THE DETERMINATION OF THE VALUE OF A BATTERY OF PSYCHOLOGICAL TESTS IN PREDICTING PILOT PROFICIENCY

By

STEVE ALLAN HECKART

Bachelor of Science

Oklahoma Agricultural and Mechanical College

Stillwater, Oklahoma

1956

Submitted to the faculty of the Graduate School of the Oklahoma State University of Agriculture and Applied Science in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE August, 1957

THE DETERMINATION OF THE VALUE OF A BATTERY OF PSYCHOLOGICAL TESTS IN PREDICTING PILOT PROFICIENCY

Thesis Approved:

William H. Rambo

Thesis Adviser

Robert W. Scofield

Dean of the Graduate School

PREFACE

This study came as a result of the need of the Oklahoma State University School of Aviation Education and Flight Training to know which students they could expect to be deficient in pilot aptitudes. The study is an attempt to predict these students by utilizing a battery of prevalent psychological tests. This need is not limited to Oklahoma State University, but instead, is widely felt in these times of rapidly increasing enrollment; the enrollment problem necessitates knowledge of student capabilities if maximum utilization of staff and resources is to be achieved.

Indebtedness is gratefully acknowledged to the following for aid in a multitude of ways: the staff and students of the School of Aviation Education and Flight Training; Dr. H. K. Brobst and his staff of the Bureau of Tests and Measurements; Robert J. Russell and his staff of the Veterans Administration Counseling Center; Dr. John W. Hamblen and his staff of the Oklahoma State University Computing Center; and to Drs. Robert W. Scofield and L. M. Gustafson for guidance on many points, both technical and practical.

A most singular source of succor has been my Thesis Advisor, Dr. William W. Rambo, who since the inception of the study has so often provided the catalyst that enabled the author to consumate the study. He has my deepest thanks; as has my brother, Coleman, who helped immeasurably with the gathering and processing of the data.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
II.	METHOD	9
III.	RESULTS	16
IV.	DISCUSSION	23
V.	SUMMARY	26
VI.	A SELECTED BIBLIOGRAPHY	28
VII.	APPENDIX A	31
VIII.	APPENDIX B	33

LIST OF TABLES

Table | Page

I. Measures of Central Tendency and Variability Obtained From the Predictor and Criterion Variables 17

II. Intercorrelations of Thirteen Tests and a Criterion 18

III. Beta Weights and b Weights for Thirteen-Variable Regression Equations 20

IV. Beta Weights and b Weights for Three-Variable Regression Equations 21

CHAPTER I

INTRODUCTION

This paper represents an attempt to predict performance in pilot training in the School of Aviation Education and Flight Training at Oklahoma State University of Agriculture and Applied Science. Being concerned with adequacy and competency of instruction in assessing the worth of the training received by the students participating in the flight program, it was felt by the School of Aviation Education staff that the students' training could be enriched and attrition reduced if some method of predicting performance in the program could be devised. This was not with the goal in view of restricting entry into the program but, instead, of isolating individuals deficient in pilot aptitudes and devoting additional time and effort to these persons early in the program when they may benefit most. It was felt that this procedure would also contribute toward the maximum utilization of the efforts of the staff of the School of Aviation Education and Flight Training.

This is not an isolated problem faced only by the staff of the flying school at Oklahoma State University, but instead, may be considered to be typical of those faced by colleges and universities in this day of rapidly expanding enrollment. In a study prepared for the Civil Aeronautics Administration by the American Council on Education (28) it was reported that 399 institutions of 1500 surveyed were offering academic work in aviation, with many of these programs culminating in the pilot's

license.

Razrah and Brown (24) stated in 1941,

> It is really difficult to think of another field of such practical importance as that of the selection of aircraft pilots in which so much confusion reigns and in which research has been attempted and interpreted by investigators of such varying background and training.

Six years later and after exceedingly extensive work by the Armed Services Lane (14) is found saying,

> . . . although there have been many attempts to devise methods for the selection of individuals who if given proper training would be able to pilot airplanes with some degree of competence, the field has been only partially explored.

A survey of the literature revealed that few investigations have been reported in the recent past by individuals on the selection and prediction of pilots except those reporting on work done by the military (12, 14, 17).

The problem, then, was one of selecting a battery of predictors and the subsequent validation of those measures selected. In order to select instruments of prediction in any but the most haphazard fashion, it was deemed necessary that one review the research that had been done in this area in order to provide background for the development of the initial battery of potential predictors.

Survey. The history of the selection and prediction of aviators goes back to World War I; however, the effectiveness of the work before World War II was, in general, scattered and of doubtful importance from the standpoint of the contemporary worker. The reader who has need of more detailed information of this period is urged to consult Razrah and Brown (24) for a rather complete bibliography prior to 1941 or Jenkins (12) for a broad inclusive resume of the approaches and results of principle investigations. Jenkin's task was not a lengthy one, for as he remarked,

Following the war, these [World War I] researches were dropped and it is a tragic fact that, during the next two decades, practically no psychological researches into problems in aeronautics were carried on in this country.

