay in contact with it for as great a time as possible. You will do best if you develop a smooth, free-swinging motion of your arm and shoulder like this." (Demonstrate.)

D. Reset clocks; check voltage.

E. "You will notice that the stylus is made so that you cannot easily put pressure on its point. Do not attempt to put pressure on the point by turning the handle up like this (demonstrate), by tilting the stylus like this (demonstrate), or by putting your finger on the metal rod like this (demonstrate). Now pick up your styluses and place them on the target." (Examine each man's grip on his stylus and correct it by repeating appropriate parts of the instructions.)

F. "You will be given a long series of tests and rests. When the disc starts, put the stylus on the target and try to keep it there. Your score does not begin to count until the buzzer sounds. When the disc stops, lift the stylus off the target and keep it off until the disc starts again. Are there any questions? There is to be no talking during the test."

G. Units 1 and 2 on.

H. "Remember, your score is the amount of time that the stylus is actually in contact with the target. It will be to your advantage to do the best you can. Ready . . ."

[illegible] COORDINATION TEST INSTRUCTIONS.

[illegible] OVERHEAD LIGHT ON; SHADE DOWN; JEWELS ON; (NO DISTRACTION LIGHTS); VOLTMETER MUST READ AT LEAST 115 VOLTS.

[illegible] S:

A During demonstration by examiner "This is a coordination test. When you sit in the chair (examiner takes seat in apparatus), push your right foot forward on the rudder bar as far as possible (demonstrate) to see if the chair needs adjusting for you. Keep your feet well up on the rudder bar so that your heels do not touch the floor."

B "Your task is to match each of these three red lights (point) with a green one. The red lights are stationary, but you can move the green lights by operating the stick and the rudder bar. If the green light goes off altogether, it means that you are between electrical contact points. Keep moving until you find out where you are. When you have matched all three of the lights, you must hold it for a half-second; and then a new setting of red lights will appear." (Demonstrate slowly with a setting No. 1.)

C. After completing setting No. 1, say: "The new setting is then to be matched as quickly as possible in the same way. You will hear a buzzing sound from my control box at intervals, but pay no attention to it."

D. While the candidate is seated, and after adjustment of seat: "Use only your preferred hand on the stick throughout the test. Do not match the lights now, but move the controls as I direct you. Move the green light in the top row by moving the stick to the right and left. Move the middle green light by moving the stick toward and away from you. Move the green light in the bottom row by moving the rudder bar. Now let go of the stick. You will do best if you match the top and middle lights with the stick first and then the bottom light."

E "Now we are ready for the test. There will be a practice period of 2 min. which will not count, followed by a rather long series of 2 min. trials with a brief rest period after each one. During this rest period, you are to simply lean back in your chair and relax, and take your feet off the rudder bar. After each rest, you are to wait until I say, "ready begin" before beginning to operate the controls again. Match as many patterns of lights as you can. Your score will be the number of patterns you match. Are there any questions?"

You are not to talk during the test It will be to your advantage to do the best you can." (Read counter.) "Ready (start clock), go."

EXPERIMENTER IS TO REMIND SUBJECT TO REMOVE FEET FROM THE RUDDER BAR BETWEEN TRIALS.

THE LIGHTS SHOULD REMAIN ON _AT ALL TIMES_ (DURING REST INTERVALS).

F. Four seconds before the beginning of any trial, subject should receive the following warning signal: "Place feet on the rudder bar." MAKE SURE THAT THE SUBJECT DOES NOT TRY TO MATCH LIGHTS DURING "READY" INTERVAL.

4 TWO-HAND COORDINATION TEST INSTRUCTIONS

To E: SHADE DOWN; OVERHEAD LIGHT ON; ASK ABOUT EXPERIENCE WITH LATHES AND RECORD EXPERIENCE UNDER REMARKS.

To S:

A. (E stands before lathe): "This is a two-hand coordination test. The black disc will rotate very slowly in a clockwise direction. This small brass target will move with the disc, and will also move in an irregular manner within this curved slot (point). Your task will be to keep the small button anywhere on top of this brass target. Your score will be the total amount of time you stay on the target. When you get off, get back on as quickly as possible."

"You are to move the button by turning the two handles at the same time. The upper handle moves the button toward you and away from you. The lower handle moves it from side to side." (Demonstrate.)

B. (Experimenter takes seat at control box and gives following instructions): "Now take your places at the apparatus. But don't touch it until I tell you what to do." (Frequently necessary to repeat this instruction.)

C. "Now, using your right hand only on the upper handle, move the button about an inch off the target and back on again. Let go of the upper handle. Now, using your left hand only on the lower handle again move the button about an inch off the target and back on again. Now let go of the handles."

"You will be given a long series of short work periods and rest periods. At the end of each work period, if you are off the target, get back on as quickly as possible and then take your hands off the handles and wait for the ready signal."

D. "When I say ready, grasp the handles but do not turn them. When I say go, the target will begin to move. You try to stay on it. Never release the handles in order to spin them (explain this statement upon request). Are there any questions? There is to be no talking during the test. It will be to your advantage to do the very best you can."

E. (At the end of first trial and later as the situation requires): "Get back on the target. Then take your hands off the handles and wait for the ready signal."

5. DIVISION-OF-ATTENTION TEST INSTRUCTIONS

To E: SHADE DOWN; OVERHEAD LIGHT OFF; BEFORE ANY S IS RUN, SET THE VOLTMETER AT APPROXIMATELY 115; IF NECESSARY WARN S DURING EARLY TRIALS TO LOOK DIRECTLY AT THE POINTER AND NOT AT THE LIGHTS; CHECK THE TIMER CONTACT-STRIPS AND CONTACTS TO BE SURE GOOD CONTACT IS MADE.

To S:

A. "This is a test of your ability to do two things at once. The first thing is to keep the tip of this pointer on the ridge which will move back and forth in the slot (point). The second thing is to keep all these lights (point) turned off by means of these switches (point). You will have to learn which switch turns out any light, but it will always be the same switch for that light. Which hand do you want to use for the switches? Then only these (point) 4 switches will operate the lights. You will make the best score if you learn quickly to keep the pointer on the ridge."

B. "Stand facing the apparatus squarely. You must always look directly at the pointer. This means that you will know which light is on only by seeing it in indirect vision, i.e., out of the corner of your eye. This will be difficult to do at first, but it is very important, and you must show how well you can follow this instruction."

C. "If any light is on, or is left on, you will not get credit for keeping the pointer on the ridge. In operating the lights, never use more than one switch at a time, but hold each correct switch until the next light comes on."

D. "You will have a series of trials and rest periods. During the rest period, put your hands down and relax. The apparatus will start 3 sec. before the counters are connected, and this will be your warning to get into action again. Trials and rests are each 30 sec. long."

