VOLUME 63
NUMBER 3

WHOLE NO. 298
1949

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Factor Analyses of Tests and Criteria:
A Comparative Study of Two AAF Pilot Populations

By

WILLIAM BURTON MICHAEL

A Report Based on a Dissertation Presented to the Faculty of the Graduate School
of the University of Southern California

Accepted for publication, August 5, 1948

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC
Publications Office
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

ACKNOWLEDGMENTS

THE WRITER wishes to express his appreciation particularly to Professor J. P. Guilford, Chairman of the Doctoral Committee, for his assistance and for his constructive criticisms in the preparation of the dissertation upon which the report is based; to the other members of the committee for many helpful suggestions: Professors D. W. LeFever, M. F. Metfessel, A. S. Raubenheimer, F. L. Ruch, and L. P. Thorpe; and to Dr. W. S. Zimmerman, not only for his personal communication of his hypotheses concerning the nature of visualization, but also for his aid in the rotation techniques employed in the factor analyses. The writer also wishes to express his gratitude to Miss Ruth Bohnett, Assistant to the Dean of the Graduate School of the University of Southern California, for her encouragement in the period preceding and accompanying the preparation of the dissertation.

WILLIAM B. MICHAEL

Princeton, January 21, 1948

TABLE OF CONTENTS

CHAPTER I

THE PROBLEM AND ITS BACKGROUND

A. Introduction

1. *Historical Background*

During World War II, psychologists put forth an immense amount of effort to improve techniques of selection and placement of personnel in the military services and in essential civilian activities. Among the most outstanding services were those of a group of psychologists who participated in the research activities of the United States Army Air Forces Aviation Psychology Program. With almost unlimited numbers of subjects, they were able to undertake many experimental studies and to evaluate the effectiveness of many tools of research which had been largely developed in the non-emergency atmosphere of college and university campuses. Statistical tools which had been of primarily academic interest were brought to bear upon the solution of numerous practical problems.

One such statistical device is factor analysis. Probably more than any other tool, factor analysis has served to expand the horizon of psychological testing in the last five years. In fact, many previous ideas concerning the requirements of a good test have been seriously challenged, and new standards of test evaluation have had to be proposed (10).

At the time of cessation of hostilities, considerable research remained to be completed upon projects already in progress—to say nothing of that research which had been reserved for future projects to be authorized. It is from such uncompleted research that the data of this study originate. In addition to the factor analyses to be presented in this investigation, numerous other research problems for which data are available remain that may throw additional light upon the persistent problems encountered in psychological and educational measurement.

2. *Prerequisite Background for Understanding the Statement of the Problem*

Since there are several approaches to factor analysis, it should be mentioned that the Thurstone system with its appropriate terminology will be employed throughout this investigation (5). This system received major emphasis during World War II in the research units of the Aviation Psychology Program of the Army Air Forces. Its value in the construction of tests for selection and placement of Air Force personnel has been demonstrated again and again. The following advantages have justified the use of Thurstone's centroid approach in the studies of human abilities:

(1) The smallest number of psychologically meaningful and relatively independent factors necessary to describe a table of (test) intercorrelations may be ascertained—factors which are much easier to interpret than those derived by several other systems.

(2) The weights of each test with respect to the identified factors, or the amount of primary ability represented by a test or a criterion, is easily determined.

(3) Regression equations may be set up to predict, from the tests, the amount of an individual's ability in each of the independent (or nearly independent) factors.

(4) Correlations between tests in terms of the sum of cross-products of paired factor loadings are easily obtained, and original intercorrelations of tests are easily reproduced. Moreover, correlations between tests and a criterion are effected which heretofore have been possible only through more cumbersome techniques.

(5) The presence of correlation between factors is not excluded and may be readily demonstrated through use of oblique (non-orthogonal) axes.[1]

B. The Problem

1. *Statement of the Problem*

The purpose of the investigation was to ascertain the contributions of factors both to the description of tests and to the predictive values of these tests in two *pilot* populations of the United States Army Air Forces. Specifically, for two groups—815 West Point Cadets and 356 Negro cadets of significantly different standings in their mean composite scores and in their mean scores in tests identical to the two batteries—the problem may be clarified in terms of the following two major questions and subquestions:

(1) What differences were present in the factorial composition of two test batteries and of a pass-fail criterion?

(a) How many factors or abilities might be found within each of the two matrices of intercorrelation? What was the smallest number of factors by which each matrix might be described?

(b) What were the factors? In other words, how might the factors be interpreted? Were they meaningful? Were the factors the same for the two matrices?

(c) What were the weights of the factors or abilities in each test in its respective matrix?

(d) For both matrices, how complex factorially was each test? In other words, how many factors were involved in each test? How nearly pure was each test?

(e) Were the factors derived from each matrix independent or uncorrelated with one another?

(f) Were the factors derived from the two matrices and their corresponding weights essentially the same as those found in previous analyses (in which a larger number of tests have been used) for a white cadet-pilot population?[2]

(g) Were the factors and their weights for those tests which were common to the two matrices noticeably different?

(h) What new factors, not uncovered by previous analyses, appeared in either matrix?

(i) For the two groups, what differences appeared in the pass-fail criterion as to the number and the identification of factors, and as to their respective loadings?

(2) What differences were present between the two groups in the prediction of the pass-fail pilot criterion from the scores of the tests in the respective batteries?

(a) What was the validity of each test battery in terms of the coefficient of multiple correlation between the pass-fail criterion and the test scores of the battery?

(b) Approximately how much did each test of a battery contribute to the coefficient of multiple determination? For those tests which were identical in the two matrices, were there marked differences in the variance-contributions to the total predicted variance?

(c) How closely might the validity of each test battery be estimated from the total common-factor variance of the

[1] For the reader who is interested in additional background material concerning the research activities of psychologists in the Army Air Forces, three publications (listed in the Bibliography as 21, 22, 23—especially 23)—are helpful.

[2] The results of several previous factor analyses have been combined in a report (24). In this report are the "weighted averages" of the factor loadings, communalities, reliabilities, and validities of tests from fifteen batteries upon which analyses had been performed. The document henceforth will be called *Composite Factor Analysis Summary*.

pass-fail criterion? In other words, how closely did the sum of the common-factor variance in the pass-fail criterion approximate the coefficient of multiple determination corresponding to each battery?

(d) What were the estimated validities (correlation coefficients) of each test with the pass-fail criterion in terms of the sum of the cross products of the paired factor loadings in the test and criterion?

(e) Did any tests used in either group contribute a marked degree of unique variance which materially influenced the validity of the battery?

3. *Importance of the Problem*

If, with two such diverse populations[3] as those represented in this investigation, the same factors with their respective weights approximately equal should appear in the analyses to be highly valid in the prediction of a pilot pass-fail criterion, then an inference could be made that the same tests or test batteries might be used in the selection of pilots despite marked differences in the characteristics of populations. Such a finding may lead one reasonably to believe that a relatively small number of empirically determined factors in certain amounts is essential to the success of pilot trainees of heterogeneous racial, intellectual, educational, and socio-economic backgrounds. Moreover, if the results of the factor analysis of the batteries administered to the two groups should approximate those of several previous analyses (reported in (24)), then the inference suggested could be made on empirical grounds with a still higher degree of confidence. In short, a plausible inference could then arise that individuals of diverse backgrounds would have to possess certain important psychological factors in sufficient amounts to be successful pilots. Such a conclusion, if valid, would permit the use of the same tests in predicting a criterion for numerous populations and would minimize considerably the work of the psychologist.

On the other hand, if the empirically determined factors in the two groups should differ considerably with respect to their types, their number, and their loadings on both the tests and the criteria, then obviously two test batteries would be necessary. Although such an outcome would limit the range of applicability of a test battery, the psychologist interested in achieving maximum validity for the batteries corresponding to the two populations would be in a position through his knowledge of the factorial composition of each criterion and of the tests to effect many improvements.

Such possibilities for improvement, of course, would be revealed whether the factor analysis demonstrated similarities or differences in the factorial composition of the two batteries. When a criterion is included with the tests of a battery, several important advantages accrue in the factor approach. With a clear-cut and quantitative picture of the factorial structure of both the tests and the criteria, the test technician has achieved a degree of satistical control over a test battery which enables him to take steps to maximize its validity through retention of certain tests—especially pure tests containing relevant variance—revision of others, elimination of some containing large amounts of irrelevant variance, and addition of tests

[3] In Chapter II the populations will be defined in operational terms with respect to their standings on a scale resembling that of standard scores (the pilot stanine).

contributing new variance held in common by the criterion. Hence, some important outgrowths of this study may be several practical suggestions for further experimentation with the tests so as to increase their validity.

C. Definitions of Terms

1. *Factor*

Nearly synonymous with the interpretation of ability and trait, the meaning of a factor may be considered a convenient theoretical construct or intervening variable which is functionally related both to antecedent *S*-variables and to consequent *R*-variables.[4] The *S*-variables refer to the test-items which sample, in a more or less controlled manner, the predictive abilities of the examinee. The *S*-variables, frequently referred to as "independent" variables, are subject to a considerable degree of manipulation or modification (as in two forms of a test designed to measure the same function, or in two tests purporting to measure different functions). The *R*-variables are the "dependent" variables. Schematically, the relationship involved may be pictured as S→ I→ R. Upon the assumption that a workable scale of measurement with an arbitrary zero and approximately equal units can be effected, scores for individuals may be determined and classified. The application of factor analysis to correlations derived from these measures serves merely to simplify the relationships among the quantified response (dependent) variables in terms of fewer, somewhat more stable, unit patterns (factors) which are functionally related to the stimulus (independent) variables. Any interpretation, however, of what these relatively independent unit patterns signify is a theoretical construct, intervening variable, or inference, capable of being related to both the antecedent and the consequent conditions described. In short, factors are considered to be functional unities of reasonable stability, describing in a simplified manner the fundamental psychological operations of the individual as he reacts to a variety of different tests, problems, or situations subject to a greater or lesser degree of experimental control.

Considerable controversy persists as to the value of introspective reports of subjective experiences in the *interpretation* of a theoretical construct. (12). Reports based upon introspection of several examinees might supply information concerning what influence the test exerts upon their actions. Since the factor analyst himself resorts to considerable introspection in his attempt to give a meaningful interpretation of a factor, the combination of reports of both the technician and examinee might yield a somewhat more universally acceptable understanding of the factor. However, as most operationists will maintain, an intervening variable to be useful need not be defined in terms other than those which relate it functionally to the operations involved in the specified situation. Labelling a construct verbally does, however, allow a more convenient mode of reference to it.

[4] The operational definition of factor follows closely the technique employed by Hull and Spence in their definitions of other constructs such as drive. For a detailed consideration of the role of intervening variables in psychological theory, see (4) and (14).

2. *Validity*

In terms of the factorial composition of a test and criterion, a relatively new approach to a validity coefficient is possible. Just as the correlation between two tests may be expressed as the sum of the cross-products of the loadings of paired factors, so may the correlation between a test and criterion be estimated. Let $k_{1t}, k_{2t}, \ldots, k_{it}, \ldots k_{nt}$, represent the factor loadings of a test, and let $k_{1c}, k_{2c}, \ldots, k_{ic}, \ldots, k_{nc}$ stand for the factor loadings of a criterion. The correlation between the test and criterion is determined as follows:

$$r_{tc} = k_{1t}\,k_{1c} + k_{2t}\,k_{2c} + \ldots + k_{it}\,k_{ic} + \ldots + k_{nt}\,k_{nc}$$

Since the reliability of a test (or criterion) is that variance which equals the sum of the common factor variance and specific variance of a test (or criterion), there is a tendency for the correlation between a test and criterion to be underestimated if the specific variance in either or both is high (upon the additional assumption that the error variances are uncorrelated with each other and with the non-error variance). If additional tests can be brought into a battery to aid in the identification of a new factor common to a given test and criterion (with a resulting drop in specific variances), then the validity coefficient can be considerably raised (upon the assumption that the loadings upon the other factors in the test and criterion do not decrease and that the loadings on the new factor in the test and criterion are of like sign). With a sufficient number of tests in a battery for identification of factors, the validity coefficients obtained closely approximate those ascertained from the use of traditional methods (23, pp. 797-850).

3. *Stanine*

A term coined and frequently used in the AAF psychological research units during the war was *stanine*. A distribution of a large number of representative composite scores from a given test battery may be scaled in terms of standard scores or according to a T-scale or C-scale technique. The stanine is an adaptation of Guilford's eleven point C-scale system (2). A stanine is a Guilford C-scale in which the 0 and 1 scaled scores and the 9 and 10 scaled scores are lumped together. Hence the *stanine* is a C-scale of nine scaled score intervals instead of the customary eleven. Therefore the percentages of scores (to the closest whole numbers) are 4, 7, 12, 17, 20, 17, 12, 7 and 4. With respect to a standard representative group, an examinee's score may be easily converted from a centile score or standard score to a stanine.

For each new classification battery, stanines were established for the scores of bombardiers, navigators, pilots, and in some instances, for the scores of other groups. The norms for these stanines were based upon the scores of several representative groups of cadets taking the classification battery at various training centers throughout the United States. The scores of all groups were converted to scale values determined from these standardizing groups. Cut-off points were established for elimination of cadets, frequently as a result of determination of the proportion of trainees at each stanine level eliminated. The basis of elimination was the *pass-fail* criterion, the next term to be defined.

4. *Pass-fail Criterion*

Whether a cadet should pass or fail frequently rested upon highly subjective grounds. The criterion was subject to considerable variation from one training unit to another. A given cadet was passed or failed on the basis of a rather unstable "weighting" of grades earned in subjects studied in ground school and of ratings reported by flight instructors and qualified judges. The decisions of boards of military personnel were based upon those grades and ratings as well as upon their impressions of the cadet gained through personal interview and general military record. In many instances periodic physical examination served to eliminate a small proportion of cadets from training. Trainees expressing dissatisfaction or fear were eliminated.

Despite the inherent weaknesses of such a criterion, it was used as a basis for validation of tests. Biserial coefficients of correlation were computed between the dichotomous pass-fail criterion (assumed to be a continuous variable normally distributed) and the more-or-less symmetrically distributed scores on a given test.

In addition to these four important terms, others might also be defined. A more functional approach appears to be that of elucidation of concepts or terms when necessary at appropriate places throughout the text. However, the four terms defined do represent a necessary and workable nucleus.

D. Review of Related Literature

Relatively few studies have been made concerning the effect of two or more populations upon the factorial composition of a test battery. A few studies concerning the effect of groups differing with respect to age (7, 8, 13) or sex (19) upon factorial composition of tests have been performed. However, these studies are not particularly applicable to the problem of this investigation, since the two male populations of Negroes and West Point Cadets are representative of nearly the same age range.

In three recent articles of a highly theoretical nature, Thomson (15), Thomson and Ledermann (16), and Thurstone (18) have investigated the influence of univariate selection and multivariate selection in the factor analyses of tests of abilities. Thurstone has used the formulas developed by Thomson, as well as his definitions of terms, and has expanded somewhat upon the work of these British investigators. The fundamental problem proposed by these three psychologists is to determine the alteration in the factor pattern of test scores of an initial group when a second comparable group (the experimental group) is selected on the basis of their homogeneity of performance on one test (or several tests) of the battery—called a selection variable or a criterion test.

Thurstone is able to show that the inclusion of one or more additional tests which are linear combinations of those tests initially present in the battery results in the addition of one or more factors that are incidental. If simple structure is apparent for the initial test battery, the addition of new tests does not alter the structure except in the instance of a large number of incidental factors, the variance of which may mask that of the common factors.

However, Thurstone's findings are at best only indirectly applicable to the two populations of West Point and of Negro pilot cadets, in that no selection test, such as the AAF Qualifying Examination, was included in the test batteries. (The cadets were selected on the basis of their total, or composite, scores on the classification battery. There was, of course, some degree of selection in pilot training as in any other occupation.) Nor can the pass-fail criterion be employed, since it was not used for selection of examinees, but followed by several months the administration of the tests. Moreover, to explore the problem as set up by Thurstone would require for best results two control populations in addition to the two experimental populations.

In short, no investigations reported in the literature seem to bear directly upon the problem of the study. At appropriate places in the following chapters, however, references will be made to pertinent investigations which may furnish needed clarifications of the text. For general background to the problem the reader is referred to the previously mentioned publications of the AAF Aviation Psychology Program (21, 22, 23).

CHAPTER II

SOURCES OF DATA

In the following three sections an attempt will be made: (1) to define in operational terms the two previously mentioned populations of West Point Cadets and Negroes selected for investigation of the problem of the dissertation; (2) to present a rationale for the inclusion of tests in the two batteries (the intercorrelations of which will be subjected subsequently to factor analysis); and (3) to describe the tests with respect to their purpose, content, scoring formulae, and time-limits.

A. An Operational Definition of the Two Populations of Negroes and West Point Cadets

In a previous paragraph (I, B, 3), reference was made to the fact that the two groups of West Point Cadets and Negroes were diverse. Although from a qualitative standpoint differences in the two groups with respect to such characteristics as race, amount of previous education, socio-economic status, and so forth, may be described, an operational approach in terms of the difference in their standings in the pilot stanine and in individual tests common to the two batteries is both meaningful and precise.

One serious difficulty arose in the comparison of the two groups with respect to their means and standard deviations on the pilot stanine, in that the composite scores of each group were based on two different test batteries. Two reasons, however, justified the computation of a t-ratio as a test of the significance of a difference between means (actually a test of the hypothesis that the two samples were drawn from the same population). First, the obtained mean stanine of each group stood in reference to a mean stanine of 5 (and approximate standard deviation of 1.96) which was derived for two representative white aviation-cadet populations. Each of these two populations was a composite of samples obtained at various training centers throughout the United States. Second, the mean stanines of two representative white cadet populations upon the November, 1943, and September, 1944, battery were regarded as relatively stable (prior to the elimination of any cadets below the stanine cut-off point). In other words, a negligible difference between the mean stanines of two representative white-cadet populations should be expected from one year to the next on two comparable test batteries.