He further stated that,

Despite the rapid growth of commercial air-transport, and despite the repeated demonstrations of the military importance of aircraft, the field was as unfamiliar to psychologists at large in 1940 as it was in 1917.

The next major emphasis on selection and prediction work after that of the military in World War I came in 1939 when the Civil Aeronautics Authority (now the Civil Aeronautics Administration) underwrote a broad program of training civilian pilots. This program has been extensively reported by Viteles (30). The CAA program, known as the Civilian Pilot Training program, was designed to function through the colleges and universities of the country, making use of available private facilities and it encouraged others to be set up. The program was very inclusive and much of the work on pilot selection, pilot training, pilot rating, evaluation of common techniques of evaluation, airsickness, accidents, and emotional disturbances was put to good use with the outbreak of hostilities in 1941 and the subsequent mobilization. The Committee on Pilot Training of the CAA found several predictors to be of value. These tests were principally pencil-and-paper and psychomotor tests. In general, the tests fell into groupings on the basis of intelligence, mechanical comprehension, perception-coordination, and interest. Personality, attitudes and other measures were taken but the results were inconclusive.

Viteles (30), in his article on the CAA investigation, did not give estimates of reliability or validity but simply stated whether the instruments are "of value" or perhaps, "does not discriminate." Some

of the measures that were mentioned by Viteles' report as being good predictors were 1) The Biographical Inventory, which was listed as being a successful attempt to predict pilot proficiency from biographical data, 2) General Intelligence Test, 3) a Mechanical Comprehension Test, and 4) Psychomotor Tests. These psychomotor tests included the Eye-Hand Coordination Test, the Two-Hand Coordination Test, The Rotary Pursuit, and others. Melton (19) reported the Rotary Pursuit Test was

> . . . originally introduced in the psychomotor classification battery [of the Army Air Forces] on a hunch supported by a bit of data which had come from a study sponsored by the National Research Council Committee on Selection and Training of Aircraft Pilots.

The testing programs of the AAF and the Navy adopted considerable of the work done by the various groups receiving CAA funds totaling over $900,000 during the five year period of the life of the Civilian Pilot Training Program.

Views of Tests, Traits, and Criteria. Pilot selection research gained much impetus at the time when it became apparent that the United States was to be embroiled in World War II and that huge numbers of qualified pilots would be a necessity (5, p. 7). Although the ultimate requirements of military flying were different from those expected of a civilian light plane pilot, in that the ultimate test of the military aviator may be his combat performance, many of the problems encountered by the military in the selection of persons with the aptitudes to learn the art of piloting were the same, as were many of the problems of the establishment of ratings and criterion. Davis (4, p. 11) reports that, in general, the military of the Allied Forces, as well as several civilian groups, cooperated in the exchange of data. For this reason

reports of military investigations were limited largely herein to the AAF Aviation Psychology Program that bore most directly on the selection procedures. It was believed that these were representative of the gamut and they appeared to be fruitful.

Perhaps the first item for consideration should be the traits that the military deemed involved in the flying process; those aptitudes and abilities that the predictive test measures. Flanagan (6), Head of the Aviation Psychology Program in the AAF, delineated the area in this way,

> In order to select the individuals for training as Aviation Cadets, it is necessary to know: The minimum level of general intelligence or academic aptitude which will enable the cadet to absorb the necessary instruction and training and to perform his duties in a satisfactory manner; the minimum requirements in alertness, speed of decision and reaction, and in the powers of rapid and accurate observations of details; the minimum level through complicated maneuvers smoothly and precisely; and finally, it is necessary to know the most suitable amounts of aggressiveness, fearlessness, calmness, and similar personality traits. To measure these traits, a well-balanced battery of tests must be applied.

Liljencrantz (15) voiced this view on the distribution of traits,

> The laws of probability preclude occurrence of individuals who excell in respect to all of many unrelated desirable traits. Our problem therefore focuses on the identification of individuals whose outstanding characteristic is a balanced broad ability, rather than those who possess single desirable traits in extraordinary degree.

Lane (14), working with civilians, took this approach,

> The abilities required in successful piloting are best described as a complex of coordination, skills, and abilities. Therefore, adequate pilot selection would best be accomplished through a method which combined various measures of the components of this complex.

Tests that the AAF has found to be of value may be grouped into the following categories, 1) Tests of intelligence, judgment, and proficiency, 2) Tests of alertness, observation, and speed of perception, 3) Tests of personality, temperament, and interest, and 4) Tests of

visual-motor coordination (16, 19, 22).

In developing tests to measure traits believed to be important for success in a particular classification as piloting, a guide was found to selection set forth by Flanagan (7, p. 63),

> . . . the points which should be specifically considered should include (1) validity, (2) independence, (3) simplicity, (4) stability, (5) objectivity, (6) acceptability, (7) practicality, (8) atypical performance, and (9) discrimination.

Also, on methodology, he says,

> From the outset it was decided that all types of testing, observations, questionaires, and interview procedures would be tried out. Because of the large numbers of persons to be tested it was agreed that objective, printed, multiple-choice tests which could be scored by machines were to be prefered if the trait was susceptible to measurement by this method. It was believed essential to include certain apparatus tests of coordination and speed of decision, at least until such time as it could be adequately measured by more efficient devices (7, p. 63).