E. "I shall now give you a demonstration of what you are supposed to do."

F. "Any questions? There is to be no talking after the test starts. Ready for the first trial. It will be to your advantage to do the very best you can."

(THE CLOCK AT THE LEFT IS CLOCK NO. 1 ON THE DATA SHEET!)

APPENDIX C

SAMPLE PAGE FROM LOGBOOK

Course ____________________

Stage ____________________

Lessons ________ to ________

Type of aircraft	NC number	1 Engine H. P.	2 Preflight check	3 Taxiing, use of brakes	4 Take-offs	5 Traffic pattern	6 Straight and level flight	7 Rectangular course	8 "S" turns across road	9 Series of eights	10 Pylon eights	11 Climbs and climbing turns	12 Series of turns	13 Steeply banked turns	14 Lazy eights or wingovers	15 Chandelles	16 Stalls	17 Spins (precision)	18 Slips	19 Spirals	20 Glides and gliding turns	21 Loops	22 Immelmann turn	23 Slow roll	24 Snap roll	25 Vertical reverses (cartwheel)	26 Half-roll (split "S")	27 Slow flight-approach to stalls	28 Trimming airplane	29 Throttle elevator coordination	30 Rate turns (gyro., clock)	31 Dead reckoning	32 Beam bracketing	33 Cone interception	34 Radio tuning	35 Fadeout method	36 90° method	37 Parallel method	38 Combination method

	DUAL		SOLO	
	Hours	Minutes	Hours	Minutes
Time this page				
Previous time				
Total stage time				
Total course time				

I hereby certify that the above entries are correct and I have received the above flight instruction.

I hereby certify that the above entries are correct as stated.

FIGURE 30

SAMPLE PAGE FROM LOGBOOK

Position finding	Loop bearings	Homing loop	Estimated arrival time	Approach let down	Final approach	Emergency pull-ups	Emergency landings	Small field procedure	90o ("S" turns to circle)	180o (slips to circle)	Power landings	Single-engine procedure	Normal landings	Flight logs	Flight plans	Use of computer	Use of weather data	Instructional technique	Aptitude	Judgment	Over-all grade	REMARKS	DUAL		SOLO		CERTIFICATE	
																							Hrs.	Min.	Hrs.	Min.	Std.	
39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60							
																						TOTAL THIS PAGE						
																						BROUGHT FORWARD						
																						TOTAL HOURS TO DATE						

(Signature of trainee)

(Signature of Instructor)

(CAA certificate No.)

(Rating)

FIGURE 30
SAMPLE PAGE FROM LOGBOOK (Cont'd.)

APPENDIX D

THE RELATIONSHIP BETWEEN FLIGHT CRITERIA AND RATES OF FITTED LEARNING CURVES, PART I

THE RELATIONSHIP BETWEEN FLIGHT CRITERIA AND RATES OF FITTED LEARNING CURVES, PART II

APPENDIX D

THE RELATIONSHIP BETWEEN FLIGHT CRITERIA AND RATES OF FITTED LEARNING CURVES[47]

PART I

INTRODUCTION

In the study entitled "An Appraisal of Certain Tests of Pilot Aptitude" data were analyzed to determine if there was any relationship between various measures of learning on a series of psychomotor tests and criteria of flight aptitude. One of the measures considered in this study was the gain score. The gain score is the increment of performance between certain arbitrarily chosen trials on the psychomotor tests. The gain score was considered to be a measure of the rate of learning, and its relationship to the criterion measures was investigated.

As a result of certain objections[48] to the use of the gain scores as they were measured, it was deemed advisable to develop some other measure of rate of learning. The other measure suggested was the slope of a curve fitted to the individual's performance scores on a particular test. This measure is obtained by taking the first derivative of the equation of the fitted curve. Another measure considered to be of value was the trial at which the fitted learning curve reached a maximum.

As a preliminary to applying these techniques to the entire test battery, this technique was applied to the data on the serial coordination test. To the data for each subject was fitted a curve of the form:

$$y = ax^3 + bx^2 + cx + d$$

The first derivative for each of these curves was taken and its value at the 2nd, 4th, and 6th trials was determined. The latter measures were the rate of learning measures that were correlated with the criterion measures. In addition, to each set of data for a subject, a curve was fitted of the form:

---

[47]Parts I and II of Appendix D were prepared by Dr. Seymour Wapner and Mr. David Bakan, Statistical Unit, University of Rochester. Acknowledgment is made to Dr. Raymond Franzen, Consultant to the Committee on Selection and Training of Aircraft Pilots, for fitting the slopes of the curves to the learning data.

[48]As pointed out in Footnote 17 the gain score was criticized on the ground that its correlation with the criterion is not statistically independent of the correlations of the initial and terminal points with the criteria, where these initial and terminal points are used in calculating the gain score.

$$y = ax^2 + bx + c$$

The first derivative of this equation was set equal to zero and the solution for x carried out. This gave the trial at which the curve reached a maximum. Maximum trial values were found only for those cases in which the fitted curve gave a maximum. There were a few cases in which the procedure outlined yielded only minima.

From the calculations four measures of learning were obtained, each of which could be correlated with the criterion measures. It is the purpose of this report to evaluate these measures as predictors of flight criteria and to note their relationship with the gain scores which were previously used.

METHOD OF ANALYSIS

The population studied consisted of a total of 153 subjects (Group A, 90 subjects; Group B, 63 subjects). The criteria used were:

1. Criterion A: mean maneuver rating, with the mean and variability of all instructors equalized.

2. Criterion B: mean percentage grade on "last-10" lessons, adjusted for leniency and variability of instructors.

3. Criterion C: mean maneuver rating on "last-10" lessons, adjusted for leniency and variability of instructors.

4. Criterion D: mean percentage grade on "last-10" lessons, uncorrected for leniency or variability of the instructors.

Correlation coefficients were computed between each of the rate measures and maximum trial and the above criteria.

Biserial correlation coefficients were computed between the criteria and the subjects divided into two groups characterized as having either a maximum or a minimum in the fitted curve.

Correlation coefficients were computed between the rate measures from the fitted curves and the gain scores, to determine the relationship between these two types of measures.

RESULTS

In Table D-1 the correlation coefficients of the four criteria with rate measures and maximum trials are indicated. Inspection of these values shows there is no reason to suppose that there is any relationship between these measures and the criteria. The correlation coefficients of the gain scores with the criteria are presented for comparison. It is evident, however, that these new measures are no more highly related to the criterion measures than the gain scores.