1. *Difference Between the Negro Population and the West Point Population Statistically Defined*

Upon the assumption that a stanine of 5 on the November, 1943, battery was the same as a stanine value of 5 on the September, 1944, battery, t-ratios were computed to obtain the degree of significance of the difference between the obtained stanine means of the two groups. Since a distinction between fighter pilot and bomber pilot was employed at the time of administration of the September, 1944, battery, two t-ratios were computed to test the significance of differences (1) between the mean pilot stanine of the Negroes and the mean fighter pilot stanine of the West Point Cadets and (2) between the mean pilot stanine of the Negroes and the mean bomber pilot stanine of the West Point Cadets. The means and standard deviations for the 815 West Point men and 356 Negroes are presented in Table 1.

The *t*-ratio for the difference between the mean of the Negro pilot stanine and the mean of the West Point fighter pilot stanine was significant beyond the one percent level ($t = 21.3$ for 1169 degrees of freedom). For the difference between the same mean for the Negroes and the bomber pilot stanine mean the degree of significance was still higher ($t = 30.8$ for 1169 degrees of freedom). Obviously, it was reasonable to assume that the two populations were not homogeneous.

TABLE 1

STANINE MEANS AND STANDARD DEVIATIONS OF NEGRO CADETS AND WEST POINT CADETS

Group	Number	Mean	Standard Deviation
I. Negro Cadets	356	3.95	1.61
II. West Point Cadets	815		
A. Fighter Pilots	(815)	6.21	1.80
B. Bomber Pilots	(815)	7.09	1.59

In eight other comparisons between the mean stanine scores of these two groups and the new stanine scores of eight representative white-cadet samples on respective test batteries, statistical significance was revealed beyond the one percent level. (The data for these comparisons were taken from Tables 3.17 and 3.20 in (21) and from (25)).

2. *Differences Between Mean Scores in Individual Tests Administered to Negroes and West Point Cadets*

Since there were only thirteen tests common to the November, 1943, and September, 1944, batteries, the number of comparisons of scores was limited. Moreover, since the raw scores of the Negroes upon the six psychomotor tests (the only scores available) did not correspond with the standard scores of the West Point group, the number of tests for which comparisons of means might be effected was limited to seven (all pencil-and-paper tests). All seven differences between means for the two groups were statistically significant beyond the one percent level.

B. RATIONALE UNDERLYING SELECTION OF TESTS FOR NOVEMBER, 1943, AND SEPTEMBER, 1944, CLASSIFICATION BATTERIES

The choice of tests to be used in classification batteries such as those of November, 1943, and September, 1944, depended largely upon the traditional multiple-regression-equation principles. As a result of this statistical work, many tests were retained; some, modified; and a few new ones, added. Since a test battery had to serve a threefold purpose of selecting pilots, navigators, and bombardiers, several compromises had to be made in choice of tests for the batteries. Some tests, for example, contributed substantially to the validity of the battery for one of the three groups mentioned, but negligibly for the two others.

During the course of the war, the rationale underlying selection of tests rested more and more upon the factorial variance which these tests held in common with that of the criterion. Wherever possible, pure tests (containing variance in the relevant factor) were preferred to complex tests (containing substantial amounts of variance in two or more relevant factors). Several exceptions to this goal of purity were apparent—especially in the use of psychomotor tests. Complex tests were retained when their absence would result in a marked reduction in the overall validity of the battery. As the factorial composition of the pass-fail criteria became better known, tests were constructed to cover the identified areas of the criteria. The addition of new tests frequently revealed new factors, and subsequently other tests were designed to describe better these factors and to augment the validity of the battery. Consequently, it became convenient to think of test areas and of families of tests covering these areas of variance.

Although as many as twenty-seven factors were revealed (at least tentatively), the major portion of the validity variance (for pilot, navigator, and bombardier groups) had been delimited to six or eight factors on the basis of previous analyses. For convenience, tests might be grouped into the following areas: mathematical, verbal, reasoning, judgment, general and mechanical information, biographical data, space and visualization, perceptual speed, and apparatus.[1]

C. Description of Tests

In order that the identification of the factors may be more meaningful a description of the tests in the two batteries follows: first, for the Classification Battery of November, 1943, administered to the population of Negroes; second, for the Classification Battery of September, 1944, administered to the population of West Point Cadets. As mentioned previously, the November battery consisted of eighteen tests, sixteen of which appeared in the September battery (with slight revisions in the composition of two and in the directions for one); in the September battery, five new tests appeared (making for a total of twenty-one tests). Hence, the following description covers systematically each test in the November battery and those eight tests in the September battery of which five were unique additions and three were modifications of tests in the November battery.

1. *Tests of the Classification Battery of November, 1943*

a. Pencil-and-paper tests. The test of *Reading Comprehension*, C1614H, planned to obtain a measure of verbal intelligence, consists of five paragraphs of technical material upon each of which the examinee must select the best answer in each of several multiple choice items of five alternatives. Paragraphs upon geography, mechanics (physics), the psychology of vision, the compass, and map projection theory represent the content. The total number of items is thirty-six. With a thirty-minute time limit, the test is scored: 2R—W/2.

Two tests of fifty items each of spatial orientation were employed. *Spatial Orientation Test I*, CP501B, was designed to test the ability of the examinee to identify parts of aerial photographs.

On the upper half of each page of the test booklet is a large rectangular aerial photograph, approximately six inches by eight inches of likely military objectives (such as industrial cities traversed by rivers, or harbor cities). Fifteen small circles of about one-fourth inch diameter are overprinted upon the aerial photograph in a random fashion. Each of these circles is lettered. On the lower half of each page are five or six larger circles (numbered) with diameters of approximately one and one-fourth inches, representing the areas to which certain ones of the fifteen small circles (lettered) on the rectangular photograph correspond. To each of the larger numbered circles (representing

[1] For further information concerning the rationale for inclusion of tests in the two batteries, see 23, pp. 851-886.

number of test item) the examinee is to choose a smaller lettered circle (response) which matches. As a rule, the position of the territory in the larger circle is enlarged or diminished considerably and rotated from its corresponding position in the rectangular photograph, because the small photograph was taken from altitudes and angles different from those of the rectangular photograph. Planned to be a five-minute test, it is scored: R—W/5.

Spatial Orientation Test II, CP503B, was designed to measure an examinee's ability to locate on a map corresponding places shown on an aerial photograph. On the upper half of each page the rectangular section of the map is approximately four inches by six inches. The relief maps (similar to those in Goode's *School Atlas*) employ various contour shadings of green and purple for ranges of elevation, yellow areas for large towns, circles for small towns, blue lines for rivers and streams, black lines crossed by short marks for railroads, purple lines for highways, and other miscellaneous symbols. Each map is sectioned off into twelve square areas labeled with capital letters (A to L). At the bottom of each page are four photographs, each enlarged ten times (bearing the number of the items) corresponding to a restricted portion of one of the square areas upon the map (correct alternative response). Both the photographs and the maps are oriented in the same manner geographically—north being the upper edge of the respective figures. In summary, the task is to find the area in the lettered block on the map corresponding to the enlarged photograph and to record the answer by blackening on the answer sheet the space containing the appropriate letter. For each double page, consisting of two maps and four respective photographs, the examinee is allowed three minutes. Since there are six double pages, the total testing time is 18 minutes. The scoring formula is: R—W/5.

The *Dial and Table Reading Test*, CP622-21A, represents a combination of tests which in earlier batteries were administered separately. In the dial reading test the purpose is to determine how quickly and accurately the examinee can read the dials on an instrument panel. Seven imitation dials are labeled R.P.M., Air Speed, Altitude, Voltmeter, Temperature (with Oil and Fuel subdials superimposed), Fuel-Air Ratio, and Amperes. Ten sets of simultaneous readings are presented throughout the test to each of which six items, one for each instrument, are used (except in the tenth set where only three items are used). Each item contains five alternative responses—readings of the respective dial—from which the examinee selects the correct answer. For the 57 items following the practice set of three, nine minutes are allowed.

The table reading test, designed to measure one's ability to read tables quickly and accurately, is divided into two parts. In Part I a large table is set up with numbers ranging from −17 to +17 across the top and up the left side. The numbers at the top and sides are respectively labeled *first value* and *second value*. In the same manner as one reads a road map to ascertain the distance between two localities, the examinee given a first value number and a second value number has to go over to the column corresponding to the first value number and down this column to the space where the row corresponding to the second value number intersects the column. At the space of intersection is either a two digit or three digit number representing the correct answer among four other incorrect alternatives in a multiple choice item, in which the first value and second value numbers stated together are the given stimuli. The testing time allowed for 42 items is eight minutes.

In Part II of the table reading test, four tables are presented. Across the top of each table is a given air speed of 100, 120, 140 and 180 miles per hour. For different wind angles written in a column at the left-hand side by steps of 10 degrees from 0 to 90, three wind velocities of 10, 15, and 20 miles per hour head three additional columns. Each of these three columns is further divided into two sets (sub-columns) of *direction correction* and *ground speed* (corresponding to the wind angle at the air speed under consideration). Given for each item the amount of air speed, of wind velocity, and of wind angle, the examinee is asked to determine from reading one of the four tables either the *direction correction* or the *ground speed*. There are five choices for each item. The total testing time is seven minutes. Hence the total time for the three sub-tests is 24 minutes. For all items upon the three sub-tests the scoring formula is: $\frac{R-W}{2}$

The *Biographical Data Test*, CE602D, is scored twice in each battery. The raw scores upon the test are differentially weighted according to whether it is used in the pilot or navigator context. An outgrowth of several previous forms, the Biographical Data test was designed to measure family and social relationships, and background experiences, which have been shown valid in predicting success of pilots. The sixty-five items with varying numbers of alternatives to which the examinee is directed to give the most accurate answer possible pertain to the following areas: the socio-economic status, racial origins, and educational experiences of the members of his family; his degree of success in various school subjects and in athletic participation; his favorite hobbies and recreational activities; previous work experiences; and reasons for choice of aviation cadet training. Timed at twenty-five minutes, the test is scored: R—W + 20 in which R is a positively weighted response and W, a negatively weighted response.

The test of *Mechanical Principles,* consisting of two comparable forms, CI903A and CI903B, in the respective batteries, was designed to measure the examinee's knowledge and understanding of mechanical principles. As in the *Bennett Test of Mechanical Comprehension,* diagrams of mechanical devices and familiar situations in which elementary principles of mechanics are involved are part of each item. These figures are labeled with letters and arrows (usually indicating a force vector or movement). Given certain relevant information concerning the problem represented by the diagram, the examinee responds to one of the alternatives, three, four, or five in number. Planned to take twenty minutes, the test of forty-one items is scored: R—W/2 + 20.

The *General Information Test,* CE505E, was intended to measure what its title suggests. The one hundred items consisting in each instance of four or five alternative words, phrases, or clauses completing the meaning of the fractional part of the given stimulus sentence are distributed as follows: Aviation Interest, 35 items; Sports and Hobbies, 13 items; Mechanical Information, 20 items; Driving Information, 12 items; Flying Information, 20 items. Designed to take thirty-six minutes, the test is scored: R—W/4.

Two tests in mathematics employed in the November, 1943, Classification Battery were designed to measure mathematical and reasoning ability. The first test, *Mathematics A,* CI702F, containing thirty-five multiple choice items, embodies elementary problems primarily in algebra with a few geometric and trigonometric concepts introduced. Emphasis upon the so-called "thought" problems is minimized in Test A. With a time limit of twenty-five minutes, this test is scored: R—W/4.

The second test, *Mathematics B,* CI206C, frequently designated as *Arithmetic Reasoning,* CI206C, was developed to measure primarily reasoning ability. The thirty multiple choice test items closely resemble the typical "word" or "thought" problems with which students in general mathematics and algebra courses in high schools seem to have difficulty. With a time limit of thirty-five minutes, this second test is scored: 2R—W/2.

Similarly, two tests in instrument comprehension were used in November, 1943, Classification Battery. The first, *Instrument Comprehension I,* CI615B, was designed to evaluate the examinee's ability to read and to interpret instrument dials. The instrument panel used in each item consists of a series of dials (altimeter, artificial horizon, compass, rate of climb, air speed, and turn-bank) shown on one page. Opposite each item are five choices that describe the actions of the plane in phrases consisting of verbal and numerical terms. With a time limit of twelve minutes, the test is scored: 20 — R + W/4. This unusual formula was used because during the short time this test was included in the classification battery, it had a negative regression weight. This formula permitted the use of a positive weight in machine combination procedures.

The second test, *Instrument Comprehension II,* CI616B, attempts to measure one's ability to interpret the position of a plane from the readings of two instruments: the compass and artificial horizon—instrument flying. As the reader knows, the compass shows the direction of the plane. The artificial horizon dial indicates whether the plane is climbing, flying level, or diving; banking left, or banking right. Combinations of one of the former three motions and of one of the latter two motions may be interpreted from this dial. Five pictures of planes in different positions (alternative responses) accompany each new set (item) of artificial horizon and compass readings. Limited to eighteen minutes, this test is scored: R—W/4.

b. Apparatus Tests. Among the six psychomotor tests employed, the most familiar to the psychologist is probably the *Rotary Pursuit,* CP410B. Designed to measure the degree of skill acquired in eye-hand coordination in which sweeping arm and hand movements are required, this test has been given under two conditions: without divided attention and with divided attention. In both the November and September batteries, the test with divided attention was used. In the simple rotary pursuit task (without divided attention) the examinee attempts to maintain his stylus on a small target on the disk, revolving at 60 revolutions per minute. Hence the arm moves in a free and gross circular manner. In the rotary pursuit task with divided attention, the distracting device consists of a second task. Inside a box are two lamps, connected to adjacent keys. During trials 6 to 15 of the simple rotary pursuit task, the lamps alternate in lighting in an irregular (but controlled) manner. In order that his pursuit score may be recorded, the examinee has to depress with his free hand the key corresponding to the lighted lamp. A total of fifteen trials, twenty seconds each, serves as the basis for the score: the total time of contact with the target in hundredths of a minute.

No penalties were used in scoring the six psychomotor tests. Simple conversion tables were employed in transforming all scores on psychomotor tests into standard scores. This eliminated test machine differences.

The *Complex Coordination Test,* CM701A, measures, as its name suggests, the degree of skill and speed with which an examinee can perform a series of complex reactions. A chair and stick are provided as well as a panel in front of the examinee containing three sets of parallel rows of potentially red and green lights. In any given trial, three red bulbs are illuminated. A lighted bulb on the top arc shows the lateral position of the stick required; one in the middle vertical row (perpendicular to the top arc and to the bottom horizontal row) signifies the required forward-backward placement of the stick; and the one along the bottom horizontal row denotes the required position of the rudder control. When the examinee moves the stick and rudder control, green bulbs illuminate in rows parallel to those in which the red lights appear. The assigned task is to match the three green lights to their corresponding red lights as quickly as possible. Whenever one problem is solved, a new combination of three red lights is shown. Following a two-minute practice period, the performance score is the number of sets matched in eight minutes.

The *Finger Dexterity Test,* CM116A, was designed to measure the degree of skill acquired in eye-finger coordination that requires rapid and precise movements of the fingers and wrist. In this test the examinee fits pegs into a peg board, turning the pegs to the right in the process. Square-shaped pegs about two inches long are painted blue on one vertical side and yellow on the opposite vertical side. At the start, either all blue or all yellow sides are toward the examinee. He is directed to lift each peg from its hole, to turn it clockwise 180° and to re-set it in the hole. Following one practice trial, five test trials of thirty-five seconds each are administered. The score is the total number of pegs turned by the right hand.

The *Discrimination Reaction Time Test,* CP611D, the origin of which obviously may be traced to the psychological laboratory, was planned to measure the speed of reaction determined by the amount of time required for the examineee to press one of four switches in response to four possible patterns of green and red lights on a panel in front of him. Two lights are green; two, red. The upper left and lower right bulbs are red. For each trial one red lamp and one green lamp are lighted in a randomized arrangement. In terms of the position of the red light with respect to the green light (four possible combinations), the examinee responds by throwing one of four switches. For example, if red is at the left of green, he throws the left-hand switch. When the correct response is made, an illuminated white bulb located above the four others, goes out. Following ten practice trials, eighty test trials of four series of twenty settings were given. Recorded automatically by electric clocks, the total of the reaction times is the score.

The *Two-Hand Coordination Test,* CM101A, was designed to measure one's ability to coordinate the movements of both hands in the performance of a complex motor task. At the slow rate of one revolution per minute a target on a disk moves in a circle and at the same time runs slowly back and forth in a slot upon the disk. In an attempt to keep the pointer upon the irregularly moving target, the examinee can move the pointer forward or backward by turning with his right hand a crank; right or left, by turning another crank with his left hand. The coordinated simultaneous movement of both hands is required for a good score. For eight trials of one minute each the total amount of time that the examinee keeps the pointer on the target represents the score. The contact time, expressed in hundredths of a minute, is automatically recorded by an electric clock.

The *Fernald Rudder Control Test,* CM120B, was designed to measure the examinee's ability to effect appropriate motor responses to changes in position indicated by visual and kinesthetic cues. The apparatus consists of a seat and rudder bar mounted on a boom which is placed on a swivel. With the center of gravity in front of the swivel, a state of disequilibrium is evidenced in the constant tendency of the apparatus to swing to one side. When force is appropriately exerted on the rudder bar, a resulting change occurs in the tension of the control springs. With equilibrium temporarily restored, the device is returned to a straight-ahead position. Other parts of the apparatus include a control stick (not functionally related to the uprighting of the device) and a cowling along which the examinee aims at a target. Following a one-minute practice trial, five test trials of one minute each are scored in the unit of number counts (recorded on a counter) in .25 seconds.