On this tack of devices, Michael (21) did a factor analysis of a number of AAF tests and states,

> The presence of a factor identified as spatial relations, in a pencil-and-paper test as well as in apparatus tests, has suggested the potential economy of pencil-and-paper tests in the measurement of human abilities frequently effected by more cumbersome devices.

With this orientation, the AAF developed a battery of psychological tests containing

> . . . 12 printed tests and 5 apparatus tests which had been especially developed to measure those traits believed to be essential to success in the various air-crew specialities [pilot, navigator, and bombadier]. This battery included four different types of mathematics tests believed to be especially important for the navigator; tests of dial and table reading also believed to be of primary importance in selecting navigators; three tests involving speed of perception and recognition of forms which were considered to be especially important to pilots and the bombadier; a test of mechanical comprehension considered to be essential for the pilot; a test of reading and judgment considered to be important for the navigator; as well as for all three positions; and a technical-vocabulary test containing separate parts for pilot, bombadier, and navigator scores. The apparatus tests of complex coordination and twohand coordination were regarded as primary requirements of the pilot . . . (7, p. 64).

Correlation of these tests for the pilot with the criterion range from approximately .30 to .50, depending upon sampling fluctuations (7, p. 81).

Any research program on problems of selection and training must become involved with general problems of evaluating proficiency, otherwise one will never be able to place any degree of confidence in his validation work. In the Aviation Psychology Program it was reported that, "The proficiency measure which appears to be clearly superior to other measures available for these types of training was graduation or elimination from the training schools" (7, p. 115). This was used as the validation criteria even though letter grades of A, B, C, D, E, or F were assigned to the student-pilots, as reported by the Staff, Psychological Research Project (Pilot) (26). They also reported that, ". . . flying performance rating of students in the AAF is almost entirely subjective (26)." And, in addition, they report, ". . . with subjective methods, considerable differences exist in the standards of various schools and of the same schools on successive classes (26)."

Early efforts to develop objective measures of flying skill uncovered three main sources of variation in the student's score that were unrelated to the skill being tested (7, p. 118). It was necessary that 1) judgment of all examiners be standardized, 2) the student know exactly what he was to do in order to maximize his score, and 3) the nature of the task be controlled (7, p. 118).

Summary. The evidence presented is of such a nature as to lead one to believe that it is possible to predict performance in pilot

training from psychological tests. Selection procedures have steadily improved for some four decades, even though progress has at times been retarded. Several traits apparently lie at the basis of the flying skill. The traits have been described as being measured by tests of intelligence, alertness, proficiency, personality, interest, and visual-motor coordination. Pencil-and-paper tests have proved to be equally as adequate for measuring many skills, as psychomotor tests, that were formerly thought to be indispensable. The tests were validated against pass-fail criterions based upon instructors' ratings. The ratings were an important aspect of the validation procedure.

It was the purpose of this study to attempt to validate a group of tests of both types for use in predicting performance in the School of Aviation Education and Flight Training at Oklahoma State University.

CHAPTER II

METHOD

Within this chapter will be found an elaboration on the subjects, the instruments used in gathering the data, and the general procedure that was followed throughout the study.

Subjects. The subjects used in this study were 62 students enrolled in three Aviation Education sections (ground school) at Oklahoma State University. Of these 37 were concurrently enrolled in the laboratory work (flight school). Twenty nine students of this group were included in the final experimental sample, the remainder having failed to complete their training for a host of reasons assumed to be not directly related to their flying ability. The subjects in the final sample were male, ranging in age from the late teens to the early thirties, the mean age being 24 years, 1 month. They represented a loose cross-section of the student body from the standpoint of major field of study and student academic level. Every school but Veterinary Medicine was represented; the mean academic level was 3.2 years in college.

Instruments. The instruments used in this investigation were of two general types: (1) the objective tests which were hoped to be predictors of pilot proficiency, (2) the rating scale of the demonstrated student proficiency at the end of training. Following is a

description of each of the instruments within each of these two categories.

Within the first category of instruments there were five objective tests. These tests were thought to be representative of the major trait areas that make up the complex of abilities associated with piloting an aircraft and were chosen because of the potential of prediction of the ratings of flying ability. The actual choice of tests was based upon a value judgment after the survey of the literature had been made; reports of the value of some of the tests in predicting pilot proficiency have been presented and these were recognized when the tests were chosen. Flanagan's (7, p. 63) guide list of attributes was followed where possible. Care was also taken to select tests readily available in order to increase the practicality of the proposed battery.

The First instrument chosen was the American Council on Education Psychological Examination for College Freshmen. This test yields three scores, quantitative, linguistic, and total, the last being the composite of the first two. The administration time is about one hour.