TABLE D-1

CORRELATIONS OF CRITERIA WITH RATE MEASURES, MAXIMUM TRIALS* AND GAIN SCORES
(Serial Coordination Test)

Criterion	Rate on 2nd Trial A	B	A+B	Rate on 4th Trial A	B	A+B	Rate on 6th Trial A	B	A+B	Maximum Trial* A	B	A+B
A	.07	-.11	-.01	.07	-.11	.00	.05	-.05	.01	.13	-.18	.04
B	.12	.01	.08	.14	.03	.10	.12	-.01	.06	.01	-.23	.04
C	.12	-.02	.06	.13	.01	.08	.10	.03	.07	.14	-.14	.07
D	.14	.01	.09	.16	.00	.10	.13	-.04	.06	.05	-.17	-.01
N	90	63	153	90	63	153	90	63	153	83	49	132

Criterion	Gain I-C A	B	A+B	Gain C-F A	B	A+B	Gain I-F A	B	A+B
A	.12	-.15	.01	.00	-.07	-.03	.10	-.21	-.02
B	.22	-.09	.09	.06	-.07	.00	.22	-.14	.08
C	.21	-.04	.11	.10	-.07	.03	.23	-.10	.11
D	.22	-.10	.10	.07	-.12	-.02	.23	-.19	.09
N	90	63	153	90	63	153	90	63	153

*Trial at which fitted curve reaches a maximum (in cases where a maximum is reached).

In Table D-2 are presented the biserial correlations of the criteria with the cases that have either a maximum or a minimum in the fitted curve. It may be observed that there is no relationship between having a maximum or a minimum in the fitted curve and the criterion measures.

Table D-3 presents the correlation coefficients of the gain scores with the new rate measures. It can be seen that the rate measures correlate highest with I-C gain; next highest with I-F gain; and to the smallest degree with C-F gain.[49]

The relatively high coefficients of correlation of the rate score with I-C gain and the I-F gain may be noted by inspection of the following dia-

---

[49] Initial score = sum of scores on trials 1 and 2.
Central score = sum of scores on trials 11 and 12.
Final score = sum of scores on trials 15 and 16.
I-C gain = central score minus initial score.
C-F gain = final score minus central score.
I-F gain = final score minus initial score.

TABLE D-2

BISERIAL CORRELATIONS OF CRITERIA WITH MAXIMUM AND MINIMUM CASES*
(Serial Coordination Test)

Criterion	Group A	Group B	Group A+B
A	.04**	-.13	-.07
B	.05	-.02	.02
C	.01	-.06	-.03
D	-.01	.15	.07
$N_{Max.}$	83	49	132
$N_{Min.}$	7	14	21

*Cases characterized as having either a maximum or minimum in the fitted curves.

**If the mean of maximum cases was higher than the mean of minimum cases, the coefficient was considered positive.

TABLE D-3

CORRELATIONS OF GAIN MEASURES WITH RATE MEASURES
(Serial Coordination Test)

	Gain I-C			Gain C-F			Gain I-F		
Criterion	A	B	A+B	A	B	A+B	A	B	A+B
Rate on 2nd trial	.62	.39	.50	.24	-.05	.11	.66	.34	.52
Rate on 4th trial	.83	.71	.77	.05	-.23	-.07	.72	.51	.64
Rate on 6th trial	.79	.79	.79	-.27	-.33	-.29	.50	.51	.50
N	90	63	153	90	63	153	90	63	153

gram. Hypothetical curves for two subjects are shown in Figure 31. For the purpose of exposition let us assume that the means of trials 1 and 2 for both subjects is the same, and that the means of the initial (1 and 2), central (11 and 12), and final scores (15 and 16) coincide exactly with the fitted curves. At trial 6 the tangents of the curves have been drawn. Curve A has an I-C gain greater than the I-C gain of Curve B by a; and its I-F gain is greater than the I-F gain of Curve B by a'. It is evident from this diagram that as a or a' are added to the gain scores of Curve B, all other factors remaining the same, the slope of the tangent of Curve B at trial 6 will tend to increase. This simultaneous increase of the slope of the tangent and gain score is reflected in the coefficients of correlation between gain scores and rate scores.

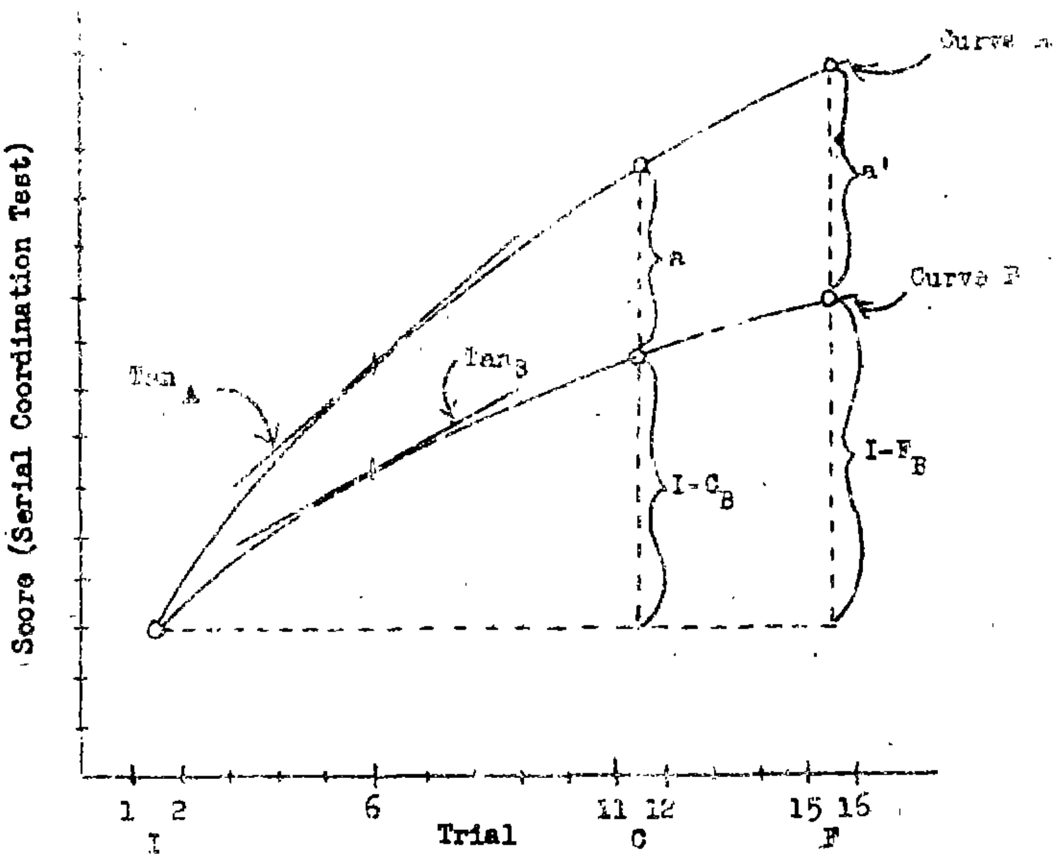

FIGURE 31

HYPOTHETICAL LEARNING CURVES

DISCUSSION

From the results presented it is fairly certain that the rate scores and maximum trial scores presented in this analysis are not very useful in predicting the criterion measures. However, there are several factors which may be operative in determining the low values of the coefficients of correlation.