2. *Tests of the Classification Battery of September, 1944*

In both the November, 1943, Classification Battery and the September, 1944, Classification Battery the same six apparatus tests were used. Among the printed tests the following were identical in the two batteries: *Reading Comprehension,* CI614H; *Spatial Orientation I,* CP501B; *Spatial Orientation II*, CP503B; *Dial and Table Reading*, CP622-21A; *Biographical Data*, CE602D (both Navigator and Pilot Weightings), and *Mathematics B (Arithmetic Reasoning)*, CI206C.

Three tests in the more recent battery were comparable to those of the earlier battery, since only slight revisions were made. Five new pencil and paper tests appeared in the second battery: *Numerical Operations Front*, CI702B; *Numerical Operations Back*, CI702B; *Practical Judgment*, CI301C; *Mechanical Information*, CI905B; and *Speed of Identification*, CP610A. The test *Mathematics A*, CI702F, of the earlier battery was replaced by two tests upon numerical operations.

For the three tests of the earlier battery only minor changes were effected. Instead of *Mechanical Principles*, CI903A, and *General Information*, CE505E, the revised tests *Mechanical Principles*, CI903B, and *General Information*, CE505F, were used in the September, 1944, Classification Battery administered to the West Point cadets. The differences between the two forms of the *Mechanical Principles* test are negligible. In the *Composite Factor Analysis Summary*, only inconsequential differences in the factor composition of these two tests are present. Aside from a few changes in certain items these two forms for all intents and purposes are equivalent.

In the test of *General Information*, CE505F, the number of multiple choice items is 110 instead of 100, as in form CE505E. In Part I of the revised test, the content of 50 items is concerned with aviation interest: definitions of terms, identification of planes, and miscellaneous facts of aviation. As for Part II, sixty items of diverse content were used requiring a knowledge of contemporary history, literature, geography, art, music, sports, and such activities as motoring, sailing, and hunting. Supposedly, correct responses to items of such content were to be indicative of the examinee's interests and hobbies. For each part the time limit is twenty minutes. Although both forms CE505E and CE505F of the *General Information Test* are limited to forty minutes, the scoring formula for the revised form is R instead of R—W/4, of the earlier form.

The last of the three tests in which relatively minor changes occurred in the revised forms was *Instrument Comprehension*. In the September battery *Instrument Comprehension I* did not appear, and *Instrument Comprehension*, CI616C, was a revision of the form CI616B. The difference in the forms of the two tests lies, not in content, but in the directions employed. In the earlier batteries in which CI616A and CI616B were used, the directions depended upon the examinee's first having read those pertaining to *Instrument Comprehension I*, CI615A and CI615B, respectively. In the September, 1944, Classification Battery the directions for *Instrument Comprehension*, CI616C, were rewritten, since *Instrument Comprehension I* was absent.

Of the five new tests added to the September 1944, Classification Battery, two were mathematics tests, designed to measure the examinee's accuracy and speed in the performance of fundamental numerical operations. In *Numerical Operations Front* and *Back*, CI702B, the front consists of simple examples in multiplication and addition; the back, of elementary problems in division and subtraction. For each item on the front an answer is given which the examinee indicates to be correct or wrong by blackening one of two spaces. On the back five answers are presented for each item, one of which is correct. The total number of items is 174. In each test the amount of time allowed is five minutes (total time: 10 minutes). The scoring formula for both front and back is: $\frac{R-3W}{2}$

Another new test in the September battery was *Practical Judgment*, CI301C, designed to measure the examinee's ability to solve practical problems. In each of the thirty items, a practical problem involving activities in the Army is described in two or three sentences with conditions specified. Many of the problems represent situational predicaments. To each item four or five choices of action are presented, the best one of which the examinee is to select. With a total time of thirty minutes allowed, the test is scored: 2R—W/2.

The fourth new test was *Mechanical Information*, CI905B, purporting to measure the amount of information the examinee has concerning the function and operation of mechanical devices. In most of the thirty items the examinee selects one of three, four, or five alternative answers that describe best the function of a familiar mechanical gadget or the probable cause of a described mechanical difficulty. With a time limit of twelve minutes, the test is scored: R—W/3.

The last additional test was *Speed of Identification*, CP610A, developed to measure the accuracy of form perception through use of airplane silhouettes. The items are presented in twelve blocks of four each—a total of 48 items. Within each block, four planes (items) in the form of blue silhouettes are oriented in the same position. Although upon initial perception they appear quite similar, minute differences are apparent in each one. Slight differences in shapes of wing, fuselage, gear, and tail, for example, appear. Corresponding to each item of a given block are five alternative choices of planes in silhouette form (also very similar in appearance) which are rotated at different angles from the given standard position of the stimulus plane. The examinee is to pick for each stimulus plane in the first block one of the five (response) planes within a second block adjacent to the first block (of stimuli). Within each of the twelve stimulus blocks the appearance of the four stimulus planes is relatively homogeneous, but between the different blocks of stimulus planes marked heterogeneity is apparent in that the planes are of diverse models. A similar statement may be made concerning the twelve blocks of response planes. Clocked at only four minutes, the test is scored as R—W.

CHAPTER III

STATISTICAL PROCEDURES

In the following two major divisions concerning statistical procedures employed in the investigation, the order of the described techniques follows closely the sequence of the two major questions and of the sub-questions in the *statement of the problem* (I, B, 1). The portion upon factor analysis parallels the first major question, and the next section upon multiple regression equations in the prediction of the pass-fail criterion scores closely conforms to the second major question. Emphasis is placed upon the statistical operations used rather than upon interpretation of results.

A. The Factor Analysis of the Two Matrices of Intercorrelations

As mentioned previously, the Thurstone system was employed exclusively in the factor analyses of the two matrices of intercorrelations of test in the classification batteries of November, 1943, and September, 1944, administered respectively to the two populations of Negroes and West Point Cadets. All intercorrelations were of the product-moment type with the exception of the biserial coefficients of correlation (equivalent to a Pearsonian r) between the test scores and the pilot pass-fail criterion. The matrices of intercorrelations appear in Table 2 and Table 3.

1. *Limitations of Intercorrelations Used in the Two Factor Analyses*

For the population of Negroes, the biserial correlations placed in the correlation matrix (Table 2) were corrected for restriction of range which resulted from the elimination of approximately 16.3 percent of the candidates (58 of the 356 examinees were disqualified for pilot training). It is not expected that this proportion of elimination would materially affect the factorial composition of the eighteen tests of the classification battery. However, the loadings of the factors in the pass-fail criterion probably would be systematically underestimated, since the biserial correlation coefficient corrected for restriction of range would be larger than one not corrected.

For the population of West Point Cadets, a marked *lack* of normality might be expected in the distribution of test scores. Fortunately, however, the ceilings of the tests were sufficiently high that in most instances approximately normal distributions of test scores resulted. Moreover, for each set of test scores, the amounts of variance present for both the West Point group and the Negro cadets were approximately the same as those of groups of representative white aviation-cadets of about the same size. For the West Point population, the biserial correlations of the pass-fail criterion with test scores were based upon a number of subjects fewer than those given the test battery in that only 355 out of 815 tested elected pilot training. However, a marked difference between the two populations with respect to the reduced number of cases in the pass-fail criterion was apparent. No systematic elimination of pilot cadets on the basis of a low stanine score resulted for the West Point group. Hence, no underestimation of the factor loadings of the pass-fail criterion should appear, although such loadings might be subject to somewhat greater sampling errors than those for the twenty-one tests which are based on more than twice as many cases.

2. *Extraction of Centroid Factors*

For both matrices of intercorrelations the technique of extraction followed precisely the procedure outlined by Guilford (3). For the November, 1943, battery (administered to Negroes) two sets of extractions were required. In the first extraction the computed communalities of five tests differed more than .10 from the estimated communalities used at the beginning of the analysis (the discrepancies being +.14, +.12, +.12, +.12, and —.12). For the September, 1944, battery (administered to West Point Cadets) only one extraction of centroid factors was necessary, since the discrepancies between estimated communalities and computed communalities were all less than .10.

Criteria for the cessation of extraction are numerous. Experiences of AAF psychologists of the research units revealed that many criteria used prior to the War were too conservative.

TABLE 2

INTERCORRELATIONS OF TESTS AND BISERIAL VALIDITY COEFFICIENTS OF TESTS WITH THE PILOT PASS-FAIL CRITERION FOR THE NOVEMBER, 1943, TEST BATTERY ADMINISTERED TO NEGROES AT THE MEDICAL AND PSYCHOLOGICAL EXAMINING UNIT NUMBER 6, KEESLER FIELD, MISSISSIPPI ($N=356$)[a]

Tests and Code Numbers		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Reading Comprehension	CI614H	..	14	14	35	04	05	19	24	40	43	30	29	−01	09	08	22	−02	00	11
2. Spatial Orientation II	CP503B	14	..	39	25	12	00	24	25	09	09	24	37	01	08	05	27	09	05	18
3. Spatial Orientation I	CP501B	14	39	..	35	04	09	05	02	16	07	25	32	06	18	23	29	18	01	16
4. Dial and Table Reading	CP622-21A	35	25	35	..	06	01	02	08	39	41	49	33	08	29	28	40	15	−03	14
5. Biographical Data—Navigator	CE602D	04	12	04	06	..	26	07	02	11	05	00	12	21	05	06	09	00	05	05
6. Biographical Data—Pilot	CE602D	05	00	09	01	26	..	23	33	−08	−15	06	14	25	20	07	07	24	22	17
7. Mechanical Principles	CI903A	19	24	05	02	07	23	..	30	00	14	10	33	16	22	10	17	31	24	17
8. General Information	CE505E	24	25	02	08	02	33	30	..	−02	−01	19	30	13	15	04	12	13	09	21
9. (General) Mathematics A	CI702F	40	09	16	39	11	−08	00	−02	..	54	19	19	−07	04	09	21	−06	01	−01
10. Mathematics (Arithmetic Reasoning) B	CI206C	43	09	07	41	05	−15	14	−01	54	..	22	18	−06	05	06	16	−03	27	05
11. Instrument Comprehension I	CI615B	30	24	25	49	00	06	10	19	19	22	..	43	10	27	23	30	23	04	17
12. Instrument Comprehension II	CI616B	29	37	32	33	12	14	33	30	19	18	43	..	18	31	20	35	23	16	36
13. Rotary Pursuit	CP410B	−01	01	06	08	21	25	16	13	−07	−06	10	18	..	31	26	12	32	27	14
14. Complex Coordination	CM701A	09	08	18	29	05	20	22	15	04	05	27	31	31	..	33	37	52	28	07
15. Finger Dexterity	CM116A	08	05	23	28	06	07	10	04	09	06	23	20	26	33	..	25	34	15	00
16. Discrimination Reaction Time	CP611D	22	27	29	40	09	07	17	12	21	16	30	35	12	37	25	..	27	10	22
17. Two-Hand Coordination	CM101A	−02	09	18	15	00	24	31	13	−06	−03	23	23	32	52	34	27	..	37	15
18. Rudder Control	CM120B	00	05	01	−03	05	22	24	09	01	27	04	16	27	28	15	10	37	..	33
19. Pass-Fail Pilot Criterion	($r_{bis.}$)	11	18	16	14	05	17	17	21	−01	05	17	36	14	07	00	22	15	33	..

[a] All decimal points are omitted in this table and in all subsequent tables.

TABLE 3

INTERCORRELATIONS OF TESTS AND BISERIAL VALIDITY COEFFICIENTS OF TESTS WITH THE PILOT PASS-FAIL CRITERION FOR THE SEPTEMBER, 1944, CLASSIFICATION BATTERY ADMINISTERED TO THE WEST POINT CLASS OF 1946 ($N=815$)

Tests and Code Numbers		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. Mathematics B (Arithmetic Reasoning)	CI206C	..	38	19	30	15	−01	29	40	54	33	16	29	37	32	12	−04	14	17	09	21	04	05
2. Dial and Table Reading	CP622-21A	38	..	41	27	15	03	52	50	25	19	13	38	11	07	31	18	25	37	14	31	20	22
3. Spatial Orientation I	CP501B	19	41	..	46	09	18	19	19	19	18	22	37	19	12	47	13	25	27	17	26	18	12
4. Spatial Orientation II	CP503B	30	27	46	..	15	19	−01	08	38	31	29	38	42	34	39	10	29	30	17	23	08	10
5. Biographical Data—Navigator	CE602D	15	15	09	15	..	12	07	04	11	12	05	10	08	04	10	−01	−03	05	−06	02	01	07
6. Biographical Data—Pilot	CE602D	−01	03	18	19	12	..	−13	−07	03	21	49	22	33	36	21	15	26	17	20	07	06	29
7. Numerical Operations Front	CI702B	29	52	19	−01	07	−13	..	63	10	03	−03	12	−17	−11	07	06	05	16	−03	19	15	01
8. Numerical Operations Back	CI702B	40	50	19	08	04	−07	63	..	13	04	01	22	−04	−06	13	07	10	11	00	24	11	09
9. Reading Comprehension	CI614H	54	25	19	38	11	03	10	13	..	44	30	34	45	43	16	08	19	22	19	18	05	11
10. Practical Judgment	CI301C	33	19	18	31	12	21	03	04	44	..	34	23	38	36	13	13	19	16	16	15	05	11
11. General Information	CE505F	16	13	22	29	05	49	−03	01	30	34	..	36	46	54	21	15	28	26	28	14	06	24
12. Instrument Comprehension	CI616C	29	38	37	38	10	22	12	22	34	23	36	..	39	29	33	17	34	36	31	30	10	36
13. Mechanical Principles	CI903B	37	11	19	42	08	33	−17	−04	45	38	46	39	..	65	18	21	44	34	40	22	09	35
14. Mechanical Information	CI905B	32	07	12	34	04	36	−11	−06	43	36	54	29	65	..	14	13	35	28	29	13	08	18
15. Speed of Identification	CP610A	12	31	47	39	10	21	07	13	16	13	21	33	18	14	..	08	22	25	14	22	14	06
16. Rotary Pursuit	CP410B	−04	18	13	10	−01	15	06	07	08	13	15	17	21	13	08	..	34	36	38	22	33	36
17. Two-Hand Coordination	CM101A	14	25	25	29	−03	26	05	10	19	19	28	34	44	35	22	34	..	51	49	34	28	32
18. Complex Coordination	CM701A	17	37	27	30	05	17	16	11	22	16	26	36	34	28	25	36	51	..	46	40	32	34
19. Rudder Control	CM120B	09	14	17	17	−06	20	−03	00	19	16	28	31	40	29	14	38	49	46	..	25	23	39
20. Discrimination Reaction Time	CP611D	21	31	26	23	02	07	19	24	18	15	14	30	22	13	22	22	34	40	25	..	28	22
21. Finger Dexterity	CM116A	04	20	18	08	01	06	15	11	05	05	06	10	09	08	14	33	28	32	23	28	..	19
22. Pass-Fail Criterion ($r_{bis.}$)		05	22	12	10	07	29	01	09	11	11	24	36	35	18	06	36	32	34	39	22	19	..

Extraction of additional factors instead of adding mere error variance actually contributed to the realization of improved rotations and hence more meaningful factors. One criterion for cessation widely used by the members of the psychological research units was that if the product of the two absolutely highest factor loadings on the nth extraction was less than the standard error of zero correlation for the size of sample used (assuming null hypothesis) then the n-1 extraction is the last one to be included. Symbolically, if $|k_{ni} \times k_{nj}| < r_o$, then the extraction should cease at the n-1 centroid factor.

For the correlation matrix of the November battery, the product of the highest two loadings of the eighth centroid factor, .179 × .176, was equal to .0315, which was less than .051, the standard error of zero correlation for a sample of 356 (354 degrees of freedom). Accordingly seven factors were rotated. (After the first seven rotations the eighth factor was introduced in order that the other factors might be better defined. Elaboration of this point follows in a subsequent paragraph: *technique of rotation.*)

For the correlation matrix of the September battery the product of the highest two loadings of the ninth centroid factor (+.146) × (—.138) was equal to .0201, which was less than .0351, the standard error of zero correlation for a sample of 815 (813 degrees of freedom). Therefore eight factors were rotated. (After 44 rotations, the ninth factor had to be introduced in order to obtain a more meaningful solution—a point subsequently to be explained.)

Another indication of the adequacy of a centroid extraction is the dispersion of the residuals of the last centroid extracted. For the matrix of intercorrelations of the November, 1943, battery the 190 seventh factor residuals ranged from —.119 to +.073. The standard deviation of this distribution was .0283.

For the matrix of intercorrelations of the September, 1944, battery the 253 eighth factor residuals ranged from —.053 to +.061. The standard deviation of this distribution was .0173. For both matrices the standard deviations of the distribution of residuals were considerably less than the standard errors of zero correlation for the respective numbers of degrees of freedom.

3. *Techniques of Rotation*

Graphical rotation by the Zimmerman method (20) was employed for both matrices of intercorrelations.[1] The purpose of the rotation was to realize both *simple structure* and *positive manifold.* In other words, an attempt was made to obtain a maximum number of entries of a value near zero and to reduce the number and size of negative loadings.