The Second instrument was the Test of Mechanical Comprehension. There were 60 items and 30 minutes were allowed the testees. Fiske (8) offered this word on the function of the test, "The purpose of the MCT is to measure knowledge of 'barnyard physics', not rote learning of textbook principles." Bennett's manual (1) expressed it this way, "The Mechanical Comprehension Tests are designed to measure the capacity of an individual to understand various types of physical and mechanical relationships." There was much evidence to attest to the value of this test in predicting pilots; it has been used by the AAF (27), the Navy

(26, 8), and in civilian work (10). Super (27, p. 353) related,

> One part of the AAF Qualifying Examination consisted of from 15 to 60, generally 30, Bennett-type items; Validity coefficients for various forms correlated with the success-failure in primary pilot training ranged from .14 to .38 . . .

The Manual (1) listed correlations ranging from .28 to .38 with Naval Aviation Cadets. The Navy used a pass-fail criterion.

The Third test selected was the Clerical Speed and Accuracy portion of the Differential Aptitude Tests battery. The administration time was listed at 15 minutes, with actual testee working time at six minutes. The test, according to the Manual (1), measured " . . . speed of perception, momentary retention, and speed of response." R. F. Berdie (3, p. 677), reviewing for Buros, described the task of the testee,

> The Clerical Speed and Accuracy test requires the testee to select combinations of letters and numerals marked in the test booklet and then to identify from a group of similar combinations the identical combination on the answer sheet.

It was believed that this test would serve as a perception, alertness, and observation test to fill one of the categories of traits to be measured as listed by the Staff of the Psychological Branch, Office of the Air Surgeon (16).

The Fourth test utilized was the Rotary Pursuit test, a psychomotor test originally designed by Dr. W. Koerth in 1922 (13). The test used in this investigation was essentially the same as the original model and that model used by the AAF. The apparatus consisted of a rotary turntable, a stylus, an electric timer, and an electric clock. In this model, the turntable was 10.9 inches in diameter, and a brass target, .75 inches in diameter was set flush with the turntable with

its center 3.12 inches from the center of rotation of the disk. The disk was powered by a phonograph motor, which for testing purposes was set at 60 r.p.m. through the adjustment of a brake on the govenor of the motor. The stylus used was a hinged brass rod, 6 inches in length, .08 inches in diameter, bent .9 inches from the end, that was mounted in a bakelite handle. A more detailed description of the complete apparatus may be found together with wiring diagrams in Melton (20, p. 333).

In this test the testee's task was to manipulate a stylus in such a manner as to maintain contact between the tip of the stylus and the target on the turntable while the turntable is in motion. The total time of contact was recorded by the electric clock during a 50 second interval regulated by the automatic timer. Each subject received ten trials, the last five of which were averaged for the recorded score on the test.

The AAF estimated the validity of the Rotary Pursuit test at .21 and gave it a multiple regression weight of 9 percent in the 1942 battery (20, p. 330). The test is thought to measure " . . . functions as perceptual-motor coordination, smoothness of control movement (20, p. 54).

The Fifth test utilized was the Guilford-Zimmerman Temperament Survey. The Guilford-Zimmerman was a 300-item, booklet form, machine-scored test covering ten facets of personality adjustment. Not a timed test the administration time was about 70 minutes. The ten traits measured by the Survey were G—General activity, R—Restraint, A—Ascendance, S—Sociability, E—Emotional stability, O—Objectivity,

F—Friendliness, T—Thoughtfulness, P—Personal relations, and M—Masculinity. The Manual (11) stated,

> The titles of the categories should be suggestive of the kind of adjustment or behavior to be expected in those with high or low scores. A high score indicates the 'positive' qualities and a low score the 'negative' scores. Extreme positive qualities do not always indicate the best adjustment, but extreme negative ones are likely to indicate trouble.

It was believed that the several traits measured by this test would facilitate the assessment of the personality-temperament trait category listed as crucial by Flanagan (6), Viteles (30), and others (16). Van Steenberg (29) says, "The survey gives a very favorable impression of a well-rounded, carefully worked out method of evaluating an important portion of the total personality."

In the second category of instruments used in this study, a rating scale of student proficiency was employed. This scale was adapted from that developed under the sponsorship of the National Research Council Committee on Selection and Training of Civilian Pilots as described by Viteles (30) and adopted by the CAA. The scale covers ten levels of efficiency in ten aspects of flying (See Appendix). This gives a theoretical range of scores from 0 to 100. The scale, as finally, was the joint effort of the author and the Chief Flight Instructor of the School of Aviation Education and Flight Training; the scale had the approval of the District Safety Inspector of the CAA. One of the Inspector's duties was to administer check-rides for applicants of Pilot License. All concerned felt that the scale represented an adequate measure of student proficiency.

Procedure. The tests were administered to the students at the beginning of the Spring semester, 1957, with the exception of the ACE. These scores were obtained from the files of the Oklahoma State University Bureau of Tests and Measurements; all students enrolling for the first time at Oklahoma State University take the test, usually when they are freshmen. The Mechanical Comprehension Test was administered in the regular class sessions. The Guilford-Zimmerman, being an untimed test, was used to fill out the hour; it was subsequently completed when the students came into the laboratory to be tested on the Rotary Pursuit Test. The Differential Aptitude Test of Clerical Speed and Accuracy was also given in this laboratory period. The test-makers' administration and scoring instructions were followed for the printed tests. The instructions used for the Rotary Pursuit Test were essentially those used by the AAF (20, p. 347 — See Appendix). Each subject was given ten 50 second trials, the last five of which were averaged for the typical performance measure of the subject. This was done in order to improve the stability of the scores on the test. Woodworth and Schlosberg (31, p. 788) refer to a "warm-up effect" that enters here and often leads to spurious results. It was believed that dropping the first five trials removed the warm-up effect and that nothing was lost by this procedure since the ten trials allowed each subject did not allow the task to be completely learned; curves plotted for the trials were still climbing at the tenth trial.