1. There is the possibility that there is actually no relation between these criteria and rate of learning as measured by the serial coordination test.

2. The statistical procedures used in this analysis may be such as not to indicate the relationship between rate of learning on the serial coordination test and the criteria because (a) the curve chosen to be fitted to the data, that is, the curve of the form $y = ax^3 + bx^2 + cx + d$, is not sufficiently similar to the form of the data to yield valid measures of rate of learning,[50] and (b) the points chosen at which to determine the rates of learning are not the ones that would give the highest relationship to the criteria.

---

[50]Evidence suggested that a good fit was made with the curve used. After fitting the learning curves, each curve was drawn. The goodness of fit was then tested by inspection. In instances where the goodness of fit was suspect, tests of goodness of fit were applied (see: Yule, G. U., and Kendall, M. G. An introduction to the theory of statistics. (12th ed.) London: Charles Griffin & Co., 1940, pp. 328-329). In no instance was there sufficient reason to reject the curve.

APPENDIX D

THE RELATIONSHIP BETWEEN FLIGHT CRITERIA AND RATES OF FITTED LEARNING CURVES

PART II

INTRODUCTION

In Part I, Appendix D, an analysis was made of the relationship between the slopes at particular points of fitted curves to learning data and certain flight criteria. The data used in this analysis were the performance records of cadets on the serial coordination test. The analysis was initiated because of certain objections raised against the use of gain scores as measures of rate of learning where these rate measures were to be correlated with another measure. It will be recalled that it was suggested a better measure of rate of learning would be the slope of a curve fitted to the learning data.

In Part I the first derivatives were taken and evaluated at trials 2, 4, and 6. The rates so determined were then correlated with the flight criteria. However, on the basis of the analysis it was found that the correlations between the slopes at the points taken and the flight criteria were very low. It was suggested that several factors may have been operative, resulting in the low correlation coefficients, among which was one factor capable of further investigation (see 2, b, of Discussion, Part I). Therefore, this analysis was undertaken to explore the possibility that the relationship might be higher between rate of learning and the criteria at other points on the fitted learning curve.

METHOD OF ANALYSIS

In order to facilitate the determination of the correlation coefficients between the rates of the curves at each of the learning trials and the flight criteria, a general formula was developed.[51] It was possible, by means of this formula, to omit the step of taking each of the derivatives and evaluating at each point. Instead a set of constants was developed for the whole set of data, and by proper substitution in the formula, the desired correlation coefficients were obtained. A check on the validity of the formula was provided by checking the "formula-derived" coefficients with the coefficients derived by the long method presented in the preliminary report.

The population studied consisted of two groups of men tested at the University of Iowa and who later received flight training. The first group

---

[51]The general term for the coefficient of correlation between the rate of a fitted curve and another measure was obtained by the following formula:

consisted of 90 subjects (Group A) and the second group consisted of 63 subjects (Group B). In the following analysis, the groups are treated separately and together.

RESULTS

Table D-4 presents the correlation coefficients between the rates of the fitted curves and the flight criteria for each trial.

Examination of the table reveals that the highest correlation coefficient obtained, disregarding sign, is .19. This value occurs at the 12th and 13th trials for Group B, Criterion D. This value is not significant.

There is no marked tendency for the correlation coefficients to be either predominantly positive or negative. There seems to be some tend-

---

[51] (Continued)

$$r = \frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\ \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}} \quad (1)$$

$$\text{Let } X = \frac{dy}{dx} = 3ax^2 + 2bx + c \quad (2)$$

which is the rate of the curve at any point x.

Let Y be the measure.

Substituting (2) in (1) we get

$$r = \frac{Ax^2 + Bx + C}{P\sqrt{Dx^4 + Ex^3 + Fx^2 + Gx + H}} \quad (3)$$

Where

$A = 3 \quad (N\Sigma aY - \Sigma a\Sigma Y)$

$B = 2 \quad (N\Sigma bY - \Sigma b\Sigma Y)$

$C = \quad N\Sigma cY - \Sigma c\Sigma Y$

$D = 9 \quad \left[N\Sigma a^2 - (\Sigma a)^2\right]$

$E = 12 \quad (N\Sigma ab - \Sigma a\Sigma b)$

$F = 6 \quad (N\Sigma ac - \Sigma a\Sigma c) + 4\left[N\Sigma b^2 - (\Sigma b)^2\right]$

$G = 4 \quad (N\Sigma bc - \Sigma b\Sigma c)$

$H = \quad N\Sigma c^2 - (\Sigma c)^2$

$P = \quad \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}$

(Continued on p. 112)

TABLE D-4

CORRELATIONS OF CRITERIA WITH RATE MEASURES (SLOPE OF FITTED CURVE) AT EACH OF THE SUCCESSIVE SIXTEEN LEARNING TRIALS

		Trials															
Criterion	Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	A	.06	.06	.07	.07	.06	.04	.03	.01	.01	.01	.01	.03	.03	.04	.04	.04
A	B	-.10	-.10	-.10	-.10	-.09	-.05	-.02	-.01	-.02	-.04	-.08	-.14	-.18	-.18	-.16	-.15
	A+B	-.01	-.01	-.01	-.00	.00	.01	.01	.01	.00	-.01	-.02	-.04	-.05	-.05	-.05	-.04
	A	.11	.12	.13	.14	.14	.12	.09	.06	.06	.06	.08	.11	.10	.10	.09	.08
B	B	.02	.02	.02	.02	.02	-.00	-.02	-.04	-.06	-.09	-.13	-.18	-.18	-.15	-.12	-.10
	A+B	.07	.08	.09	.09	.09	.06	.03	.01	-.01	-.02	-.03	-.03	-.02	-.01	.00	.01
	A	.11	.12	.12	.13	.12	.10	.07	.05	.05	.07	.11	.15	.16	.14	.13	.12
C	B	-.02	-.02	-.01	.01	.03	.04	.03	.02	.01	-.01	-.04	-.08	-.11	-.11	-.10	-.09
	A+B	.06	.06	.07	.08	.09	.07	.05	.04	.03	.03	.04	.05	.05	.05	.04	.04
	A	.13	.14	.15	.16	.16	.13	.09	.07	.06	.07	.11	.14	.15	.14	.13	.12
D	B	.01	.01	.01	.00	-.02	-.04	-.05	-.06	-.08	-.11	-.15	-.19	-.19	-.15	-.12	-.09
	A+B	.08	.09	.10	.10	.10	.07	.04	.01	.00	.00	.00	.01	.02	.03	.04	.04

ency for the coefficients of Group B to be negative, and the coefficients of Group A to be positive. However, in view of the low magnitude of the coefficients, little significance should be assigned to this fact.