For the matrix of intercorrelations of the November, 1943, battery (administered to Negroes) seven factors were used in the first seven rotations (in adherence to the criterion of cessation of extraction). However, since at the conclusion of the seventh rotation tests known from previous analyses to be loaded with numerical, reasoning, and verbal variances, were clustered about an axis, the eighth centroid factor was introduced in order to separate, if possible, any possible components of what appeared to be a general intellectual factor. Variance in the so-called intellectual factor did separate: variance identified as verbal remaining on what had been the intellectual factor and the remainder of the variance, evidently number and reasoning, going over to the eighth *orthogonal factor.* Since the two tests of the battery most heavily loaded in this "doublet" factor had been revealed in previous analyses to be complex with respect to both reasoning and number, and since no test known to be relatively pure with respect to either factor was included in the battery, the outcome was not surprising. The final factor loadings at the conclusion of the forty-fifth rotation are presented along with test communalities in Table 4.

For the matrix of intercorrelations of the September, 1944, battery (administered to West Point Cadets) the first eight centroid factors were subjected to forty-two rotations. The seventh factor appeared difficult to define, since it contained a high loading (.480) in the pass-fail criterion and a substantial loading (.429) in only one test (*Biographical Data—Pilot,* CE602D). Without success, several trial rotations were undertaken to increase the loadings of this factor on other tests in order that identification might be facilitated (and at the same time to fulfill the requirements of both positive manifold and simple structure). Finally, the ninth factor, which according to the criterion employed for cessation need not be retained, was introduced in several trial rotations with the seventh factor and others in the expectation that both factors seven and nine might be identified. Again several trial rotations failed to reveal a solution satisfactory to the requirements of positive manifold and simple structure

[1] This outstanding innovation enables an experienced technician to perform as many as fifteen rotations of pairs of axes per hour on which twenty test variables have loadings. It is highly probable that the Zimmerman technique will soon replace the more cumbersome and inefficient procedures now commonly employed by many factor analysts.

TABLE 4

ROTATED FACTOR LOADINGS FOR TESTS OF THE NOVEMBER, 1943 CLASSIFICATION BATTERY ADMINISTERED TO NEGROES AT KEESLER FIELD AND FOR THE PILOT PASS-FAIL CRITERION[a]

Test and Code Number		I(PM)	II(V)	III(P)	IV(S)	V(PI)	VI(ME)	VII(K)	VIII(RN)	h^2 Communality
1. Reading Comprehension	CI614H	−090	526	024	090	013	313	−030	285	473
2. Spatial Orientation II	CP503B	−100	−030	551	219	043	325	−019	081	478
3. Spatial Orientation I	CP501B	065	046	519	280	061	005	−026	166	386
4. Dial and Table Reading	CP622-21A	−033	435	208	487	181	−114	−025	397	675
5. Biographical Data—Navigator	CE602D	134	086	234	−175	404	018	073	031	280
6. Biographical Data—Pilot	CE602D	176	−106	−023	080	562	264	075	−053	444
7. Mechanical Principles	CI903A	216	−010	069	061	031	543	165	081	386
8. General Information	CE505E	−041	130	−045	207	226	528	098	−112	417
9. General Mathematics A	CI702F	−041	443	149	−083	018	005	−075	551	537
10. Mathematics B (Arithmetic Reasoning)	CI206C	−027	420	026	−115	−097	050	225	628	648
11. Instrument Comprehension I	CI615B	012	387	144	512	036	063	092	096	456
12. Instrument Comprehension II	CI616B	105	290	374	369	000	372	238	014	566
13. Rotary Pursuit	CP410B	497	081	038	095	252	047	186	−127	381
14. Complex Coordination	CM701A	495	−031	−054	503	073	160	−003	207	577
15. Finger Dexterity	CM116A	454	158	130	273	015	−058	−025	092	341
16. Discrimination Reaction Time	CP611D	158	100	260	400	051	137	010	278	362
17. Two-Hand Coordination	CM101A	533	−190	−018	425	051	160	173	156	584
18. Rudder Control	CM120B	311	−097	−104	020	074	146	554	240	508
19. Pass-Fail Criterion		−050	098	260	190	060	130	500	−060	381

[a] The capital letters in parentheses adjacent to the Roman numerals placed at the heads of the columns represent abbreviations for the names of the factors, as follows:

PM—Psychomotor Coordination
V—Verbality
P—Perceptual (Speed)
S—Spatial Relations
PI—Pilot Interest
ME—Mechanical Experience
K—Kinesthesis
R-N—Reasoning and Number (an intellectual factor)

TABLE 5

ROTATED FACTOR LOADINGS FOR TESTS OF THE SEPTEMBER, 1944, CLASSIFICATION BATTERY ADMINISTERED TO WEST POINT CADETS OF THE CLASS OF 1946 AND FOR THE PILOT PASS-FAIL CRITERION[a]

Test and Code Number		I(ME)	II(N)	III(PM)	IV(P)	V(R)	VI(S)	VII(Res)	VIII(V)	IX(PI)	h^2 Communality
1. Mathematics B (Arithmetic Reasoning)	CI206C	164	373	−087	015	513	137	025	404	−046	622
2. Dial and Table Reading	CP622-21A	−096	587	155	262	076	307	−137	237	094	631
3. Spatial Orientation I	CP501B	011	189	056	627	−029	214	−003	181	111	524
4. Spatial Orientation II	CP503B	223	−001	−035	549	181	219	−110	305	018	538
5. Biographical Data—Navigator	CE602D	−056	064	−044	086	160	−087	025	163	250	141
6. Biographical Data—Pilot	CE602D	441	−043	075	173	−060	030	−003	−065	531	523
7. Numerical Operations—Front	CI702B	−127	784	120	−006	−019	088	−003	057	−056	659
8. Numerical Operations—Back	CI702B	−175	708	037	017	217	166	123	108	090	643
9. Reading Comprehension	CI614H	261	098	018	053	305	148	007	625	−042	589
10. Practical Judgment	DI301C	298	052	113	101	147	−029	−047	486	137	394
11. General Information	CE505F	580	047	025	112	−095	180	098	244	352	588
12. Instrument Comprehension	CI616C	105	129	002	230	103	498	050	288	284	507
13. Mechanical Principles	CI903B	550	−159	094	081	404	358	108	233	160	727
14. Mechanical Information	CI905B	699	−040	058	063	219	185	−025	226	075	638
15. Speed of identification	CP610A	065	118	000	627	−002	166	126	078	137	480
16. Rotary Pursuit	CP410B	073	−010	548	015	−082	259	−098	053	156	416
17. Two-Hand Coordination	CM101A	328	015	404	186	123	463	057	−017	010	538
18. Complex Coordination	CM701A	172	105	473	227	070	453	−035	108	−013	540
19. Rudder Control	CM120B	218	−120	406	016	010	496	058	094	104	496
20. Discrimination Reaction Time	CP611D	000	160	351	231	200	256	130	102	037	336
21. Finger Dexterity	CP116A	013	011	508	188	056	069	060	−051	016	321
22. Pass-Fail Criterion		040	−053	305	−040	075	415	−120	050	436	484

and still open to a psychologically meaningful interpretation. Since all loadings on the seventh factor other than those of the pilot pass-fail criterion and one test were negligible, the only remaining approach appeared to be that of attempting to build up factor nine and to reduce factor seven to the role of a residual (a factor containing small positive and negative loadings). After twenty-three additional rotations, factor nine was loaded highly (.436) in the criterion and substantially in two tests (.531 and .352), and factor seven became a residual with both positive and negative loadings within the narrow range of —.137 to .130. A psychologically meaningful interpretation of factor nine appeared feasible. The factor loadings and communalities for tests at the conclusion of the final rotation are given in Table 5.

On the basis of several factor analyses the weighted loadings of the factors and communalities, as well as reliabilities and validities of the test, have been reported in the *Composite Factor Analysis Summary* (24). These loadings will be taken to represent those of a representative white aviation-cadet population.

4. *Satisfactory Fulfillment of the Requirements of Rotation*

The criterion of positive manifold was reasonably well satisfied in the final rotations of both matrices. After the final rotation (the forty-fifth) of the eight centroid factors extracted for the November, 1943, battery, only eight loadings were equal to or algebraically less than —.100 (the highest negative loading being —.190). In all, thirty-two negative entries appeared for the eight rotated factors.

For the final rotation of the matrix of the West Point population the highest negatives in excess of —.100 in factors (other than the residual) were —.175, —.159, and —.127. Twenty-seven other negative loadings in the eight meaningful factors ranged between —.099 and .000, with only nine being between —.099 and —.050. In the residual factor eleven loadings were positive, eleven negative, although the highest negative was only —.137 (and the highest positive only .130).

In the rotation procedure for both matrices an attempt was made wherever possible to maximize the number of vanishing loadings. In view of the fact that most of the tests in both batteries were complex in previous analyses, it was not surprising that final rotations for both matrices of intercorrelations revealed a not too high degree of simple structure. If loadings absolutely less than .150 are considered to be insignificant, inspection of Table 4 shows that for the November battery (administered to Negroes) the number of insignificant weights for each of the eight factors was respectively as follows: 11, 11, 12, 7, 14, 11, 12, and 10. Similarly, for the September battery the number of insignificant loadings for the eight real factors was respectively: 10, 15, 14, 12, 14, 7, 11, and 15. Despite the complexity of the tests, a fair degree of simple structure was evident for both matrices in the final rotation.

Another condition to be fulfilled which in mathematical terms is absolute was that the sum of common-factor variances (communality) of a test after rotation must equal its common-factor variance before rotation (centroid factor variance). In other words, the sum of common-factor variances of a test, or the magnitude of a test vector, remains invariant under rotation. In the Zimmerman method of rotation, negligible discrepancies accrue after numerous rotations to the extent that any inaccuracies in scales employed or any degree of bluntness of a pencil point are present.

At the conclusion of the rotations of the factors of the November, 1943, battery, minor discrepancies between the communalities of tests before rotation and after rotation existed. With the communality of each test in the centroid factors taken as the standard of reference, the common-factor variance following the forty-fifth rotation deviated from the assumed standard between —.012 and +.002. In general, the discrepancies were negative.

In the rotations of the second matrix the discrepancies ranged between —.009 and +.006, again with a tendency for a slight underestimation. However, these discrepancies were of no practical consequence. The requirement of invariance was considered to be satisfied for both matrices.

B. Multiple Regression Equations in the Prediction of Each Pilot Pass-Fail Criterion from Tests of the Respective Batteries

In order that the validity of each test battery in terms of the multiple correlation between the pass-fail criterion and the test scores (optimally weighted) might be determined and in order that beta weights of these tests might be ascertained for inclusion in multiple regression equations, the Doolittle method as outlined by Guilford (2, pp. 263-268) was employed.

Ezekiel's correction (1) was employed for the coefficient of multiple determination (R^2) and the coefficient of multiple correlation (R), in order that an unbiased estimate of the most probable values of the parameters in the universe corresponding to these statistics might be obtained. (These data are presented in Table 15.)

CHAPTER IV

INTERPRETATION OF THE STATISTICAL RESULTS

In Chapter I, two major questions were proposed in the statement of the problem. In Chapter III, the survey of statistical procedures closely followed the sequence of these two major questions. Emphasis was placed upon a rationale for, and description of, the statistical operations employed. In the current chapter, the three major divisions are concerned with an interpretation of the statistical results. These divisions also parallel the organization of the statement of the problem presented in Chapter I.

In part A, for the two populations of West Point Cadets and Negroes, the factors will be tentatively identified and interpreted. Among the three groups of West Point Cadets, Negroes, and the white aviation-cadets[1] (in general, pilots), comparisons will be made of the amounts of loadings of the identified factors in tests and in the pilot pass-fail criteria. Hypotheses will be set up to account for marked differences in the weights of the factors in tests common to either two of, or three of, the groups and for differences in the factorial composition of the three pilot criteria.

In part B, an interpretation of the results of the traditional multiple-regression equation technique for the two populations of West Point Cadets and Negroes will be undertaken. The two points of major emphasis will include (*a*) a comparison of the validity of each test battery in terms of the coefficient of multiple correlation between the pilot pass-fail criterion and the test scores of the battery optimally weighted, with the factorial validity estimated from the total common-factor variance of the respective pilot criteria, and (*b*) a comparison of the proportion of variance-contribution of each optimally weighted test to the coefficient of multiple determination between the pilot criterion and the tests of each battery.

A. An Interpretation of the Factors and a Comparison of Their Loadings in Tests and in the Pilot Criteria for Three Aviation-Cadet Populations

For convenience, seven of the nine factors to be identified and interpreted are somewhat arbitrarily grouped as intellectual functions of verbality, number, and reasoning; as factors of perception and spatial relations; and as factors of mechanical experience and pilot interest. Psychomotor-coordination and kinesthesis are considered independently. For the three populations of West Point Cadets, of Negroes, and of representative white aviation-cadets, mostly pilots, comparisons will be made of the factor loadings in tests administered to the West Point Cadets and Negroes and of factor loadings in the pilot pass-fail criterion. In order that each factor may be more readily defined and discussed, a weight of a final rotated factor equal to or exceeding +.175 in any test administered to either the West Point population or the Negro group is listed in accompanying tables, along with the loadings in the test for the other two populations (whether these loadings be greater than, equal to, or less than the arbitrarily

[1] As mentioned previously, the data for the representative white aviation-cadet population were taken from (24). In most instances the loadings are for pilot cadets.

chosen weight of .175). In those instances in which a test is unique to either the November, 1943, or September, 1944, battery, an additional factor loading, of course, can be reported only for the white cadet population. As an economy measure in the tables to follow the roman numerals I, II, and III corresponding respectively to the groups of West Point Cadets, Negroes, and (representative) white cadets are to be placed at the head of each of the last three columns in which the loadings with respect to a given factor are listed for the several tests. At the conclusion of each list of tests for which the factor weights are reported is an entry for the pilot criterion.

Throughout the interpretation of the factors, hypotheses will be set up to rationalize noticeable differences among the loadings in the same test or in the same set of tests in two of, or three of, the groups. As a conclusion to the discussion of each factor, brief mention will be made of the importance of the factor for each of the three populations to the predictive value, or validity, of the test battery primarily in terms of the magnitude and the sign of the loading in the criterion.

One of the fundamental weaknesses of the Thurstone system of factor analysis is that the difference between two (or more) factor loadings derived from either the same matrix or from different matrices cannot be tested for statistical significance. In other words, there is no means for computing the standard error of a factor loading. The reason for this is that in rotation an infinite number of solutions is possible.[2] The slightest rotation results in a change in the magnitude of the projections of the test vector upon the axes entering into the rotation. To the extent, however, that the Thurstone requirement of positive manifold and simple structure is fulfilled, the range within which the factor loadings may vary is somewhat limited. If these two criteria are fulfilled, limited comparisons are feasible. As the number of observations entering into the computations of the original intercorrelations increases, the greater, of course, is the stability of the factor loadings derived. Even with a sample of several thousand, the actual significance of the difference between two loadings can never be known in terms of any stated degree of probability. Therefore, in any comparisons to be made between the loadings of a given factor in the same test or in different tests for two or more of the groups, considerable caution and restraint should be exercised.

In the first chapter, the importance of a factor as an intervening variable, or theoretical construct, was emphasized. Although the naming, or labeling, of a factor is convenient for purposes of reference, the emphasis in an operational approach is placed upon the existence of a real variable of a reasonable degree of stability which stands in functional relationship to certain antecedent conditions such as test items (independent variables) and consequent conditions of human behavior such as responses to the items (dependent variables). The basis of description of a factor rests upon certain more or less common properties, or requirements, which are common to

[2] In fact, with n axes, the number of possible solutions is ∞^{n-1}. After the first centroid axis is determined, each of the n-1 axes may assume an infinite number of positions.

certain tests (the antecedent conditions, or *S*-variables, referred to in I, C, 1) and *not* common to others. In reference to these properties, certain communalities of responses (the consequent conditions, or *R* variables) are present. Through application of statistical procedures of correlation and factor analysis, these communalities of response appear in the form of loadings on factors. The names applied to these factors are subject to change and to modification, but the presence of the factor by the operational definition is relatively stable. Hence, any of the descriptions of factors to follow might better be considered as *tentative* hypotheses, subject to modification, which serve to communicate to the reader the presence of useful and perhaps endurable categories.

In short, whereas the definition of a factor is more or less arbitrary (often depending upon what psychological terms are in vogue), the existence of the factor is relatively certain, since its presence usually can be repeatedly demonstrated under controlled conditions of testing. Therefore, the following labels of factors stand for abbreviated hypotheses. Finally, introspective references are employed wherever their use makes the interpretation of a factor more comprehensible.

For convenience of reference to the size of numerical values of factor loadings, five descriptive terms, arbitrarily chosen, are to be used. A loading of .500 or greater is said to be *high;* one between .400 and .499, *substantial;* one between .300 and .399, *moderate;* one between .175 and .299, *slight;* and finally, one less than .175, insignificant or *negligible.*

1. *The Intellectual Factors of Verbality, Reasoning, and Number*

a. Verbal factor. One of Thurstone's primary mental abilities, the verbal factor, is identified with those tests in which the comprehension of meanings of words, singly or collectively, and of the ideas associated with them is present (17). Tests of reading comprehension, of vocabulary, and of intelligence (scholastic aptitude) in the past have revealed high loadings in this factor. Among the tests of the two classification batteries, fourteen are loaded .181, or more, for one or more of the three groups. The loadings of these fourteen tests in the verbal factor are presented for the three groups in Table 6.

The test most indicative of a verbal factor is that of *Reading Comprehension,* with comparable factor weights of .625, .526, and .600 for the three groups of West Point Cadets, of Negroes, and of white cadets. The substantial saturations of the two mathematics tests in the verbal factor are not surprising, in view of the fact that a verbal statement of most problems is presented. Despite the fact that the verbal factor loadings of the three groups in *Mathematics B* appear to be slightly lower than those of *Mathematics A,* the importance of verbal comprehension in *Mathematics B* is particularly apparent in the problems designed to measure reasoning—word problems similar to those encountered in high school algebra. Similarly, in the three tests: *General Information,* CE505E, *General Information,* CE505F, and *Mechanical Information,* all the multiple-choice items consist of verbal statements.