The rating scale used was adopted from that of the CAA. This, for the reason that the CAA issued check-rides and subsequent Pilot Licenses to students if the check-rides were passes. It was deemed necessary

that the students learn as a minimum those things necessary for qualification for the license. The material covered on this check-ride was well-standardized and covered items as safety, flying regulations and practices, as well as the actual handling of the aircraft. These items became the criterion; the rating scale that was developed for the use of the study was designed to measure these items, with the emphasis on the handling of the aircraft. The actual rating was to be made following the final check-ride administered by the school staff. If the student passed this hurdle he was deemed ready to be given a check-ride by the CAA in qualifying for his license. It was deemed appropriate that the rating be made as soon as possible after the item or manuever under surveillance was completed. Accordingly, the staff agreed to make the rating immediately following the check-ride. Even though any rating may be essentially subjective, efforts were made to keep the ratings as objective as possible. The Coordinator of the School of Aviation Education and Flight Training took the final check-ride and made the ratings.

A description of the statistics used throughout the study will be found in the following chapter.

CHAPTER III

RESULTS

The first step in the treatment of the data was the examination of the distributions in terms of central tendency and variability. These measures, together with an assessment of the standard error of the mean of each variable, may be found in Table I. The formulas by which the means, standard errors of the means, and the standard deviations were computed may be found in Garrett (9, pp. 28, 55, 200), formulas 1 and 14.

The next procedure involved the computation of a matrix of inter-correlations which was composed of the13 predictor variables and the criterion measure. These results may be found in Table II. The correlation formula used is Garrett's formula 26 (9, p. 142). As may be expected, varying degrees of relationship are found, both positive and negative, between the several variables.

In this study, the five percent level of significance was taken as the minimum level of acceptable significance; under these experimental conditions it was necessary that a correlation coefficient reach the magnitude of .367 before it could be assumed that chance alone could not be the basis of the correlation more than five times out of 100 trials. Only one of the coefficients between a variable and the criterion reaches the level.

TABLE I

MEASURES OF CENTRAL TENDENCY AND VARIABILITY OBTAINED FROM THE PREDICTOR AND CRITERION VARIABLES

List of Tests (N - 29)

C - Quality criterion
1 - American Council on Education
2 - Clerical Speed and Accuracy
3 - Mechanical Comprehension Test
4 - Pursuit Rotor
5 - G — General activity
6 - R — Restraint
7 - A — Social boldness
8 - S — Social interest
9 - E — Emotional stability
10 - O — Objectivity
11 - F — Friendliness
12 - T — Thoughtfulness
13 - P — Personal relations

(Tests 5 through 13 are subtests of the Guilford-Zimmerman)

Test	Mean	S. E. of Mean	S. D.
C	62.45	±2.918	15.44
1	89.03	±3.392	17.95
2	60.45	±2.177	11.52
3	37.83	±1.355	7.17
4	382.69	±8.666	45.86
5	18.55	±1.086	5.76
6	17.52	± .775	4.10
7	17.66	± .998	5.28
8	19.76	±1.124	5.95
9	18.62	± .884	4.68
10	19.34	±.843	4.46
11	15.86	± .845	4.47
12	20.24	±.724	3.83
13	18.72	± .799	4.23

TABLE II

INTERCORRELATIONS OF THIRTEEN TESTS AND A CRITERION

List of Tests (N - 29)

C - Quality Criterion
1 - American Council on Education
2 - Clerical Speed and Accuracy
3 - Mechanical Comprehension Test
4 - Pursuit Rotor
5 - G — General activity
6 - R — Restraint
7 - A — Social boldness
8 - S — Social interest
9 - E — Emotional stability
10 - O — Objectivity
11 - F — Friendliness
12 - T — Thoughtfulness
13 - P — Personal relations

	1	2	3	4	5	6	7	8	9	10	11	12	13
C	.38*	.22	.13	.18	.16	-.25	.20	.17	.11	-.06	-.32	-.06	-.29
1		.05	.22	-.01	.25	-.24	.09	-.02	.52*	.45*	.09	-.17	.10
2			.20	-.21	.34	-.15	.52*	.48*	.35	.15	-.24	-.09	-.07
3				.04	-.004	.21	-.18	-.03	.24	.16	.18	-.01	.10
4					-.02	-.20	-.30	-.32	.03	-.17	-.30	-.36	-.14
5						-.008	.29	.48*	.33	-.02	-.08	-.10	.16
6							-.18	.19	-.06	-.12	.44*	-.05	.41*
7								.65*	.34	.32	-.34	.15	-.37*
8									.20	.10	-.13	.18	.11
9										.63*	.009	-.24	.21
10											.28	-.12	.42*
11												-.16	.60*
12													-.22

* — Significant at the 5% level of significance.