The largest coefficients for Group B tend to lie somewhere in the neighborhood of the 13th trial; and the largest coefficients for Group A in the neighborhood of the 4th trial. Again, due to the low magnitude of the correlation coefficients, little significance may be assigned to this fact.

CONCLUSION

The low relationship between rate of learning and the flight criteria is not due to an unfortunate choice of points at which the rates were taken in the previous analysis.

Assuming that the curves fit the data adequately, the conclusion suggests that there is no relationship between the rate of learning as measured by the serial coordination test and the flight criteria (with the experimental procedures and the technique of curve fitting employed in this research).

51(Continued) The chief advantage of this formula (3) is that the values of the constants in it need be determined but once for the set of data, and the value of r may be determined for any point on the fitted curve between the slope at that point and the measure by substituting the value of the point in the formula. It should also be indicated that the values needed to get the constants in formula (3) can be very readily determined with the use of an I.B.M. tabulator, sorter, and an adding machine. The critical values that need be punched into cards are only a, b, c, and Y. From this all the sums of cross-products and sums of squares can be obtained.

Examination of the constants in formula (3) indicates that the correlation coefficient between the fitted curve and the measure is a function of the intercorrelations of the constants a, b, and c, and of the correlations of a, b, and c with the measure. It is apparent that the larger the values of r_{aY}, r_{bY}, or r_{cY}, the larger will be the value of the correlation between the rate of the fitted curve and the measure. Furthermore, the larger the value of r_{ab}, r_{ac}, and r_{bc}, the lower will be the value of the correlation between the rate of the fitted curve and the measure.

APPENDIX E

TABLES OF DATA

TABLE E-1

MEAN RAW SCORES FOR EACH TRIAL, PSYCHOMOTOR TESTS

Rotary Pursuit (.001 min.) Group			Serial Coordination Group			Two-Hand Coordination (.001 min.) Group		
A	B	A+B	A	B	A+B	A	B	A+B
21.5	25.6	23.2	17.8	19.95	18.85	206.8	198.3	202.6
30.2	34.8	32.1	20.1	21.70	20.92	181.8	208.6	195.2
32.9	37.5	34.8	21.9	23.69	22.81	255.6	277.6	266.7
36.3	40.7	38.1	23.3	24.93	24.13	275.2	282.0	278.6
37.9	40.3	38.9	24.4	25.84	25.14	274.6	273.6	274.1
40.5	43.2	41.6	25.2	25.98	25.58	277.3	291.2	284.3
41.5	48.5	44.3	25.5	27.79	26.64	300.6	311.1	305.9
44.3	50.4	46.8	26.3	28.22	27.25	296.9	308.0	302.5
51.6	51.8	51.7	27.2	28.94	28.06	304.1	321.8	313.0
51.0	52.5	51.6	27.6	29.27	28.45	279.3	278.4	278.9
52.1	59.1	55.0	27.8	30.11	28.96	342.5	337.9	340.2
52.8	60.2	56.1	28.7	30.88	29.77	357.9	366.8	362.4
54.7	65.1	59.0	28.7	31.02	29.87	329.8	338.5	334.2
56.8	64.0	59.7	29.4	31.78	30.61	335.6	349.8	342.7
58.9	67.0	62.2	29.9	32.09	30.99	355.6	357.2	356.4
61.4	69.3	64.6	30.0	32.45	31.21	342.2	358.9	350.0
65.8	73.6	69.0				352.2	354.8	353.5
65.7	74.5	69.3				329.9	340.7	335.3
67.3	77.8	71.6				371.9	376.8	374.4
68.9	77.7	72.5				384.1	392.8	388.5
70.0	79.8	74.0				367.7	366.0	366.9
75.2	83.6	78.7				376.4	380.6	378.5
74.6	83.2	78.2				389.9	381.3	385.6
76.5	85.8	80.3				377.5	386.9	382.2
77.3	88.3	81.8				371.0	386.2	378.6
82.2	85.1	83.4				354.5	369.3	361.9
81.4	91.1	85.4				401.9	393.2	397.6
81.5	89.2	84.7				399.4	415.8	407.6
83.5	94.5	88.0				388.2	402.5	395.4
83.9	94.2	88.1				398.3	403.6	401.0
86.2	96.4	90.4						
86.6	94.6	89.8						
86.9	98.3	91.6						
89.5	102.0	94.6						
90.2	99.4	94.0						
93.2	99.2	95.7						
93.8	102.3	97.3						
91.5	102.4	96.0						
95.3	103.5	98.6						
98.2	103.3	100.3						
96.5	108.8	101.5						
100.1	103.8	101.6						
99.3	104.9	101.6						
99.5	111.0	104.2						
104.1	108.0	105.7						
102.2	110.6	105.6						
101.5	110.0	105.0						
103.4	105.0	106.4						
102.8	110.8	106.0						
100.8	109.8	104.5						

TABLE E-1 (Continued)

MEAN RAW SCORES FOR EACH TRIAL, PSYCHOMOTOR TESTS

Division-of-Attention Test (.01 Sec.)