In such a complex test as *Dial and Table Reading,* the loadings for the three groups in the verbal factor are

TABLE 6

VERBAL-FACTOR LOADINGS IN DEFINITIVE TESTS, AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION-CADETS

Test and Code Number		Group		
		I[a]	II[b]	III[c]
Reading Comprehension	CI614H	625	526	60
Mathematics B (Arithmetic Reasoning)	CI206C	404	420	27
(General) Mathematics A	CI702F		443	37
Dial and Table Reading	CP622-21A	237	435	10
General Information	CE505E		130	43
General Information	CE505F	244		37
Instrument Comprehension	CI616C	288		17
Instrument Comprehension I	CI615B		387	22
Instrument Comprehension II	CI616B		290	24
Practical Judgment	CI301C	486		46
Spatial Orientation I	CP501B	181	046	08
Spatial Orientation II	CP503B	305	−030	14
Mechanical Information	CI905B	226		26
Mechanical Principles	CI903B	233		03
Pilot Criterion		050	098	−05

[a] West Point cadets.
[b] Negro aviation cadets.
[c] Representative white aviation-cadets.

readily explained by the lengthy directions and illustrative problems which are expressed in verbal terms. Moreover, words are overprinted upon the dials and tables. Unless an examinee is able to associate the detailed directions at the beginning of every part of the test with subsequent items, and to relate verbal symbols in the items themselves to the numbers presented upon the dials and tables, he obviously would be unable to respond in an adequate manner to the simplest test item. In the tests upon spatial orientation and instrument comprehension, the detailed directions again appear to account for the slight verbal factor weights, inasmuch as a successful performance upon these tests (novel to most of the examinees) requires a thorough understanding of the directions. The loading of .387 in *Instrument Comprehension I,* somewhat higher than those loadings found in the other two forms of this test, may be attributed in part to the verbal descriptions of plane positions in the multiple-choice responses. In the other two forms of the test, pictures of planes in different positions make up the multiple-choice responses.

In the test, *Practical Judgment,* the substantial loading of .486, is not surprising inasmuch as the situational problems are entirely verbal in statement, as are the multiple-choice solutions. Viewed superficially, this test is as highly verbal as any other test in the battery, in that the description of each problem requires three or four sentences of relatively complex construction and of varied vocabulary.

In the test *Mechanical Principles,* the slight weight of .233 may be accounted for, as in some of the other tests, perhaps by the tendency of the examinee to "talk sub-vocally" or to verbalize as he attempts to solve the item. Such an introspective conjecture does seem somewhat reasonable in that as the examinee studies each diagram with its arrows and vectors he may project himself into the situation and set out to solve the problem through the aid of verbal symbols. Of course, the verbal statement of each item, as well as the directions, may account for the slight loading. It may be that the West Point group with more academic training tended to verbalize their actions while matching photographs with larger aerial photographs or with maps. The directions for these two tests are relatively short.

A somewhat surprising result is the tendency for the Negro group (with the exception of the two tests upon spatial orientation) to place somewhat higher than the West Point and white cadet groups in their loadings upon the verbal factor in tests other than *Reading Comprehension*. Just the opposite outcome might be expected for the West Point Cadets, who in most instances have had more experience with verbal material through their academic training. An hypothesis is suggested to rationalize this apparent difference; namely, that the weights in the verbal factor are a function of the level of difficulty of the verbal material with respect to each of the populations. Specifically, it appears that the level of difficulty is pitched perhaps too high for Negroes in the *Reading Comprehension* test (as evidenced by a loading of .526 compared with those of .625 and .600 for groups I and III respectively). The level of difficulty appears to be more favorable to the Negroes in two places: first, in the lengthy directions of such a test as *Dial and Table Reading* (as evidenced by a loading of .435 compared with .237 and .100 for groups I and III respectively) and second, in the verbal material of the items in tests intended to be non-verbal, such as *Instrument Comprehension I, Instrument Comprehension II, Mathematics A,* and *Mathematics B.* In general, the directions for most tests are designed to be considerably below an examinee's level of reading comprehension.

Evidence contrary to the hypothesis is suggested by the test *General Information,* CE505E, with a loading of .130 for Negroes, compared with that of .430 for the white cadet population. No reason can be given for this difference unless it be (as seems unlikely) that the loading of .528 in the factor of mechanical experience has absorbed a disproportionate part of the relatively low communality of .417.

Another interesting result, though likely an insignificant one, is the slight positive contribution of the verbal factor to the validities of the two classification batteries administered to West Point Cadets and Negroes, as evidenced by the respective factor weights of .050 and .098 for Negroes. These two positive loadings in the pilot criterion stand in contrast to that −.050 for the white cadet population. A workable, or plausible, hypothesis is lacking to account for these differences which may likely be only chance fluctuations easily attributed to sampling errors. (It is possible, however, that verbal instructions may play a more dominant role in one pilot training program than in another.)

b. Number factor. One of the easiest factors to identify, the number factor, also one of Thurstone's primary abilities, occurs in those tasks in which the simple fundamental numerical operations of arithmetic: addition, subtraction, multiplication, and division are involved. The use of numbers, as in the test of *Dial and Table Reading,* seems to give rise to the factor. In fact, the mere presence of numbers in a test seems to be frequently a sufficient condition for the appearance of a small loading in a factor as indicated by *Spatial Orientation I,* in which a numbered serial photograph is matched with one of the lettered portions of a larger aerial photograph (although the variance in this factor may be attributed to the counting of objects common to the two photographs in the matching of them).

In the list of tests weighted for the number factor (see Table 7), attention should be called to the fact that the factor loadings reported for the Negroes (Column II) are probably inflated with reasoning variance, since only the verbal variance was separated from the "general intellectual" factor.[3] Hence each comparison of loadings to be made is limited in its scope. The most heavily weighted tests in the number factor appear in Table 7.

Most illuminating in the identification of the number factor are the two tests in numerical operations, not included in the November, 1943, battery. That the presence of these pure tests was needed for the identification of a number factor in the November battery is evidenced in part by the failure of the numerical and reasoning components to separate in the Negro group. All tests included in the November battery evidently capable of identifying numerical or reasoning factors were in terms of previous analyses, for the white cadet population, extremely complex with respect to numerical, reasoning, and verbal components. Hence, as pointed out in the previous chapter (III, A, 3) the difficulty of separating the verbal factor from the other two intellectual factors was evident in the factor-analysis procedure.

In those tests not primarily concerned with mathematics, the use or the presence of numbers in the items results in a loading in the number factor. As in a preceding paragraph, the *Dial and Table Reading Test,* with a high weight in the number factor for all three groups demands the use of numbers in one way or another (see the description of the test in II, A, 1). Even in the *Reading Com-*

TABLE 7

NUMBER-FACTOR LOADINGS IN DEFINITIVE TESTS AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION CADETS[a]

Test and Code Number		Group		
		I	II	III
Numerical Operations—Front	CI702B	784		78
Numerical Operations—Back	CI702B	708		81
(General) Mathematics A	CI702F		551	51
Mathematics B (Arithmetic Reasoning)	CI206C	373	628	48
Dial and Table Reading	CP622-21A	587	397	53
Spatial Orientation I	CP501B	189	166	18
Reading Comprehension	CI614H	098	285	12
Discrimination Reaction Time	CP611D	160	278	18
Complex Coordination	CM701A	105	207	05
Rudder Control	CM120B	−120	240	−03
Pilot Criterion		−053	−060	00

[a] The loadings for Negro cadets are for a general intellectual factor consisting of number and reasoning variance.

[3] A possible indication that the factor has absorbed most of the number and reasoning variance is that the square root of the sum of the number variance and the reasoning variance of tests for the representative white aviation-cadet group approximates the factor loadings of this bifurcated factor in the tests more heavily loaded in it.

prehension Test, highly saturated in the verbal factor, slight weights appear for the number factor, apparently because of the numerical quantities employed in the two reading selections about the Mercator projection and about the air speed meter, and in the items upon these two selections. Small loadings in the two apparatus tests, *Discrimination Reaction Time,* and *Complex Coordination,* may be attributed perhaps to the counting of the number of lights (and to the counting during the process of testing of the number of keys, four in all, in the latter test). The appearance of a weight of .240 for the test of *Rudder Control,* administered to Negroes, cannot be explained in terms of either a numerical or a reasoning component; although by a negative argument the reasoning component, perhaps, would be more likely to occur than the numerical (since no numbers are present in the testing situation).

For the West Point Cadets and white cadets, the loadings in the number factor stand in exceedingly close agreement for all tests. When allowance is made for the presence of a reasoning component as well as a numerical component in the loadings listed for Negroes, the overall agreement is close.

With respect to the pilot criterion the contribution of the number factor is zero, or slightly less than zero, for all three groups. A part of the negative loading for the Negro group is probably made up of reasoning variance which, as will be seen subsequently, is loaded about zero in the pilot criterion for the other two populations.

c. Reasoning factor. Previous analyses by psychological research units of the AAF revealed three reasoning factors with the degree of correlation among them assumed to be zero, although possibilities of absorption are numerous. Reasoning I has been called "general reasoning." The other two forms of reasoning are even more obscure. Highest loading for reasoning II appeared in tests in which comprehension of analogies is apparently required. An interpretation of reasoning III will not be undertaken. That the factorial composition of reasoning has not been too clearly defined by factor analysts is further indicated by the lack of agreement of the AAF results with Thurstone's distinction between deductive reasoning and inductive reasoning.

The reasoning factor appearing for the West Point Cadets seems to resemble that of general reasoning in the white cadet population. The factor weights in the reasoning factor for the two groups stand in relatively close agreement. Of course, direct comparisons of the loadings in the reasoning factor of these two groups with those of the bifurcated (reasoning-number) factor of the Negroes are impossible. If some of the number variance can be discounted, the agreement is fairly close with the possible exception of the test of *Mechanical Principles,* CI903B. The factor weights of the tests most heavily loaded in the reasoning factor for the three groups are presented in Table 8.

The appearance of the factor in such tests as *Mathematics B,* suggests that the factor may be interpreted as an examinee's ability to relate the essential properties, characteristics, or requirements of a problem into that unique combination or pattern of steps necessary for its solution. In common sense terms, it resembles that form of scholastic aptitude required for successful progress in the solution of word problems encountered in intermediate algebra. It may even be another term for insight.

Comparison of the factor weights reveals several slight differences for the groups with respect to the tests: *Mechanical Principles* (both forms); *Reading Comprehension,* and *Mechanical Information.* The somewhat higher loadings of these three tests for the West Point group stand in essential agreement with the higher weight for the reasoning factor in the pilot criterion. The inference to be made is that the West Point Cadets, a highly selected group with respect to academic achievement, tend to use reasoning both in answering test items and in pilot training to a greater extent than do indiduals of the other two groups. Such a tendency appears reasonable. Inasmuch as West Point Cadets represent a group of students of exceptionally high scholastic achievement, and inasmuch as the factor of reasoning is probably only second in importance to the verbal factor in successful college-preparatory and college work, its appearance in more substantial amounts than in the other groups, is indicative of a circumstance of selectivity. In the test of *Mechanical Principles,* and to a considerably lesser degree in the test of *Mechanical Information,* the West Point Cadets who, as a group, throughout childhood and adolescence may have had less interest in mechanical gadgets and a greater interest in symbolic materials, may be able to compensate for their lack of mechanical experience by reasoning out the problems presented them. For example, an examinee whose interest and experience with mechanical devices may be extremely slight, can solve from his knowledge of the principles of high school physics (or mechanics) many items in the test of *Mechanical Principles.* In pilot training, an analogous set of circumstances may be present.

TABLE 8

REASONING-FACTOR LOADINGS IN DEFINITIVE TESTS AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION-CADETS[a]

Test and Code Number		Group I	Group II	Group III
Mathematics B (Arithmetic Reasoning)	CI206C	513	628	47
Mechanical Principles	CI903A		081	34
Mechanical Principles	CI903B	404		34
Reading Comprehension	CI614H	305	285	19
Mechanical Information	CI905B	219		01
Spatial Orientation II	CP503B	181	081	17
Numerical Operations—Back	CI702B	217		11
Mathematics A	CI702F		551	24
Dial and Table Reading	CP622-21A	076	397	16
Discrimination Reaction Time	CP611D	200	278	11
Pilot Criterion		075	−060	00

[a] The loadings for Negro cadets are for a general intellectual factor consisting of number and reasoning variance.

The greater amount of loading in the reasoning factor in the test of *Discrimination Reaction Time,* may again merely represent the facility of West Point Cadets to employ reasoning in unfamiliar tasks. The largest proportion of the reasoning variance, which is only .040, may be apparent at the earlier stages of learning for the West Point Group and not in the later series of trials. In other words, the contribution of reasoning to the variance may be so substantial in early trials that a loading as high as .200 must appear. Unfortunately, no comparisons may be made with the Negro group, since the number variances cannot be partialled out of the loading .278.

As mentioned in a preceding paragraph, the validity of the reasoning factor appears to exist in a small degree for the West Point Cadets, whereas it does not for the other groups. In view of the presence of sampling errors, the loadings in the reasoning factor in the criterion, all absolutely less than .075, should be interpreted with caution. It may be concluded, however, that reasoning variance along with that of the other intellectual factors does not contribute materially to the validity of the test battery for pilot populations.

2. *The Factors of Perceptual Speed and Spatial Relations*

a. Perceptual speed. Another one of Thurstone's primary mental abilities, the factor of perceptual speed, has been investigated further by Thurstone. (6) In his exhaustive studies of perception not only perceptual speed reappeared, but also several other non-orthogonal factors (sub-factors of the super-factor of perception) among which were "the ability to form a perceptual closure against some distraction," (6, p. 101), a common factor for optical illusions, a factor identified to represent reaction time, a factor involving alternation effects (such as their rates in ambiguous figures), a factor "concerned with the manipulation of two configurations simultaneously or in succession . . .," (6, pp. 110-111), the reappearance of the factor of perceptual speed, and several other less clearly defined factors.

The factor of perception revealed in the two factor analyses of the November, 1943, battery and the September, 1944, battery has been identified as perceptual speed, although one or more of the non-orthogonal factors uncovered by Thurstone may be present. Guilford (9) and Guilford and Zimmerman (11) have pointed out that the factor of perceptual speed is in evidence when rapid comparisons of small, detailed, or complex visual figures and visual forms are made and when accurate discriminations of the similarities and differences in them are required. The factor of perceptual speed is likely to be in evidence "when in a multiple-choice response the correct figure is hard to discriminate from among its distractors, the rule or principle having been easy to apprehend. . . ." (9, p. 390).

The descriptions of the three tests: *Spatial Orientation I, Spatial Orientation II* (see II, C, 1, a) and especially *Speed of Identification* (see II, C, 2), would lead one to believe that the perceptual speed factor should be heavily loaded in these tests. The expectation that the three tests should be highly saturated in the perceptual speed factor is supported by the high factor loadings in the three tests. The weights of these three tests in the factor identified as perceptual speed, along with those in eleven other tests, are listed in Table 9.

The high loadings are comparable, test by test, for the three groups. In the test *Dial and Table Reading,* the comparable loadings appear to be due to the common need for all examinees to pick out quickly the relevant number among a conglomeration of others in one table, and to select in other tables from several detailed, complex, and similar-appearing numerical readings the correct one.

For the three tests of instrument comprehension a certain discrimination in the selection of correct figures (or of the correct verbal description in *Instrument Comprehension I*, CI615B) corresponding to instrumental readings appears to be necessary. This perceptual operation, however, does not seem to be the major function required for successful performance upon this test. In general, the higher loadings in these tests are for the spatial-relations factor. An exception to this general result occurs in the instance of a weight of .374 in the perceptual factor (and a loading of .369 in spatial-relations factor) for *Instrument Comprehension II*, CI616B, administered to Negroes. A possible hypothesis for this occurrence is that the factor to be identified as kinesthesis with a loading of .238 minimizes in part the rôle played by the spatial-relations factor, and that the function of perception (in combination with kinesthesis) is maximized. As will be pointed out subsequently, the presence of a loading in kinesthesis in pencil-and-paper tests may represent a projection of the kinesthetic factor—perhaps a factor of empathy.

In the test of *Discrimination Reaction Time*, slight loadings, of approximately the same size for the three groups, are present in the perceptual speed factor. These factor-weights may actually be related to the factor of perception, reaction time, which Thurstone identified. Inasmuch as it is necessary for the examinee to make a rapid sensory discrimination among lights spaced alike from one test trial to the next in their arrangement on the panel, but heterogeneous with respect to color (white, red and green), and inasmuch as it is necessary for the examinee to select one among four switches identical in appearance, the presence of a perceptual speed factor should be expected. However, other factors are probably apparent such as spatial-relations and psychomotor coordination. It would seem that Thurstone's reaction time designation is somewhat too inclusive a term.

TABLE 9

PERCEPTUAL-FACTOR LOADINGS IN DEFINITIVE TESTS AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION-CADETS

Test and Code Number		Group		
		I	II	III
Speed of Identification	CP610A	627		64
Spatial Orientation I	CP501B	627	519	62
Spatial Orientation II	CP503B	549	551	54
Dial and Table Reading	CP622-21A	262	208	31
Instrument Comprehension	CI616C	230		20
Instrument Comprehension I	CI615B		144	18
Instrument Comprehension II	CI616B		374	17
Discrimination Reaction Time	CP611D	231	260	22
Finger Dexterity	CP116A	188	130	20
Complex Coordination	CM701A	227	−054	20
Biographical Data—Navigator	CE602D	086	234	10
Biographical Data—Pilot	CE602D	173	−023	14
Pilot Criterion		−040	260	15

Although the factors of spatial-relations and psychomotor coordination are to be considered individually, mention of them at this point is appropriate with respect to the test of *Discrimination Reaction Time.* Moreover, it will be helpful at this point to distinguish, if possible, between the factors of perceptual speed, spatial relations, and visualization, inasmuch as the test of *Discrimination Reaction Time* illustrates rather well the differences between the psychological interpretations of these factors. Classification of these terms at this point will tend to minimize to an extent, possible confusion in subsequent sections.