Two regression equations and resultant multiple correlation coefficients were computed from the zero order correlation coefficients. In the first instance, using the complete test battery, this involved solving a set of 13 simultaneous equations in order to arrive at the regression weights carried by each of the 13 predictors. This solution gives the weighting that each of the tests carries in the equation expressing the predictive capacity of the battery. Garrett (9, p. 393) calls these values "partial regression coefficients," "Beta coefficients," or 'beta weights.' A bit of confusion arises over the use of the term "b" and the use of the term "Beta." Both terms are used to designate the coefficients of the variables in the equation for the multiple correlation coefficient; however, the "b" is used in the raw score form of the equation and the "Beta" is used in the standard score form. The two forms of this equation may be found in Garrett (9, p. 391), formulas 98 and 99. The b's and Betas for the 13 predictor battery may be found in Table III. The general regression equation, deviation form for "n" variables is: $Y_o = B_1x_1 + B_2x_2 + B_3x_3 + \ldots + B_nx_n$. The general regression equation, raw score form for "n" variables is: $Y_o = b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n$. The general equation for computing the multiple R^2 in terms of Beta coefficients and zero order r's is: $R^2_{y_o x_1 x_2 \ldots x_n} = B_1r_{o1} + B_2r_{o2} + B_3r_{o3} + \ldots + B_nr_{on}$. The multiple correlation coefficient computed in this manner was .652. The standard error of Y_o was 11.71 (the standard error of a rating predicted from this battery). Although the multiple correlation coefficient appeared to be rather high, it proved to be not significant at the five percent level of significance.

TABLE III

BETA WEIGHTS AND b WEIGHTS FOR THIRTEEN-VARIABLE REGRESSION EQUATIONS

Test	Beta Weight	b Weight
1	.544	.478
2	.070	.094
3	.144	.310
4	.219	.074
5	-.110	-.295
6	-.069	-.259
7	.183	.534
8	.258	.668
9	-.161	-.531
10	-.227	-.751
11	-.050	-.172
12	-.085	-.342
13	-.166	-.606

TABLE IV

BETA WEIGHTS AND b WEIGHTS FOR THREE-VARIABLE REGRESSION EQUATIONS

Test	Beta Weight	b Weight
1	.370	.318
2	.196	.263
3	.009	.020

It was thought that an abbreviated test battery might prove useful, provided that it proved to be significant, since the time of the staff and students is important. Accordingly the matrix of intercorrelations was examined in search of three variables whose correlation with the criterion was as high as possible and whose intercorrelations with each other was low. Weight was also given to the amount of time involved in administration and scoring of the tests. The American Council on Education, Mechanical Comprehension Test, and Clerical Speed and Accuracy tests were chosen for this abbreviated battery. Since the American Council on Education test scores will normally be on file for all students, it would only be necessary for the aviation school students to be tested on the Mechanical Comprehension Test and the Clerical Speed and Accuracy. The total time of administration of these two tests would be 45 minutes or less; less than half of the total time required if the Guilford-Zimmerman Temperament Survey were also administered, giving the other 9 variables to make up the 13 found in the first battery.

The same statistical procedure was followed using the three variables that was followed using the 13 variables of the first battery, except that only the correlation coefficients of the three selected tests were used in computing the multiple correlation coefficient and other values. The b's and Betas for the three predictor battery may be found in Table IV. The same general equations used with the 13 predictor battery apply here as well. The multiple correlation coefficient computed for this battery was .423; it proved not to be significant at the five percent level of significance. The standard error of a rating predicted from this battery was 13.99.

CHAPTER IV

DISCUSSION

The results have been presented; no zero-order correlation coefficients between the tests and the criterion were found to be significant at the five percent level of significance except that of the American Council on Education. This was despite a multiple correlation coefficient of a magnitude verbally described by Garrett (9, p. 173) as "marked" it was found that it too was lacking in significance. There were two questions that could be asked here; how could the multiple correlation be significant if the zero-order correlation coefficients were not, and why is it that the multiple correlation coefficient was not significant, since it was large. Garrett (9, p. 399) justified the use of a non-significant test in a battery by saying,

> A test may also add to the validity of a battery by acting as a "suppressor" variable. Suppose that Test A correlates .50 with a criterion—has good validity—while Test B correlates only .10 with the criterion but .60 with Test A. The $R_1(23)$ .56 despite the low validity of Test B. This is because Test B acts as a suppressor—takes out some of Test A's "non-valid" variance, thus raising the criterion correlation of the battery.

In a study of this type a basic assumption is made. It is assumed that the criterion is characterized by a factor structure that is measurable. It is further assumed that the predictors bear some relationship to the factor structure of the criterion. Assuming perfect reliability, the predictors measure all of the factors of the criterion,

then the relationship is expressed by a correlation coefficient of a ±1.00. If none of the factors are measured then the expected relationship is expressed by a correlation coefficient of zero. The low intercorrelations of the variables indicates that the predictor variables were relatively independent of each other. However, even though several of the zero-order correlation coefficients were of size and the predictor variables appeared to be relatively independent, it was found that the multiple correlation coefficient was not significant at the five percent level of significance.