Reaction Time			Track Time			Net Time		
Group			Group			Group		
A	B	A+B	A	B	A+B	A	B	A+B
17.6	18.58	18.11	11.38	12.08	11.73	5.67	5.91	5.79
15.1	15.00	15.06	14.17	14.52	14.35	8.08	8.52	8.30
13.7	14.87	14.31	14.90	15.46	15.18	9.46	9.00	9.23
13.6	13.52	13.55	14.81	15.78	15.30	9.47	10.10	9.78
12.8	13.88	13.33	15.20	15.67	15.44	9.98	9.84	9.91
12.7	12.59	12.63	15.69	16.20	15.95	10.35	11.44	10.90
12.4	12.88	12.65	15.95	16.08	16.02	10.41	10.82	10.62
12.6	11.87	12.21	16.08	16.75	16.42	10.93	11.68	11.31
12.3	11.68	12.00	16.38	17.14	16.76	11.12	11.91	11.52
12.0	11.19	11.58	16.91	17.37	17.14	11.55	12.43	11.99
12.0	11.19	11.59	17.05	17.94	17.50	12.02	12.64	12.33
11.5	10.99	11.24	17.48	17.90	17.69	12.18	12.99	12.59
11.2	10.54	10.84	17.53	18.74	18.14	12.33	13.52	12.92
11.5	9.73	10.64	17.54	19.01	18.28	12.25	14.27	13.26
10.6	10.09	10.37	18.10	19.36	18.73	12.93	14.17	13.55
10.7	10.15	10.41	18.27	19.14	18.71	13.12	14.23	13.68
10.2	9.55	9.89	18.74	20.26	19.50	13.58	15.04	14.31
10.5	9.70	10.12	18.74	19.80	19.27	13.44	14.94	14.19
10.3	9.40	9.84	18.75	20.20	19.48	13.61	15.16	14.39
10.5	9.30	9.89	18.81	20.06	19.44	13.54	15.41	14.48
10.1	9.24	9.68	19.05	20.18	19.62	13.78	15.18	14.48
10.2	9.23	9.75	19.27	20.17	19.72	14.04	15.35	14.70
9.9	8.92	9.39	19.53	20.45	19.99	14.36	15.65	15.01
9.7	8.96	9.32	19.54	20.29	19.92	14.48	15.76	15.12
9.6	9.03	9.31	19.68	20.33	20.00	14.54	15.58	15.06
9.5	8.64	9.05	20.02	20.60	20.31	14.92	16.07	15.50
9.6	8.78	9.19	19.88	21.06	20.47	14.65	16.10	15.38
9.5	8.35	8.91	19.99	20.94	20.47	15.01	16.50	15.76
9.3	8.36	8.22	20.17	21.16	20.67	14.86	16.38	15.62
9.4	8.23	8.83	20.26	20.80	20.53	14.95	16.40	15.68

TABLE E-2

MEANS AND STANDARD DEVIATIONS FOR ALL TYPES OF SCORE* ON ALL TESTS AND MEASURES

GROUP A

Test**		Unit	Initial Score	Central Score	Final Score	Total Score	Gain I-C	Gain C-F	Gain I-F
R.P.	M	.001 min.	233.8	517.0	616.8	3668.0	284.9	99.2	384.1
	σ	.001 min.	92.8	130.7	118.7	865	100.3	68.5	91.5
S.C.	M	Matches	37.9	56.5	59.9	414.7	18.5	3.5	22.0
	σ	Matches	6.5	7.0	7.2	48.7	5.6	4.0	6.7
2-H	M	.001 min.	1081.6	1445.8	1597.2	9568.0	353.6	161.7	515.2
	σ	.001 min.	204.0	198.4	168.6	1096	192.8	162.2	193.4
N.T.	M	.01 sec.	3944	5458	5982	35,980	1513	528	2041
	σ	.01 sec.	1008	1141	1153	6910	868	699	1028
T.T.	M	.01 sec.	6083	7470	8003	50,670	1410	541	1951
	σ	.01 sec.	1088	1371	1449	8270	936	928	1131
R.T.	M	.01 sec.	5317	4171	3796	30,870	1154	386	1540
	σ	.01 sec.	1107	972	801	5990	986	641	995

*Status scores represent sums, viz., on the rotary pursuit test, 6 trials of 10 sec. each; on the serial coordination test, 2 trials of 2 min. each; on the two-hand coordination and division-of-attention tests, 4 trials of 30 sec. each. Gain scores are gains from the sum of one segment of trials to the sum of another segment of trials.

**R.P. (rotary pursuit test), S.C. (serial coordination test), 2-H (two-hand coordination test), N.T. (net time, division-of-attention test), T.T. (track time, division-of-attention test), R.T. (reaction time, division of-attention test).

TABLE B-2 (Continued)

MEANS AND STANDARD DEVIATIONS FOR ALL TYPES OF SCORE ON ALL TESTS AND MEASURES

GROUP B

Test		Unit	Initial Score	Central Score	Final Score	Total Score	Gain I-C	Gain C-F	Gain I-F
R.P.	M	.001 min.	260.6	579.5	659.5	4014	319.7	80.5	400.2
	σ	.001 min.	119.4	135.2	125.0	994	93.2	55.0	101.0
S.C.	M	Matches	41.6	60.9	64.4	44[illegible].3	19.3	3.5	22.8
	σ	Matches	7.0	7.0	7.3	50.2	5.6	4.6	5.7
2-H	M	.001 min.	1128.2	1470.2	1613.3	9795	343.0	146.5	489.5
	σ	.001 min.	219.0	207.0	169.9	1222	154.3	142.3	171.2
N.T.	M	.01 sec.	4043	6071	6559	38,950	2032	487	2519
	σ	.01 sec.	899	1073	968	6130	852	682	910
T.T.	M	.01 sec.	6383	8113	8465	53,460	1746	351	2097
	σ	.01 sec.	1023	1183	1149	7340	922	648	1063

GROUPS A and B Combined

Test		Unit	Initial Score	Central Score	Final Score	Total Score	Gain I-C	Gain C-F	Gain I-F
R.P.	M	.001 min.	244.8	542.7	634.4	3310	299.2	91.5	390.7
	σ	.001 min.	105.4	136.1	123.1	936	98.9	64.0	95.9
S.C.	M	Matches	39.4	58.3	61.8	426.9	18.8	3.5	22.3
	σ	Matches	6.9	7.3	7.6	51.7	5.6	4.3	6.3
2-H	M	.001 min.	1100.8	1455.8	1603.9	9673	349.2	155.4	504.6
	σ	.001 min.	211.6	202.4	169.3	1154	178.0	154.5	215.4
N.T.	M	.01 sec.	3985	5710	6220	37,200	1727	511	2238
	σ	.01 sec.	966	1154	1117	6760	899	692	1009
T.T.	M	.01 sec.	6207	7735	8193	51,820	1548	463	2011
	σ	.01 sec.	1072	1335	1353	8020	945	830	1106

TABLE E-3

INTRA-TEST CORRELATIONS (PSYCHOMOTOR)

	Test	Central Score Group A	Central Score Group B	Central Score Group A+B	Final Score Group A	Final Score Group B	Final Score Group A+B	Gain (I-C) Group A	Gain (I-C) Group B	Gain (I-C) Group A+B
Initial Score	R.P.	.66	.73	.70	.65	.65	.65	-.07	-.22	-.11
	S.C.	.63	.68	.68	.51	.68	.62	-.37	-.40	-.35
	2-H	.47	.74	.59	.40	.61	.49	-.47	-.42	-.45
	N.T.	.68	.64	.66	.56	.53	.54	-.27	-.26	-.24
	T.T.	.74	.65	.71	.64	.52	.60	-.08	-.27	-.13
	R.T.	.56	-	-	.50	-	-	(.56)*	-	-
Central Score	R.P.				.86	.91	.89	.70	.51	.63
	S.C.				.84	.79	.84	.48	.40	.45
	2-H				.60	.71	.65	.48	.30	.41
	N.T.				.82	.78	.82	.51	.58	.57
	T.T.				.78	.84	.81	.62	.55	.61
	R.T.				.75	-	-	(-.37)	-	-
Final Score	R.P.							.52	.49	.52
	S.C.							.42	.14	.31
	2-H							.16	.09	.13
	N.T.							.42	.43	.46
	T.T.							.41	.50	.46
	R.T.							(-.19)	-	-
Gain (I-C)	R.P.									
	S.C.									
	2-H									
	N.T.									
	T.T.									
	R.T.									
Gain (C-F)	R.P.									
	S.C.									
	2-H									
	N.T.									
	T.T.									
	R.T.									
Gain (I-F)	R.P.									
	S.C.									
	2-H									
	N.T.									
	T.T.									
	R.T.									