The decision as to whether to throw the left-hand or right-hand switch (in his motor choice) is thought to describe the examinee's ability to perceive the spatial order among the switches. Hence, the appraisal of relationships between the switches in their spatial arrangement in this test may be taken as an interpretative definition of the construct *spatial relations.* Prior to the choice decision (as to which switch to throw), a discrimination is made as to the position of the red light with respect to the green light (left or right, above or below). The perception of the spatial relationships or order among the lights furnishes the mental set required for the decision. Hence, another feature of the spatial relationships is that on the sensory side. Of course, a complete distinction between the sensory and motor choice is impossible. For the three groups of West Point cadets, Negroes, and white cadets, loadings in the spatial relations factor for this test are .256, .400, and .420 respectively.

Closely related to space is visualization. The independence of these two factors Thurstone has not distinguished. Factorial analyses by psychologists of the AAF did appear to separate space from visualization and to discredit the term spatial-visualization. The sensory discrimination in the test of *Discrimination Reaction Time,* might possibly be considered visualization, just as the motor, or choice, reaction might be interpreted as psychomotor coordination. To a degree, such a distinction may be valid. However, in the definition rendered by Guilford and Zimmerman, a new position of the stimulus objects must result following (its mental) manipulation, rotation, rearrangement, or inversion:

> The tests most heavily saturated with it all seem to involve a visual manipulative ability. In solving the problems it is necessary mentally to move, turn, twist, or rotate an object or objects and to recognize a new appearance or position after the prescribed manipulation has been performed. (11, p. 157)

In the instance of the test of *Discrimination Reaction Time,* the rôle of visualization may exist, but the stimuli do not appear to be in the need of manipulation suggested by the definition. Examples of visualization tests would include those of counting blocks, and of mechanical principles (illustrated by a test of the Bennett type).

Zimmerman[4] has suggested that a test may be either spatial or visual for an examinee, depending upon the activities involved. If, in a test of flags the subject is able, so to speak, to pick up the flag, move it, turn it about as if he actually had a model in his hands, then visualization is dominant. On the other hand, if the examinee has to move himself to a different position, as in cocking his head to one side or "standing upon his head," then a spatial factor is involved.

For neither of the two test batteries

---

[4] In a personal communication to the writer, Zimmerman has proposed this hypothesis.

did the visualization factor emerge in the analyses. It is possible that a considerable portion of its variance was absorbed by the factor identified as space—especially in the instance of the test, *Mechanical Principles,* CI903B. For the test of *Discrimination Reaction Time,* the loading in the visualization factor reported in (24) is .200.

The presence of the perceptual speed factor in small amounts in the two tests, *Complex Coordination,* and *Finger Dexterity,* may be explained in terms of discriminations made among the detailed stimuli in the respective patterns in the visual field. In the former test the discrimination between the two different colors of light may account in part for the small loading for the West Point Cadets and white cadets. (Distinctions in the positions of the lights, of course, would be more nearly related to the spatial factor.) The negative loading for Negroes cannot be satisfactorily rationalized, especially in view of the fact that the perceptual factor is slightly more dominant for this group in the other tests than it is for either of the other two groups. In the latter test, during the rapid fitting of pegs into the holes, the perception of holes in the board in relation to the background about them, may account for the variance in this factor. On the other hand, speed rather than perception may be what the factor weight represents. For the three groups, the loadings are comparable.

Loadings in the test *Biographical Data,* are with the exception of the Negro group in the navigator form, insignificant. In the test the responses to several items are repetitious in their wording, with only one word being different in each of several responses consisting of many words. Under the conditions of testing, the selection of the appropriate answer among so many similar appearing clusters or groups of items may account for the small loading in perceptual speed. A more likely explanation may be that the scoring weights employed artificially induced the appearance of this factor for Negro pilots.

Whereas the factor weights in perceptual speed in the tests for the three groups are roughly the same, considerable differences exist in the pilot criterion. Next to the factor called kinesthesis that of perceptual speed is the most valid one for Negroes with its loading of .260 in the pilot criterion. For the West Point Cadets the factor weight actually is negative (—.040). Between these two extremes falls the loading of .150 for the white cadet population. Inasmuch as the loadings in the tests do not differ noticeably, the only place for speculation as to what could account for the difference is in the pilot-training activity itself. It is possible that the types of teaching methods employed at the various training units might account, to a degree, for the discrepancies. Similarities in the perceptual factor weights of the tests would argue against differences in environmental influences of a perceptual nature, as sampled by the test items. It is possible that previous experiences in perceptual activities not covered by the test items may be important in pilot training.

Some light may be shed upon the problem in terms of the loading in the pilot criterion in the spatial relations factor. In this instance the rôles of Negroes and West Point Cadets are reversed from what they are in the perceptual factor. For West Point Cadets, Negroes, and white cadets the factor weights are .415, .190, and .320 respectively.* Although it is possible that variances may have been absorbed from one factor to another, or that any variances associated with visualization may have been distributed by chance in a manner to accentuate the differences in variances between the factors of perceptual speed and spatial relations, a reasonable conclusion is that the spatial relations factor actually is *substantially* valid for West Point Cadets and perhaps *negligibly* valid for Negroes and that the perceptual-speed factor actually is *slightly* valid for Negroes and perhaps *negligibly* invalid for West Point Cadets.

b. Spatial relations. In the previous section (I, A, 2, a) concerning the factor of perceptual speed, an attempt was made to distinguish between space and visualization. During the course of the war, AAF psychologists tentatively identified three factors of space as well as the factor of visualization. Only the first space factor, spatial relations, has clearly appeared in most analyses. An hypothesis that kinesthetic imagery is present in some tests has been proposed for one of the space factors (11). At a subsequent point support for this hypothesis will be indicated. Guilford and Zimmerman have defined the factor and have indicated its importance as follows:

> Space I seems to be an ability to perceive the spatial order or the relationships among objects. In several psychomotor tests, the decision of the examinee as to which way to move—right or left, up or down, forward or backward—depended on a correct appraisal of the stimulus arrangement. This ability probably outweighed all others in the pilot criterion. It may be a prominent requirement in any machine-operating job that requires decisions as to direction of movement dependent on signals. (11, p. 157)

The consistent appearance of the spatial relations factor in both pencil-and-paper tests and psychomotor tests in previous analyses is an important discovery with many practical implications.

High loadings in the spatial relations factor for two groups of West Point Cadets and Negroes are in essential agreement with those of the representative white cadet population derived from previous analyses. In the accompanying Table 10 in which loadings in the spatial-relations factor in sixteen tests are given, a reasonable degree of similarity of the weights is present for the three populations.

Moderate to high loadings in the spatial relations factor for the test of *Complex Coordination,* are to be expected from the nature of the task (see description of the test in II, C, 1, b). The presence of a spatial relations factor is clearly indicated by the decisions required by the examinee as to whether to move the stick forward or backward, left or right, in conformity to the pattern of the two green lights in their (spatial) relation to the two red lights and as to how to move the rudder control into the required position indicated by the relation of a third green light to a third red light. Of course, the presence of some weight in a psychomotor coordination factor should be expected in that movements of the arms and hands effect the adjustment to the visual cues. In fact, a loading of .473 for West Point Cadets and a loading of .495 for Negroes did result in the psychomotor factor.

Another apparatus test which aids materially in the identification of the factor of space is *Two-Hand Coordination.* Like the test of *Complex Coordination,* it is loaded about as much in psychomotor coordination as in space for all three groups. A review of the description of this *Two-Hand Coordination* test (see II, C, 1, b) should indicate that the decision to use either the right hand or left hand in either right or left movements is a choice reaction which should be represented by the spatial relations factor. That some reasoning variance might be expected in the early stages of learning is indicated by small loadings of .123 for West Point Cadets and of .156 for Negroes (which of course may include number variance), although for the white cadet population, a loading of zero exists.

A third apparatus test, the factorial complexity of which with respect to perceptual, spatial, and psychomotor factors was revealed in several previous analyses (24), is *Discrimination Reaction Time,* loaded .400 and .420 in the spatial factor for Negroes and white cadets respectively. It is weighted only .256 for West Point Cadets. This loading differs sufficiently from that for the two other populations that an explanation is required.

What are some possible reasons to account for loading of only .256 for West Point Cadets? First of all, in view of the low communality of .336, a weight of .351 for the West Point group in the psychomotor coordination factor may account in part for the lower loading in the spatial relations factor (despite the apparent validity of this factor indicated by the presence of a substantial weight of .435 in the pilot criterion). That the residual factor with a loading .130 in this test may represent a portion of the variance which has split off from the spatial factor is one possible explanation. If another factor (not representing mere error variance) could be extracted in order to permit further rotation, the residual factor might emerge as a second space factor, or even as a visualization factor. Of course, it may be true that the factor of psychomotor coordination for West Point Cadets actually assumes a more important rôle than the space fac-

TABLE 10

SPATIAL-RELATIONS FACTOR LOADINGS IN DEFINITIVE TESTS AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION-CADETS

Test and Code Number		Group I	Group II	Group III
Complex Coordination	CM701A	453	503	49
Two-Hand Coordination	CM101A	463	425	41
Discrimination Reaction Time	CP611D	256	400	42
Instrument Comprehension	CI616C	498		41
Instrument Comprehension I	CI615B		512	44
Instrument Comprehension II	CI616B		369	53
Dial and Table Reading	CP622-21A	307	487	42
Mechanical Principles	CI903A		061	12
Mechanical Principles	CI903B	358		12
Spatial Orientation I	CP501B	214	280	10
Spatial Orientation II	CP503B	219	219	16
General Information	CE505E		207	23
General Information	CE505F	180		10
Mechanical Information	CI905B	185		02
Rudder Control	CM120B	496	020	13
Rotary Pursuit	CP410B	259	095	14
Finger Dexterity	CP116A	069	273	12
Pilot Criterion		415	190	32

tor. Evidence for the dominance of the psychomotor factor over the others in apparatus tests administered to West Point Cadets is available in the presence of its loadings in five out of the six of these tests higher than those for the white cadet population.

Finally, practice effects in the training received with the four preceding psychomotor tests of *Rotary Pursuit, Two-Hand Coordination, Complex Coordination,* and *Rudder Control,* may have rendered psychomotor coordination to be more important and spatial relations less significant in performance of the tests. In other words, the relative contribution of the visual cues to the execution of a quick motor response is far less important than the coordination involved. Stated in still another manner, at their stage of learning, the West Point Cadets, having obtained maximum proficiency in use of visual cues, performed successfully in reaction-time situation primarily to the extent that they could effect a motor response smoothly.

In the three other apparatus tests, several differences are present in the loadings in the space factor. Among the most marked dissimilarities in factor weights are those in the *Rudder Control Test:* loadings of .496 for West Point Cadets, of .020 for Negroes, and of .130 for white cadets. Since most of the common-factor variance for the Negro group is accounted for by the saturation of the test in the factor identified as kinesthesis, which may well be the name for a second space factor, the loading of .020 in the spatial relations factor is not too surprising. Moreover, the weight in the psychomotor factor is moderate (.311). Whether a second space factor might be brought out for West Point Cadets, through the type of rotation suggested in the next-to-the-last paragraph is problematical. It does seem reasonable to believe that the loading of .496 embraces more than spatial-relations variance, especially when the weight is compared with that of .130 for white cadets. That the field of visual cues in the rudder control task is less restricted than those fields of visual cues in tests of reaction time and coordination might account for the negligible weight of .130 in the white cadet population.

Somewhat the same set of circumstances as appeared in the *Rudder Control* test is present on a smaller scale in the test *Rotary Pursuit.* For the two groups of Negroes and West Point Cadets, the factor weights of .259 and .095 are proportional respectively to those obtained in the test of *Rudder Control.* Another parallel is the slight weight of .186 in the kinesthetic factor (for test of *Rudder Control*). The sweeping circular movements of the arms may account in part for this small loading. Primarily the test is one of eye-hand coordination—a fact well substantiated by the high loadings in the psychomotor-coordination factor for all three populations. What variance in the spatial-relations factor does appear may likely be attributed to the decision as to which key to depress in the distraction task.

In the sixth psychomotor test *Finger Dexterity,* the pattern of loadings in the spatial-relations factor .069, .273, and .120 for the three groups (West Point Cadets, Negroes, and white cadets respectively) resembles that of .256, .400, and .420 for the test *Discrimination Reaction Time.* Inasmuch as the *Finger Dexterity* test is the last to be administered, the same hypothesis of practice effects suggested to explain the low loading of .256 in the test of *Discrimination Reaction Time,* may account for the weight of .069 for West Point Cadets. At the particular level of learning of the West Point group, improvement in performance may depend almost entirely upon the physiological limits set by neuro-muscular mechanism. Further evidence that primarily a psychomotor function is involved is the relatively high degree of purity of the test of *Finger Dexterity.*

Among the numerous pencil-and-paper tests loaded in the spatial-relations factor, the most definitive are the three forms of the test of instrument comprehension. With the exception of the loading of .369 for the Negro group in *Instrument Comprehension II,* the factor weights are comparable among all forms for the three populations.

That the spatial-relations factor can be as well represented in pencil-and-paper tests as in apparatus tests is highly significant. If, for example, success in a given vocational task demands primarily a spatial-relations factor, the need for costly and time-consuming apparatus tests is minimized. Through symbolic representation of the possible positions of planes corresponding to instrument readings, the examinee is forced to make a decision as to whether the plane has banked left or right and has moved up or down (has climbed or has descended). Although these tests might be expected to emphasize intellectual factors, their relatively high degree of purity with respect to the spatial relations factor indicates that they successfully measure with symbolic material one of the same factors as do the psychomotor tests (without the presence of large amounts of other variance found in the apparatus tests).

In the factorially complex test, *Dial and Table Reading,* the moderate and substantial loadings obtained in the spatial relations factor for the three groups are somewhat difficult to rationalize. To a small degree, visualization variance may be present, although the format of the dials would not seem to require a considerable amount of the so-called mental-manipulation or rotation of the perceived objects. To some extent the interpolation process required in reading the dials may account for the loadings, inasmuch as the position of the needle stands in spatial relation to the numbers on the dial to its left and to its right. In the decision as to the reading represented, an oscillation in the perception between the numbers at the left and at the right, or above and below, is necessary. Similarly the part of the test upon table reading requires both "right-left" and "top-bottom" points of orientation to guide in the selection of the relevant number (from the conglomeration of others).

That the factorial composition of the test may be a function of the intellectual factors entering into scholastic aptitude is indicated by the lower loading (.307) in spatial relations factor for the West Point Cadets. Higher weights (.587 and .237) in the numerical and verbal factors than those of .530 and .100 for white cadets may reveal that the importance of the spatial relations factor in the test is reduced; first, because of the previous amount of experience of the West Point Cadets with tables and quantitative data encountered in high-school and college mathematics-and-science courses and, second, because of the tendency of academically trained people to verbalize. (The verbal loading for Negroes of .451 was rationalized in terms of the requirement of a reading level of relatively low difficulty in the lengthy directions.)

For the test of *Mechanical Principles,* forms CI903A and CI903B, dissimilarities in the weights (.358, .061, and .200 for three groups) furnish an opportunity for several hypotheses. The factor weight of .358 for West Point Cadets is further evidence of the possible complexity of the space factor extracted for this group. Inasmuch as previous analyses have revealed a loading as high as .54 in the visualization factor for this test, and inasmuch as visualization might be expected to play an important rôle in any science or engineering curriculum, the inflation of the loading in the spatial relations factor (through the presence of visualization variance) is a reasonable occurrence. For the Negro group, a negligible weight of .165 in kinesthesis furnishes a possible clue to the presence of another source of variance in the spatial factor for the West Point group. For the Negro group, the appearance of the kinesthetic factor may account in part for the small loading of .061 in the spatial relations factor. What little variance does appear in the space factor for the test of *Mechanical Principles,* CI903B is highly suggestive that the hypothesis proposed by Guilford and Zimmerman for interpretation of a spatial relations factor is a meaningful one, inasmuch as a decision of left-right, up-down, or forward-backward movement is lacking for most items. Considerable rotation, manipulation, and inversion of the visual images seem to be demanded.

In the two tests, *Spatial Orientation I* and *Spatial Orientation II,* comparable factor weights in spatial relations are present for the West Point Cadets and Negroes. Loadings in the vicinity of .200 are somewhat lower than what the "armchair" estimates of a person looking at this test might be. Primarily, the test is perceptual in line with the previously mentioned hypotheses set up by Guilford (9, p. 390) for the appearance of a perceptual speed factor. Study of the test does indicate that a decision of the directions of movement is lacking, although the stimuli in the aerial photographs and maps do stand in spatial relation to one another. It is quite possible that if most of the intricate detail were eliminated and if only a few clearly indicated landmarks (symbols) stood in relationship to one another, the factorial composition would be considerably altered in the direction of larger loadings in the spatial relations factor.

In the three tests of information, *General Information,* CE505E, *General Information,* CE505F, and *Mechanical Information,* CI905B, negligible and slight loadings are present in the spatial relations factor. In these tests, the factor may function vicariously in the verbal items. In other words, some of the responses to items may depend to a slight degree upon the examinee's reenacting so to speak, an experience suggested by the item. An example of such a possibility is furnished by the following item taken from the *General Information Test,* CE505F:

In making a turn to the left, pressure is applied:
A. first with the rudder and then with the stick,
B. first with the stick and then with the rudder,
C. with the stick and rudder at the same time,
D. with the rudder alone,
E. don't know.