This finding brought into question the assumption that the criterion was characterized by a structure of measurable factors. This assumption was thought to be valid. However, it was probable that the smallness of the number of cases (29) would not permit the most sensitive of predictors to demonstrate the relation that might exist. It was not forseen at the inception of the study that the attrition would necessitate a sample of this size; the attrition rate in this particular context is not highly predictable.

The level of significance necessary in order to claim acceptability is governed primarily by the size of the sample (N) and the number of variables (m) in the battery (9, p. 397). In this study, the non-significance of the multiple correlation coefficient may have been due to the fact that entering the table of correlation significance with (N - m) degrees of freedom proved an impossible burden for the battery to bear considering the large number of predictors and the small number of cases.

As regards the sample, 62 persons were enrolled in the three sections of ground school; 45 of these persons were induced to take the entire battery of tests; but only 29 of the total group had finished the flight aspects of the training over which they had been rated by the time the Spring semester closed, a necessary condition for inclusion in the experimental group. There may have been enough evidence present here to warrant the feeling on the part of the author that if the number of cases had been larger, the trend would have been extended into significance. As the situation actually existed, however, it must be reported that the multiple correlation coefficients were not significant at the five percent level of significance. The author would like to see a follow-up study of perhaps one year's duration, thereby gaining a much larger group of subjects.

Even though neither the abbreviated battery not the full battery can be accepted as valid, the author feels that they should not be rejected as being of no value. Instead, more work should be done in order to establish their degrees of usefulness.

CHAPTER V

SUMMARY

This study represents an attempt to predict pilot proficiency from a battery of common psychological tests. The work was done in, and with the cooperation of, the School of Aviation Education and Flight Training at Oklahoma State University. The need for this type of information grows ever greater in this era of rapidly expanding enrollment.

The literature was examined in order to gain insight into the problem and to see what answers were posited to the questions at hand. It was found that psychological tests have been used by both the military and civilian investigators to predict pilot performance. Selected findings of these investigations were presented. The tests found to predict pilot proficiency fell into the general groups of intelligence, alertness, proficiency, personality, interest, and visual-motor coordination. The tests were most often validated against pass-fail criterions.

The students in three Aviation Education Ground School courses were given a battery of five psychological tests. These tests were the American Council on Education Psychological Examination for College Freshmen, the Test of Mechanical Comprehension, the Clerical Speed and Accuracy portion of the Differential Aptitude Tests battery, the Rotary Pursuit test, and the Guilford-Zimmerman Temperament Survey. These

tests were thought to be adequate measures of the trait areas listed above. The scores on these tests were correlated with performance ratings given in the flight school to those students participating in both ground school and flight school.

Only the American Council on Education Test had a zero-order correlation coefficient that proved to be significant at the five percent level of significance. Multiple regression equations and multiple correlation coefficients were computed. The first multiple regression equation was for the full thirteen-test battery (the Guilford-Zimmerman Temperament Survey was broken into the sub-tests); the multiple correlation coefficient for this battery was of the order of .65 but it was not significant at the five percent level of significance. The second battery was an abbreviated one utilizing the American Council on Education test, the Test of Mechanical Comprehension, and the Clerical Speed and Accuracy. The multiple correlation coefficient for this battery was found to be not significant at the acceptable level of significance.

Since no significance was found, then the conclusion must be reached that the .65 multiple correlation coefficient of the full battery could have arisen by chance. However, the author feels that the trend for this battery to be an adequate measure of the criterion is strong and that additional validation work is well justified.

A SELECTED BIBLIOGRAPHY

1. Bennett, G. K. A Manual: Mechanical Comprehension Test, Form BB. (Rev. ed.) New York: The Psychological Corporation, 1951.

2. Bennett, G. K., Seashore, H. G., and Wesman, A. G. A Manual for the Differential Aptitude Tests. (2nd ed.) New York: The Psychological Corporation, 1952.

3. Berdie, R. F. [a review of] The Differential Aptitude Test Battery. In O. K. Buros (Ed.) The Fourth Mental Measurement Yearbook. Highland Park, N. J.: The Gryphon Press, 1953. pp. 679-80.

4. Davis, F. B. (Ed.) The AAF Qualifying Examination. Aviat. Psychol. Res. Rep. No. 6. Washington: U. S. Government Printing Office, 1947.

5. Du Bois, P. H. (Ed.) The Classification Program. Aviat. Psychol. Res. Rep. No. 2. Washington: U. S. Government Printing Office, 1947.

6. Flanagan, J. C. The selection and classification program for aviation cadets (aircrew—bombadiers, pilots, and navigators). J. consult. Psychol., 1942, 6, 229-239.

7. Flanagan, J. C. (Ed.) The aviation psychology program in the Army Air Forces. Aviat. Psychol. Res. Rep. No. 1. Washington: U. S. Government Printing Office, 1947.

8. Fiske, D. W. Validation of Naval Aviation Cadet Selection Tests against criteria. J. appl. Psychol., 1947, 31, 601-614.