*Values in parentheses should be read with signs reflected.

TABLE E-3 (Continued)

INTRA-TEST CORRELATIONS (PSYCHOMOTOR)

		Gain (C-F)			Gain (I-F)			Total Score		
		Group			Group			Group		
	Test	A	B	A+B	A	B	A+B	A	B	A+B
Initial Score	R.P.	-.13	-.33	-.22	-.19	-.38	-.26	.77	.84	.81
	S.C.	-.17	.06	-.06	-.41	-.35	-.35	.76	.86	.82
	2-H	-.17	-.32	-.23	-.62	-.64	-.63	.68	.82	.74
	N.T.	-.19	-.26	-.22	-.36	-.43	-.36	.78	.77	.77
	T.T.	-.09	-.26	-.15	-.14	-.40	-.22	.85	.77	.82
	R.T.	(.23)*	-	-	(.71)	-	-	.75	-	-
Central Score	R.P.	-.42	-.40	-.43	.45	.25	.37	.93	.96	.95
	S.C.	-.22	-.26	-.22	.27	.19	.25	.93	.87	.91
	2-H	-.55	-.58	-.56	.02	-.22	-.08	.83	.92	.87
	N.T.	-.27	-.47	-.35	.25	.19	.28	.92	.93	.93
	T.T.	-.25	-.32	-.29	.30	.28	.30	.90	.95	.92
	R.T.	(.57)	-	-	.00	-	-	.90	-	-
Final Score	R.P.	.09	.02	.04	.63	.46	.56	.91	.91	.91
	S.C.	.35	.39	.35	.56	.45	.51	.88	.89	.89
	2-H	.31	.15	.25	.42	.21	.33	.78	.84	.81
	N.T.	.32	.18	.25	.57	.54	.59	.90	.85	.89
	T.T.	.41	.24	.34	.68	-.58	.64	.88	.88	.88
	R.T.	(-.11)	-	-	-.26	-	-	.84	-	-
Gain (I-C)	R.P.	-.46	-.15	-.37	.75	.84	.79	.50	.31	.43
	S.C.	-.06	-.39	-.21	.80	.67	.75	.25	.01	.16
	2-H	-.42	-.34	-.39	.65	.62	.64	.12	.09	.10
	N.T.	-.15	-.31	-.22	.74	.70	.74	.30	.35	.36
	T.T.	-.26	-.12	-.23	.61	.80	.68	.34	.35	.37
	R.T.	-.31	-	-	.79	-	-	(-.06)	-	-
Gain (C-F)	R.P.				.24	.41	.29	-.20	-.30	-.25
	S.C.				.55	.42	.49	-.35	.10	.03
	2-H				.42	.53	.46	-.21	.32	-.26
	N.T.				.55	.46	.49	-.28	-.25	-.12
	T.T.				.60	.51	.56	.05	-.16	-.03
	R.T.				.34	-	-	(.32)	-	-
Gain (I-F)	R.P.							.40	.12	.28
	S.C.							.19	.09	.16
	2-H							-.06	-.19	-.12
	N.T.							.24	.14	.24
	T.T.							.33	.19	.29
	R.T.							(.15)	-	-

*Values in parentheses should be read with signs reflected.

Table [illegible]

INTER-TEST CORRELATIONS (PSYCHOMOTOR)

Test	Score	Serial Coordination A	Serial Coordination B	Serial Coordination A+B	Two-Hand Coordination A	Two-Hand Coordination B	Two-Hand Coordination A+B	Net Time A	Net Time B	Net Time A+B	Track Time A	Track Time B	Track Time A+B	Reaction Time A
Rotary Pursuit	I	.02	.27	.17	.27	.10	.15	.05	-.01	.03	.25	.06	.18	(.14)*
	C	.12	.37	.20	.07	.14	.11	.27	.03	.22	.32	.06	.26	(-.05)
	F	.20	.33	.25	.32	-.05	.05	.24	.06	.20	.21	.09	.19	(-.03)
	T	.12	.39	.26	.15	.11	.14	.31	.06	.24	.34	.11	.27	(-.05)
	I-C	.21	.13	.19	.03	-.11	-.02	-.02	.02	.05	.09	.01	.09	-.17
	C-F	.00	-.16	-.06	-.02	.08	.02	-.09	.10	.07	.00	.10	.04	.13
	I-F	.19	.24	.21	.10	-.22	-.03	-.13	.04	-.04	-.09	.00	-.05	-.29
Serial Coordination	I				.12	.35	.23	.02	.32	.24	.19	.20	.23	(-.17)
	C				.34	.31	.33	.35	.27	.37	.28	.29	.33	(-.26)
	F				.13	.28	.19	.30	.43	.39	.24	.35	.31	(-.31)
	T				.24	.44	.34	.33	.37	.39	.30	.30	.33	(-.28)
	I-C				.21	.00	.13	.13	.10	.13	.16	.16	.17	.11
	C-F				-.05	-.28	-.15	-.03	-.16	-.12	-.01	-.22	-.09	-.06
	I-F				.05	-.15	-.02	.09	.14	.12	.01	.11	.05	.12
Two-Hand Coord.	I							.12	.36	.22	.17	.27	.23	(-.08)
	C							.17	.28	.22	.07	.31	.17	(-.12)
	F							.12	.31	.19	.06	.28	.15	(-.10)
	T							.23	.35	.29	.19	.37	.27	(-.20)
	I-C							-.08	.09	-.03	-.10	.11	-.03	.00
	C-F							.10	.22	.15	.10	.15	.12	-.12
	I-F							.12	.11	.10	.01	.01	.00	.18
Net Time	I										.81	.77	.80	(-.74)
	C										.86	.86	.87	(-.62)
	F										.88	.80	.85	(-.59)
	T										.89	.84	.87	(-.65)
	I-C										.75	.81	.78	.66
	C-F										.61	.85	.68	.62
	I-F										.79	.85	.80	.64
Track Time	I													(-.35)
	C													(-.32)
	F													(-.28)
	T													(-.30)
	I-C													.27
	C-F													.22
	I-F													.25

*All values in parentheses should be read with signs reflected.