The validity of the spatial relations factor is indicated by the loadings of .415, .190 and .320 in the pilot criterion for the three groups of West Point Cadets, Negroes, and white cadets. That components of visualization, and even kinesthesis, may be present in the loading of .415 for West Point Cadets is possible. The somewhat lower weight for the Negroes may be attributed to the loading of .376 in the kinesthesis factor. Experience of AAF research psychologists tended to indicate that the spatial relations factor is one of the most valid factors, if not the most valid factor, for prediction of pilot success. The spatial-factor loadings in the pilot criterion for the West Point Cadets and Negroes are consistent with the results of previous analyses.

3. *The Factors of Mechanical Experience and Pilot Interest*

In tests of the informational and biographical type, the factors of mechanical experience and pilot interest are well-defined. For convenience, these two factors are considered together, inasmuch as several tests are complex with respect to them.

a. Mechanical experience factor. In previous analyses the factor of mechanical experience has appeared repeatedly in tests consisting of mechanical content. The examinee's familiarity with automobile parts, shop tools, machinery, and miscellaneous gadgets apparently accounts for high loadings in tests made up of such informational items. Although most clearly defined for the three populations in the six tests: *Mechanical Information,* CI905B, *Mechanical Principles,* CI903A, *Mechanical Principles,* CI903B, *General Information,* CE505E, *General Information,* CE505F, and *Biographical Data—Pilot,* CE602D, the factor of mechanical experience is present to a small degree in other pencil-and-paper tests, in one psychomotor test, and to a doubtful degree in a second psychomotor test, as indicated by the loadings for each of the three populations in Table 11.

For all three groups, the factor weights in the first six tests in Table 11 stand in close agreement with two exceptions. The high loading of .580 for the West Point Cadets in the test *General Information,* CE505F, is due to the existence of a spurious correlation of this form with another test, *Mechanical Information,* CI905B, with which it has several items in com-

TABLE 11

MECHANICAL-EXPERIENCE FACTOR LOADINGS IN DEFINITIVE TESTS AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION-CADETS

Test and Code Number		Group		
		I	II	III
Mechanical Information	CI905B	699		64
Mechanical Principles	CI903A		543	58
Mechanical Principles	CI903B	550		58
General Information	CE505F	580		34
General Information	CE505E		528	53
Biographical Data—Pilot	CE602D	441	264	50
Instrument Comprehension II	CI616B		372	03
Instrument Comprehension	CI616C	105		14
Two-Hand Coordination	CM101A	328	160	40
Spatial Orientation II	CP503B	223	325	15
Practical Judgment	CI301C	298		12
Reading Comprehension	CI614H	261	313	04
Rudder Control	CM120B	218	146	01
Pilot Criterion		040	130	27

mon. As for the small loading of .264 in *Biographical Data—Pilot,* administered to Negroes, an hypothesis is difficult to formulate. One possibility to account for this fact is that the scoring key placed, for Negroes, a disproportionate amount of emphasis upon items concerned with scholastic interests, hobbies, and other activities somewhat more common to the culture of the West Point Cadets and white cadets than to the culture of the Negroes. Many items given a positive weight are concerned with curricular and extracurricular activities of the school in which many a Negro may never have had an opportunity to participate.

Small positive loadings in the mechanical experience factor for other pencil-and-paper tests, with the exception of *Spatial Orientation II,* may be accounted for by the presence of items of mechanical content. In the test of *Reading Comprehension,* for example, one selection is concerned with principles of forces in mechanics; whereas another is built about a familiar gadget, the compass.

For the test of *Spatial Orientation II,* some of the variance may represent that of another sort of background factor—possibly experience with maps, figures, drawings, blueprints, and so forth. It might be expected, however, that the loading (.325) for Negroes should be somewhat lower than those loadings (.223 and .150) for West Point Cadets and white cadets, in view of the probably greater emphasis in map-reading and drawing for the latter two groups in their formal education. No reasonably satisfactory hypothesis for the differences is apparent.

Loadings in the two psychomotor tests, *Two-Hand Coordination,* and *Rudder Control,* may be explained to an extent in terms of a probable positive transfer effect of the examinee's experience with other mechanical devices with which the two apparatus tests have several elements in common. Higher loadings for the West Point group in these two tests may possibly be attributed to their conscious attempts at transfer. In their physical education activities, apparatus resembling these two psychomotor tests might have been available for practice (although no positive information concerning the types of training equipment at the West Point Academy is attainable).

That the mechanical-experience factor shows approximately a zero loading (.040) in the pilot criterion for the West Point Cadets is further evidence of the likelihood of the strong academic interests of the group and of their lack of experience with shop tools and machinery. The mechanical experience factor contributes somewhat to the validity of the respective test batteries for the Negro and white cadet populations.

b. Pilot interest factor. For both the West Point Cadets and Negroes, the test of *Biographical Data—Pilot,* CE602D, makes a substantial and unique contribution to the validity of the respective test batteries. (For further information upon this point, see IV, B, 3.) In the two factor analyses of the November, 1943, and September, 1944, classification batteries high loadings (.531 and .562) in the test of *Biographical Data* scored with pilot weights appeared in a factor which, for the West Point Cadet group, received a loading no higher than .352 in any other test and which, for the Negro group, received a loading of .404 in the navigator form of the *Biographical Data* test and a loading no higher than .252 in any other test. This factor, apparently unique to the test of *Biographical Data,* is tentatively called pilot interest. Loaded .436 in the pilot criterion for West Point Cadets, it proves to be the most valid factor for this group. To what extent the range of pilot interest in the criterion variable was restricted by the presence of only those cadets who elected pilot training cannot be determined. Therefore the loading of .436 may be subject to some degree of error.

The loading in the criterion for the Negro group is small with respect to the pilot interest factor. Apparently the factors of kinesthesis, perception, and spatial relations for Negroes are sufficient to account for most of the common-factor variance in the pilot criterion. It may be that in terms of the intensity of motivation the interest factor was primary for success of the West Point Cadets who ordinarily would elect other activities. For the Negroes, on the other hand, the presence of a moderate degree of interest would result in its subordinate rôle in relation to the other factors for pilot success. The extent to which factors of temperament may have been significant is undoubtedly important, but objective evidence upon this point is lacking.

The negative loadings in the test, *Biographical Data,* CE602D, for the white cadet group are unexplainable. It may be that the factor identified as mathematical background in previous analyses may have absorbed most of the variance of the pilot-interest factor. A question might be raised as to the accuracy of identification of the factor of pilot interest in previous analyses. Since no entries in this factor are present in the *Composite Factor Analysis Summary* for the four pencil-and-paper tests included in either the November, 1943, battery or the September, 1944, battery, it may be that tests were not present in those earlier matrices which could aid in the identification of the factor. The small weights in the factor of pilot-interest for the four non-biographical tests in Table 12 aid in its identification, inasmuch as the items of these pencil-and-paper tests suggest an interest in flying.

TABLE 12

PILOT-INTEREST FACTOR LOADINGS IN INFINITIVE TESTS AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION-CADETS

Test and Code Number		Group		
		I	II	III
Biographical Data—Pilot	CE602D	531	562	−09
Biographical Data—Navigator	CE602D	250	404	−14
General Information	CE505E		226	...
General Information	CE505F	352		...
Rotary Pursuit	CP410B	156	252	...
Instrument Comprehension	CI616B		000	
Instrument Comprehension	CI616C	284		...
Pilot Criterion		436	060	250

4. *The Factor of Psychomotor Coordination*

Apparently measurable only by apparatus tests, the factor of psychomotor-coordination has proved to be valid for pilots, as one might expect. Psychomotor tests do possess considerable so-called face validity in that they mimic overt operations involved in flying. Close agreement in three populations among the factor weights for the several apparatus tests is apparent in Table 13. Their factorial composition is complex for all three groups.

In previous sections concerned with intellectual factors and with those of perception and spatial relations, hypotheses have been suggested to account for the loadings of the apparatus tests in factors other than psychomotor coordination. Hence, no further consideration of the factorial pattern will be undertaken.

TABLE 13

PSYCHOMOTOR-COORDINATION FACTOR LOADINGS IN DEFINITIVE TESTS AND IN RESPECTIVE PILOT CRITERIA IN THE THREE POPULATIONS OF WEST POINT, NEGRO, AND WHITE AVIATION-CADETS

Test and Code Number		Group		
		I	II	III
Complex Coordination	CM701A	473	495	40
Rotary Pursuit	CP410B	548	497	53
Rudder Control	CM120B	406	311	48
Two-Hand Coordination	CM101A	404	533	34
Finger Dexterity	CM116A	508	454	34
Discrimination Reaction Time	CP611D	351	158	12
Biographical Data—Pilot	CE602D	075	176	22
Mechanical Principles	CI003A		216	18
Mechanical Principles	CI903B	094		18
Pilot Criterion		305	050	20

The presence of small factor weights in the paper-and-pencil tests of *Biographical Data—Pilot,* CE602D, and *Mechanical Principles,* CI903B, may well be a vicarious functioning of the psychomotor factor. Possibly, ideomotor responses accompany the reactions of some examinees to the content of certain items—content which suggests to the examinee manipulation of mechanical devices.

In the pilot criterion are noticeable dissimilarities for the three groups in the loadings in the psychomotor-coordination factor. Why should this factor be valid for the West Point group and white cadet group and not for the Negroes? That the Negroes do possess the factor in substantial amounts is indicated by the weights for it in the apparatus tests. One possible hypothesis, open to serious question, is that from previous training in tasks involving psychomotor coordination, the stage or level of attainment with respect to the factor is so high that it is relatively unimportant in pilot-training in comparison with other factors, kinesthesis, for example. In fact, the presence of too high a degree of psychomotor coordination may tend at the earlier stages of flying to result in the executions of improper movements or, more simply, in the inhibition of learning. Moreover, a certain amount of "unlearning" of psychomotor responses, or perhaps better, relearning, may be necessary to effect maximum use of kinesthetic cues.

Part of the difficulty may be in the definition of psychomotor coordination. Three factors of psychomotor coordination appeared in the results of the AAF psychological program during the course of the War. The first factor is thought to involve the somewhat grosser movements of the trunk and limbs; whereas the second factor is considered to embrace the finer, or more delicate, movements of hand and wrist. Hence, it is tentatively identified as psychomotor precision. The third factor has been called psychomotor speed. If such a differentiation is possible, it may be that different components of the psychomotor factor are apparent for the three groups, or that in the instance of the Negro group too high a degree of psychomotor speed results in unfavorable movements. The tendency to employ gross movements of trunk and limbs may be a handicap. Only more extensive research efforts in this direction can give the answer to the disconcerting picture of the pilot-criterion.

5. *The Factor of Kinesthesis*

For the Negro group only, a factor to which reference has been made several times previously appeared with a heavy loading (.554) in the *Rudder Control* test and with a large weight (.500) in the pilot criterion. For the Negro cadets, this factor viewed with respect to loadings of other factors in the criterion is obviously the most valid. Toward the close of the War, what is evidently the same factor emerged in another analysis. Inasmuch as the loading in psychomotor coordination factor is moderate (.311) for the *Rudder Control* test, this factor can not readily be interpreted as one of psychomotor coordination. Presence of slight weights in other tests believed to contain variance in space suggests that the factor is one of space.

Selection of the name kinesthesis fits reasonably well the description of the movements required by the examinee in his attempt to return the rudder-control device to a straight-ahead position. As an examinee participates in the rudder-control test, the sense of movement (a change of position) is conveyed to the higher neural processes from sensory stimulation received in muscles, tendons, and joints. In the psychomotor tests, *Rotary Pursuit* and *Two-Hand Coordination,* these same kinesthetic cues are probably present to a minor degree.

In the pencil-and-paper test *Instrument Comprehension II*, the presence of a loading of .238 suggests that various positions of planes shown in the multiple-choice responses have taken over in part several aspects of the situation encountered in the *Rudder Control* test. This apparently non-voluntary projection of the kinesthetic experience of the apparatus test into the pencil-and-paper test may be called *kinesthetic empathy.*

Somewhat disconcerning is the weight of .225 in the test, *Mathematics B.* Inspection of the items of the test does reveal in the word problems the representation of situations in which kinesthetic cues be involved. Differential velocities in the word problems and frequent statements of direction may give rise to involuntary and largely subliminal neural excitations which are conditioned to the more active phases of a kinesthetic experience. Let it be said, however, that the preceding statement, like many others, is merely an hypothesis subject to modification and possible rejection in light of any other evidence which may be forthcoming. Never to be minimized is the substantial rôle which sampling errors can play in the magnitude of small factor weights.

In Table 14 are the principal tests in which the weights in this new kinesthetic factor appears for Negroes.

TABLE 14

KINESTHESIS-FACTOR LOADINGS IN DEFINITIVE TESTS AND IN THE PILOT CRITERION OF NEGRO CADETS[a]

Test and Code Number		Group II
Rudder Control	CM120B	554
Instrument Comprehension II	CI616B	238
Mathematics B	CI206C	225
Rotary Pursuit	CP410B	186
Two-Hand Coordination	CM101A	173
Mechanical Principles	CI903A	165
Pilot Criterion		500

[a] The kinesthesis factor appeared only in the analysis of the intercorrelations of tests in the battery administered to Negroes.

B. An Interpretive Comparison of the Traditional Multiple-Regression and Factorial Approaches to the Prediction of a Criterion

The traditional procedure for the determination of the validity of a test battery has been that of the multiple-regression equation in which the maximum possible validity (multiple correlation) between the criterion variable and the tests has been achieved through optimal weighting of each test. Although mathematically sound, this approach in the hands of many a technician has led to an unnecessary duplication of coverage in several tests in his attempt to maximize the validity of each test of the battery. As a rule, factorially complex tests are the result. On the other hand, factor-analysis techniques do permit an improved control over a test battery in that; with the factorial structure of the criterion known, the technician is able to construct a few relatively pure tests, each one of which makes a unique contribution to the validity of the battery. Although each test individually may contribute only moderately to the validity of the battery, the combined contribution of only a few such relatively pure tests can exceed considerably that of double the number tests which have been selected by the traditional multiple-regression practices. In conjunction with factor-analysis procedures the multiple-regression technique, however, is a valuable supplement in that it shows: first, how the test maker should optimally weight each test for maximum validity and, second, when this is done, how much variance each test contributes to the coefficient of multiple determination (the multiple correlation, or validity, coefficient squared). In the following paragraphs an attempt is made to show the complementary nature of the techniques of the multiple-regression equation and factor analysis.

As the factorial composition of a criterion becomes better known, the maximum possible validity of the test battery is enhanced. In other words, the sum of the common-factor variance in the criterion represents the maximum possible amount of variance which these factors, optimally weighted in a test composite, can yield. As the reliability of tests and their relative degree of purity in the relevant factors are increased, the more closely is the maximum potential validity approached. Of course, the addition of any tests to a battery which may aid in the identification of other factors in a criterion and in the determination of the amounts of variance of these factors merely raises the ceiling of the validity to a new maximum.

For the West Point group, the sum of the common-factor variance (communality) in the pilot criterion is .493 for nine centroid factors extracted and .470 after rotation for eight real factors (if the residual factor variance of .014 is discounted and if the discrepancy of .009 between communality after rotation to that before rotation is overlooked). The coefficient of multiple determination R^2, the amount of variance in the criterion associated with or predicted from the most favorable weighting of the tests in the composite is .3610, or about three-fourths of the valid factorial variance in the criterion. Another way of stating this comparison is in terms of the coefficient of multiple correlation, R, the square root of R^2. With respect to the variance in the pilot criterion accounted for by only eight factors, the maximum derivable R is .692; whereas the obtained R is .601.

What are some of the probable reasons which account for the missing fourth of the total potential variance permitted by the factorial composition of the criterion? First of all, to some degree, the lack of reliability of the tests themselves may limit somewhat the contribution of tests to R^2. This argument is difficult either to expand or to defend rigorously. Correction for attenuation of the original intercorrelations among tests would permit the maximum variance associated with true scores to be predicted. On the other hand, corrections could be made for the fallibility of the pilot criterion scores, which are probably less reliable than the test scores. If this was done, the sum of the common factor variance would be considerably augmented—in fact, it is likely that it would increase proportionately more than the true variance of the test scores following their correction for fallibility. Then, perhaps about as much as one-third of the potential factorial variance in the infallible criterion might not be predicted from the optimal weighting of tests (the scores of which are corrected for attenuation).

A second, and more fruitful, approach seems to be a detailed consideration of the factorial composition of each test. Among the tests of the battery, an adequate measure may be lacking for a factor which is highly loaded in the criterion. Moreover, in several tests, substantial positive loadings may be present for factors which are negatively loaded in the criterion. In other words, a given test in some instances may correlate negatively with a criterion. If a test which correlates negatively with the criterion receives a positive beta weight, its variance contribution to the coefficient of multiple determination, which is the product of its beta weight and its coefficient of correlation with the criterion, is negative. Obviously, in this instance pilot-selection is biased in a direction antagonistic to the requirements of the practical situation.