9. Garrett, H. E. Statistics in Psychology and Education. (4th ed.) New York: Longmans, Green and Co., 1954.

10. Gordon, T. The airline pilot's job. J. appl. Psychol., 1949, 33, 122-131.

11. Guilford, J. P., and Zimmerman, W. C. A Manual: Guilford-Zimmerman Temperament Survey. Beverly Hills, California: The Sheridan Supply Co., 1949.

12. Jenkins, J. G. Selection and training of aircraft pilots. J. consult. Psychol., 1941, 5, 228-234.

13. Koerth, W. A pursuit apparatus; eye-hand coordination. Psychol. Monogr., 1922, 31, 288-292.

14. Lane, G. G. The prediction of success in learning to fly light aircraft. Psychol. Monogr., 1947, 61 (5), 1-18.

15. Liljencrantz, E. Problems in the selection of aviators. J. Aviat. Med., 1942, 13, 107-120.

16. Marquis, D. G. (Ed.) Psychology and the war; the aviation psychology program of the Army Air Forces. Psychol. Bull., 1943, 40, 759-769.

17. Marquis, D. G. (Ed.) Psychology and the war; the research program on psychomotor tests in the Army Air Forces. Psychol. Bull., 1944, 41, 307-322.

18. Matheny, W. G. The effectiveness of aptitude testing in the civilian pilot training program. Trans. Kans. Acad. Sci., 1941, 44, 363-365.

19. Melton, A. W. The selection of pilots by means of psychomotor tests. J. Aviat. Med., 1944, 15, 116-123.

20. Melton, A. W. (Ed.) Apparatus Tests. Aviat. Psychol. Res. Rep. No. 4. Washington: U. S. Government Printing Office, 1947.

21. Michael, W. B. Factor analyses of tests and criteria; a comparative study of two AAF pilot populations. Psychol. Monogr., 1949, 63 (3), 1-55.

22. Miles, W. R. Psychological aspects of military aviation. Amer. Scientist, 1945, 33, 146-158.

23. Payne, R. B., Rohles, R. H., and Cobb, B. B. The pilot candidate selection research program. USAF Sch. Aviat. Med., 1952, Proj. No. 21-29-008, Rep. No. 4.

24. Razrah, G. H. S., and Brown, H. C. Aviation. Psychol. Bull., 1941, 38, 322-330.

25. Staff, Psychological Research Project (Pilot). Psychological research on pilot training in the AAF. Amer. Psychologist, 1949, 1, 7-16.

26. Stuit, D. B. Personnel research and Test Development in the Bureau of Naval Personnel. Princeton: Princeton University Press, 1947.

27. Super, D. E. Appraising Vocational Fitness; By Means of Psychological Tests. New York: Harpers, 1949.

28. Survey of collegiate courses in aviation and related fields. Washington: U. S. Government Printing Office, 1944.

29. Van Steenberg, N. [a review of] The Guilford-Zimmerman Temperament Survey. In O. K. Buros (Ed.) The Fourth Mental Measurement Yearbook. Highland Park, N. J.: The Gryphon Press, 1953. pp 96-97.

30. Viteles, M. S. The aircraft pilot; five years of research. Psychol. Bull., 1945, 42, 489-526.

31. Woodworth, R. S. and Schlosberg, H. Experimental Psychology. (Rev. ed.) New York: Henry Holt, 1954.

APPENDIX A

Instructions for the Rotary Pursuit Test

"This is a test of your ability to follow a moving target. Your task will be to keep the point of this stylus on the round brass target while it is moving. Hold the handle like this and keep the stylus on the target as it goes around (demonstrating). You will do best if you develop a smooth, free-swinging motion of the arm and shoulder."

APPENDIX B

Student Rating Scale

	Excellent 10	9	8	7	Average 6	5	4	3	2	Poor 1
Taxiing, Parking										
Take Off, Landing										
Straight & Level										
Turns										
Climbs, Glides										
Turns Around										
Coordination										
Emergencies										
Desired Track										
Planning, Judgment										

Student _____

Total Hours Dual_____ Solo_____ Total_____

Flight Check Pass_____ Fail_____

Recheck Pass_____ Fail_____

Rated by _____

VITA

Steve Allan Heckart

Candidate for the Degree of

Master of Science

Thesis: THE DETERMINATION OF THE VALUE OF A BATTERY OF PSYCHOLOGICAL TESTS IN PREDICTING PILOT PROFICIENCY

Major Field: Psychology

Biographical:

Personal data: Born in Vernon, Texas, November 8, 1930, the son of Arlie and Clara Rose Heckart.

Education: Attended grade school in Canton, Kansas; Ida, Louisiana; and Fairfax, Oklahoma; graduated from Fairfax High School in 1948; received the Bachelor of Science degree from the Oklahoma Agricultural and Mechanical College, with a major in Psychology, in May, 1956; completed the requirements for the Master of Science in August, 1957.

Experiences: Served in the United States Army during 1950-52. Acted as a graduate assistant in the College Reading Clinic during the academic year 1956-57.

Member of: Society of Personnel Management; Psychology Club; Psi Chi; and the Oklahoma State Psychological Association.