TABLE E-5

CORRELATIONS BETWEEN CRITERIA AND PSYCHOMOTOR TESTS

Test	Criterion	Initial Segment Group A	B	A+B	Central Segment Group A	B	A+B	Final Segment Group A	B	A+B
Rotary Pursuit	A	.12	.33	.22	.17	.32	.23	.17	.24	.20
	B	.15	.26	.20	.22	.26	.23	.19	.21	.20
	C	.20	.59	.38	.27	.27	.26	.26	.21	.23
	D	.13	-	-	.15	-	-	.15	-	-
Serial Coordination	A	.15	.42	.27	.24	.30	.26	.23	.24	.23
	B	.06	.30	.16	.23	.26	.23	.27	.21	.24
	C	.12	.32	.19	.28	.25	.25	.31	.20	.24
	D	.08	-	-	.26	-	-	.29	-	-
Two-Hand Coordination	A	.25	.35	.29	.32	.31	.32	.08	.33	.19
	B	.15	.34	.23	.28	.30	.29	.06	.34	.17
	C	.22	.33	.27	.35	.24	.30	.00	.30	.13
	D	.17	-	-	.29	-	-	.13	-	-
Net Time	A	.18	.27	.21	.27	.31	.28	.27	.25	.26
	B	.11	.30	.18	.22	.28	.23	.31	.16	.24
	C	.08	.62	.29	.20	.29	.22	.21	.13	.16
	D	.09	-	-	.20	-	-	.20	-	-
Track Time	A	.26	.17	.22	.20	.29	.24	.31	.20	.27
	B	.16	.24	.19	.15	.30	.20	.27	.16	.23
	C	.19	.19	.19	.17	.32	.22	.23	.16	.19
	D	.15	-	-	.11	-	-	.14	-	-
Reaction Time*	A	-.02			-.11			-.15		
	B	-.06			-.22			-.27		
	C	.01			-.10			-.12		
	D	-.06			-.25			-.26		

*All values for reaction time should be read with signs reflected.

TABLE E-5 (Continued)

CORRELATIONS BETWEEN CRITERIA AND PSYCHOMOTOR TESTS

Gain (I-C)			Gain (C-F)			Gain (I-F)			Total Score		
Group			Group			Group			Group		
A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
.12	.10	.11	-.04	-.32	-.14	.10	-.08	.02	.16	.36	.25
.15	.04	.11	-.07	-.16	-.10	.11	-.05	.04	.20	.28	.23
.17	.01	.10	-.06	-.18	-.10	.14	-.09	.03	.26	.28	.26
.07	-	-	-.01	-	-	.07	-	-	.13	-	-
.12	-.15	.01	.00	-.07	-.03	.10	-.21	-.02	.27	.37	.31
.21	-.04	.11	.10	-.07	.03	.23	-.10	.11	.24	.31	.25
.22	-.09	.09	.06	-.07	.00	.22	-.14	.68	.28	.32	.28
.22	-	-	.07	-	-	.23	-	-	.26	-	-
.07	-.09	.01	-.32	-.04	-.21	-.20	-.12	-.17	.29	.38	.33
.09	-.08	.03	-.28	-.03	-.19	-.14	-.10	-.13	.22	.38	.29
.10	-.15	.01	-.42	.02	-.24	-.25	-.12	-.20	.23	.31	.26
.11	-	-	-.24	-	-	.09	-	-	.28	-	-
.14	.10	.13	.02	-.13	-.04	.13	.00	.08	.25	.30	.27
.14	.02	.09	.17	-.21	.02	.23	-.14	.09	.24	.28	.25
.15	.10	.12	.03	-.26	-.09	.15	-.11	.04	.16	.26	.19
.15	-	-	.01	-	-	.13	-	-	.18	-	-
.01	.19	.09	.18	-.16	.06	.15	.06	.12	.30	.26	.28
.04	.13	.07	.19	-.27	.04	.18	-.05	.09	.23	.27	.25
.02	.57	.25	.09	-.30	-.04	.10	.31	.18	.21	.27	.23
-.01	-	-	.04	-	-	.03	-	-	.16	-	-
.08			.01			.08			-.08		
.14			-.01			.14			-.20		
.11			-.01			.10			-.06		
.18			-.06			.14			-.20		

TABLE [illegible]-6

VARIABLES USED IN SECURING MULTIPLE CORRELATION COEFFICIENTS
(Additional Intercorrelations)

	Group	Hidden Figures I	Path Distance	Mechanical Comprehension	Net Time (F)	Serial Coordination (F)	Net Time (F)	Serial Coordination (F)	Two-Hand Coordination (C)	Rotary Pursuit (F)	Serial Coordination (I-C)	Two-Hand Coordination (C-F)	Track Time (F)
Path Distance	A	.18											
	B	.28											
	A+B	.21											
Mechanical Comprehension	A	.21	.36										
	B	.24	.46										
	A+B	.21	.37										
Net Time (F)	A	.14	.25	.32									
	B	.09	.33	.28									
	A+B	.11	.25	.31									
Serial Coordination (F)	A	.05	-.01	.23	.30								
	B	.17	.19	.26	.45								
	A+B	.09	.11	.25	.41								
Two-Hand Coordination (C)	A	.14	.02	.[illegible]	.13	.29							
	B	-.05	.12	.36	.33	.33							
	A+B	.06	.10	.32	.21	.33							
Serial Coordination (F)	A						.30						
	B						.45						
	A+B						.41						
Two-Hand Coordination (C)	A						.13	.29					
	B						.33	.33					
	A+B						.21	.32					
Rotary Pursuit (F)	A						.24	.20	.14				
	B						.09	.33	.09				
	A+B						.22	.29	.13				
Serial Coordination (I-C)	A						.21	.42	.18	.17			
	B						.11	.17	.07	.26			
	A+B						.15	.33	.34	.22			
Two-Hand Coordination (C-F)	A						-.01	-.20	-.55	-.01	-.05		
	B						-.10	-.21	-.58	.[illegible]	-.03		
	A+B						-.05	-.21	-.56	-.02	-.02		
Track Time (F)	A						.88	.24	.05	.21	.15	.27	
	B						.80	.36	.34	.11	.14	-.15	
	A+B						.81	.31	.17	.19	.17	-.0[illegible]	
Hidden Figures (I)	A						.14	.05	.14				
	B						.09	.17	-.05	-.06	-.19	.04	-.07
	A+B						.11	.09	.05				
Mechanical Comprehension	A						.34	.23	.28				
	B						.23	.25	.36	[illegible]	[illegible]	[illegible]	[illegible]
	A+B						.31	.26	[illegible]				