Although detailed consideration subsequently will be given to the contribution of individual tests to R^2, it is illuminating at this point to mention that eight tests of the battery administered to West Point Cadets actually yield negative variances. (See Table 15 for listing of the relative contributions of tests to the total predicted variance R^2 for the November, 1943, and September, 1944, batteries.) In large measure these negative contributions can be explained in terms of the presence of high positive loadings of tests in factors of perception and number which are weighted $-.040$ and $-.053$ in the pilot criterion. The sum of the variance contribution of these eight tests is $-.0490$. For purposes of illustration only, if the factor structure of the criterion and the contributions of the remaining thirteen tests to R^2 could be assumed constant, the elimination of

RELATIVE CONTRIBUTIONS OF TESTS OF THE NOVEMBER, 1943, AND SEPTEMBER, 1944, CLASSIFICATION BATTERIES TO THE PREDICTED VARIANCE R^2 (COEFFICIENT OF MULTIPLE DETERMINATION) OF THEIR RESPECTIVE BATTERIES

Test, Code Number, and Date of Battery		Beta Weight	Biserial Correlation (Validity Coefficient) r_{bis}	Relative Contribution of Test to Predicted Variance R^2	Beta Weight	Biserial Correlation (Validity Coefficient)[a] r_{bis}	Relative Contribution of Test to Predicted Variance R^2
Pencil-and-Paper Tests Common to 1943 and 1944 Batteries				$\beta \cdot r$			$\beta \cdot r$
1. Mathematics B (Arith. Reasoning)	CI206C	011	05	0005	−067	04	−0027
2. Dial and Table Reading	CP622-21A	139	22	0306	103	11	0113
3. Spatial Orientation I	CP501B	−037	12	−0044	052	12	0062
4. Spatial Orientation II	CP503B	−099	10	−0099	−055	14	−0077
5. Biographical Data—Navigator	CE602D	049	07	0034	−091	04	−0036
6. Biographical Data—Pilot	CE602D	176	29	0510	105	11	0115
7. Reading Comprehension	CI614H	−006	11	−0007	016	14	0022
Apparatus Tests Common to 1943 and 1944 Batteries							
8. Rotary Pursuit	CP410B	159	36	0572	010	08	0008
9. Two-Hand Coordination	CM101A	−006	32	−0019	040	08	0032
10. Complex Coordination	CM701A	076	34	0258	−229	−02	0046
11. Rudder Control	CM120B	151	39	0589	185	15	0277
12. Discrimination Reaction Time	CP611D	031	22	0068	105	16	0168
13. Finger Dexterity	CM116A	037	19	0070	−085	00	0000
Pencil-and-Paper Tests Unique to 1943 Battery							
14. Mechanical Principles	CI903A				000	11	0000
15. General Information	CE505E				088	15	0132
16. (General) Mathematics A	CI702F				−110	−02	0022
17. Instrument Comprehension I	CI615B				018	14	0025
18. Instrument Comprehension II	CI616B				287	31	0890
Pencil-and-Paper Tests Unique to 1944 Battery							
19. Numerical Operations—Front	CI702B	−067	01	−0007			
20. Numerical Operations—Back	CI702B	072	09	0065			
21. Practical Judgment	CI301C	−039	11	−0043			
22. General Information	CE505F	003	24	0007			
23. Instrument Comprehension	CI616C	209	36	0752			
24. Mechanical Principles	CI903B	247	35	0864			
25. Mechanical Information	CI905B	−110	18	−0198			
26. Speed of Identification	CP610A	−122	06	−0073			
			$R^2 = \Sigma\beta \cdot r =$	3610		$R^2 = \Sigma\beta \cdot r =$	1772
			$R =$	.601 ($Rc = .587$)		$R =$	.421 ($Rc = .361$)

[a] The validity coefficients have not been corrected for restriction of Range. R_c stands for Ezekiel's correction.

these eight tests would mean that R^2 would be .410—not too much less than the ceiling of .470.

Among six tests yielding negligible, but positive amounts of variance (less than .01) to the coefficient of multiple-determination, a marked degree of complexity is apparent among five which are loaded positively in factors containing both positive and negative weights in the criterion. Hence, further cultivation of these tests through increasing their loadings in valid factors would allow a further narrowing in the difference between the coefficient of multiple-determination, R^2, and the factorial variance in the pilot criterion.

For the Negro group, the sum of the common factor variance for the eight factors in the pilot criterion after final rotation, is only .381 compared with .470, identified by the eight factors in the pilot criterion of the West Point Cadets. In terms of the proportion of the factorial variance of the criterion which the coefficient of multiple-determination represents, the result is somewhat less for the Negro group than for West Point Cadets. The value of .1772 for R^2 is approximately one-half of .381. Considerably less than that for the West Point group, the total amount of negative variance contributed by three tests is -.0140. In part, the negative contributions of these tests are explained by the presence of negative loadings of -.050 and -.060 in the pilot-criterion for the respective factors of psychomotor coordination and reasoning-number which appear in the tests. In part, however, these negative contributions are indicated by negative beta coefficients which are a function of not only the correlation between a test and the pilot criterion, but also the correlation of a test with all other tests in the battery.

However, the need for new tests to aid in the identification of other factors in the pilot-criterion of the Negroes is apparent. As suggested previously, improvements of tests in the battery through an increase in the degree of purity whenever feasible and through the addition of tests which describe better the relevant factor-variance of the pilot-criterion are required. For example, the presence of the factor tentatively identified as kinesthesis in the pilot criterion suggests the importance of cultivation of tests loaded in this factor. In turn, these new tests may furnish additional information concerning the factorial composition of the pilot-criterion for Negroes. Finally, the lack of reliability of the pilot pass-fail criterion itself for the Negro group is probably greater than that for the West Point Cadets in view of the lower communality in the pilot-criteron. In other words, the difference between the communalities of the two criteria of the West Point group and of the Negroes cannot be reasonably attributed just to specific, or unidentified, factor variances.

1. *Relative Contributions of Tests in the Two Batteries to the Predicted Variances*

For each of the tests in the two batteries administered to West Point Cadets and Negroes, a summary in Table 15 is presented of the beta weights, of the biserial correlations of each test with the pilot criterion, and of the relative contribution of each test to the predicted variance. Since the coefficient of multiple-determination R^2 is the sum of the products of the beta weight of each test and of the coefficient of correlation of the test with the criterion variable: $R^2 = \beta_{c1.234} \ldots {}_n r_{c1} + \beta_{c2.134} \ldots {}_n r_{c2} + \ldots + \beta_{ci.12} \ldots {}_j \ldots {}_n r_{ci} + \ldots + \beta_{cn.12} \ldots {}_{n-1} r_{cn}$, $j \neq 1$, the contribution to the predicted variance of each test is clearly its beta weight multiplied by its coefficient of correlation with the criterion:

$\beta_{ci.12} \ldots {}_j \ldots {}_n r_{ci}$
where c = the criterion,
where i = the ith test,
where n = the number of tests,
where r_{ci} = the correlation between the ith test and the criterion,
where β_{ci} = the beta (optimal) weight applied to test in the composite,
where R^2 = the coefficient of multiple-correlation squared (the coefficient of multiple-determination).

Among the tests furnishing the largest amounts of variance (in excess of .0100 to the predicted variance of .3610 (R^2) in the battery administered to the West Point Cadets are the following: *Mechanical Principles,* CI903B, .0864; *Instrument Comprehension,* CI616C, .0752; *Rudder Control,* CM120B, .0589; *Rotary Pursuit,* CP410B, .0572; *Biographical Data—Pilot,* CE602D, .0510; *Dial and Table Reading,* CP622-21A, .0306; and *Complex Coordination,* CM701A, .0258.

Seven tests together contribute a variance of .3851 which exceeds the predicted variance of .3610 derived from the optimal weighting of twenty-one tests in the battery. The negative variances in eight tests account for this rather surprising fact.

With the exception of the tests of *Rotary Pursuit* and *Instrument Comprehension,* inspection of the factor loadings of the other five tests (see Table 5) shows them to be complex. Inasmuch as most loadings in these tests are in factors with high weights in the pilot criterion, the high degree of variance-contribution is not surprising. However, the fact that two of the most nearly pure tests contributed .0752 and .0572 to the total predicted variance indicates that four or five such tests (one for each factor with a substantial loading in the criterion), could jointly yield as much variance as several complex tests which overlap in their function.

For the Negro population the following tests supply the largest amounts of variance to the total predicted variance: *Instrument Comprehension II,* CI616B, .0890; *Rudder Control,* CM120B, .0277; *General Information,* CE505E, .0132; *Biographical Data—Pilot,* CE602D, .0115; and *Dial and Table Reading,* CP622-21A, .0113.

The sum of the variances for these five tests is .1527, an amount slightly less than that of the total predicted variance of .1772. The test in instrument comprehension actually contributes more than half the variance to the coefficient of multiple determination. Although this test is relatively pure for the West Point Cadets, it is for the Negroes, extremely complex, with loadings of .374, .372, .369, .290, and .238 in the factors of perceptual speed, mechanical experience, spatial relations, verbality, and kinesthesis, which in the criterion are weighted respectively .260, .130, .190, .098, and .500. Similarly, it can be shown that the other four tests are relatively complex.

2. *Factorial Estimates of the Validity of Tests*

The validity of each test in a given factor is indicated by the correlation of the test with that factor; in other words, by the loading of the test in the factor. With respect to the criterion, instead of the factor, the validity of a test has been defined previously as the sum of the cross products of the paired factor loadings in the test and the criterion.

Discrepancies between the biserial correlation coefficients and the factorial estimates of validity are negligible. For the West Point group the largest single discrepancy between a reproduced coefficient of correlation and the original is +.030. With the exception of two discrepancies of −.0240 and −.0201, all others are less than |.0200|. In the Negro group, somewhat larger discrepancies occur in the reproduction of the biserial correlation coefficients, although the largest is −.0490. Three other reproduced coefficients deviate from the original biserial coefficient more than |.040|. Since the cross-products of the paired loadings in eight factors were used instead of nine (as was done for the West Point group), the somewhat larger discrepancies of the Negro group seem reasonable.

Comparisons may be made of the validities of different tests in the same battery as well as of the same test in different batteries (although the latter comparisons should be made with greater caution). For example, the consistently lower validity coefficients for the psychomotor tests in the Negro group are in essential agreement with the presence of a slight negative loading (−.050) in the psychomotor-coordination factor in the Negro pilot-criterion. In contrast, the higher validity coefficients of the apparatus tests in the West Point group correspond to the substantial positive loading of .305 in the psychomotor factor in the pilot-criterion.

3. *Tests containing Substantial Amounts of Unique Variances*

In a battery a test loaded with a large portion of variance not found in any other test is said to contain a substantial amount of *unique* variance.[5] Such a test need not be pure, although the probabilities are that it is relatively pure if the amount of unique variance is great. Arbitrarily two conditions may be stated to define operationally a test which contains a *substantial amount of unique variance:*

(1) that the loading (variance) of the factor in the test be equal to, or greater than, .500 (.250).

(2) that the loading (variance) of no other test in the battery with respect to the factor be equal to or greater than .300 (.090).

The fulfillment of such a set of stringent conditions tends to make the naming of a factor difficult, if not impossible.

However, as was pointed out in (IV, A, 3b and 5), three such circumstances arose in the identification of factors (pilot interest and kinesthesis). For both the Negro group and the West Point group the presence of a large amount of variance in the test *Biographical Data-Pilot* served primarily to identify the factor of pilot interest. Similarly, for the factor identified as kinesthesis, only one test (*Rudder Control*) contained a substantial amount of variance in this ability. The fact that in the circumstances mentioned two substantial load-

[5] By this definition any test relatively pure with respect to a factor contains a substantial amount of unique variance provided that no other test in the battery contains a large amount of variance in the factor. If two or more tests (either complex or pure in their factorial composition) yield a large portion of variance in the same factor, a unique contribution is no longer said to be made, since neither test is the only one supplying variance in the factor.

ings: one in a single test and the other in the criterion exist is indicative that a marked contribution to the validity of the battery is associated with a factor, be it common or specific, named or unnamed. Whether a construct such as a factor can be named or not is of little consequence if its presence may be inferred from the operations of the testing situation.

CHAPTER V

SUMMARY AND CONCLUSIONS

A. Summary

The purpose of the investigation was to ascertain the contributions of factors both to the description of tests and to the predictive values of these tests in two pilot populations of the United States Army Air Forces. To two samples, consisting of 815 West Point Cadets and of 356 Negro cadets, were administered respectively the November, 1943, Classification Battery, consisting of twelve pencil-and-paper tests and of six apparatus tests and the September, 1944, Classification Battery consisting of fifteen pencil-and-paper tests (of which seven were identical with those of the first battery) and of the same six psychomotor tests. That the two samples were non-homogeneous was demonstrated by the application of the Fisher *t*-test to the differences between mean composite scores (stanine standings) of the two groups, and to the differences between mean scores of tests identical to the two batteries. All differences between sample means were significant beyond the one per cent level.

In Chapter I, the problem of the investigation was clarified in terms of two major questions and in terms of sub-questions pertaining to them. In the first question the following points were proposed: (1) the identification and interpretation, for the two groups, of factors derived from two matrices of intercorrelations; (2) a comparison of the weights of the identified factors in the two groups with the loadings in the same factors for a third representative white aviation-cadet population; and (3) a comparison of the factorial composition of the pass-fall criteria (each of which was included in respective matrices of intercorrelations) of the three populations. In the second question the differences between the two groups in the prediction of the pass-fail criterion from scores in the tests optimally weighted was proposed for consideration, along with the relationship of the traditional multiple-regression techniques to approaches employed in factor analysis.

In addition to the definition of the two populations in operational terms, both a brief rationale for the inclusion of tests in the two batteries and a detailed description of the tests with respect to their purpose, content, scoring-formulae, and time limits were presented in Chapter II. Such background material was deemed important to a meaningful interpretation of factors.

In Chapter III followed a survey of the statistical procedures, which closely paralleled in sequence the two major questions proposed in the statement of the problem in Chapter I. Two factor analyses following the Thurstone system were made of the two matrices of intercorrelations consisting of nineteen and twenty-two variables respectively for the two groups of Negroes and West Point Cadets. For these two matrices, eight centroid and nine centroid factors, respectively, were extracted. By the Zimmerman method, a psychologically meaningful rotation of eight factors was effected for each matrix.

The Doolittle method was employed in the prediction of the pass-fail criterion from the tests of each battery. The relative contribution of each optimally weighted test to the total predicted variance of the pilot criterion (coefficient of multiple determination) was computed by the multiplication of the beta weight of the test and its coefficient of correlation with the criterion. Comparisons were then made of the variance-contributions of each test in a given battery to the total predicted variance.

In Chapter IV, the sequence of the two major divisions concerned with the interpretation of the statistical results was the same as that set up in Chapter III. For the West Point Cadets the eight rotated factors were identified as mechanical experience, number, pilot interest, psychomotor coordination, perceptual speed, reasoning, spatial relations, and verbality. For the Negro group, variance in seven of these eight factors appeared. A general intellectual factor made up of both numerical and reasoning components was evident. An eighth factor, tentatively identified as *kinesthesis* emerged.

Detailed comparisons were made of the factor-weights in the tests administered to the two groups. Moreover, loadings in eight factors for a white aviation-cadet population, available from previous analyses, were compared with those in tests administered to either or both of the two groups of West Point Cadets and Negroes.

Unfortunately, any difference between factor loadings in the same test or in different tests administered to two or more groups cannot be tested for statistical significance. The impossibility of determining the standard error of a factor loading is probably the major disadvantage of the Thurstone system. Nevertheless, hypotheses were set up to rationalize what appeared to be noticeable differences among the three groups in loadings of any test with respect to a factor under consideration.

Similarly, attempts were made to rationalize the differences in the factorial composition of the pilot-criteria of the three groups—differences much more pronounced than those appearing in the tests. For the West Point Cadets the three most valid factors were pilot interest, spatial relations, and psychomotor coordination, with loadings of .436, .415, and .305, respectively, in the pilot criterion. For the Negro cadets the three most valid factors were kinesthesis, perceptual speed, and spatial relations with loadings of .500, .260, and .190, respectively, in the pilot criterion. Previous analyses revealed that for representative white aviation-cadets (mostly pilots) the most valid factors were spatial relations, mechanical experience, and psychomotor coordination.

Loadings in intellectual factors such as number, reasoning, and verbality were low in the pilot criterion for all three groups; in fact, the number factor actually contributed negatively to the prediction of pilot success for the three groups. Negative contributions of variance to the total predicted variance in tests emphasizing numerical operations substantiated the importance of these negative weights in the criterion.

Two rather surprising results in the factor analyses were: first, the relatively high loadings for Negroes in the verbal factor in pencil-and-paper tests not intended to measure the verbal factor; second, the negative loading of the psychomotor factor in the pilot criterion for Negroes. In the former instance, the hypothesis was proposed that the level of reading difficulty of the lengthy directions of several pencil-and-paper tests more nearly approximated the true level of the Negroes than did the test in reading comprehension. For the second result, an hypothesis was suggested of inhibition in learning, or negative transfer effect. Carried over from previous experiences in which a high degree of psychomotor coordination had been developed, certain patterns of motor response might have blocked the successful execution of new motor reactions required in the early stages of pilot training.

Relating the traditional multiple-regression techniques to the newer approaches of factor analysis proved illuminating. Stress was placed upon the complementary contributions of the two procedures. For the two groups of West Point Cadets and Negroes, the respective amounts of total predicted variance (or the coefficients of multiple determination) were approximately three-fourths and one-half of the communality (sum of the common-factor variance) of the pilot criterial. For the two groups the coefficients of multiple determination were .3610 and .1772 respectively, compared with the respective communalities of .484 and .381 in the pilot criteria after rotation. In both groups, several tests contributed negative amounts of variance to the total predicted variance. To a considerable extent the presence of negative loadings of factors in the pilot criterion with which these tests were loaded accounted for the negative contributions of variance. Although psychomotor tests were highly valid for West Point Cadets, the slight negative weight of the psychomotor factor in the Negro criterion accounted for the small positive portions of variance contributed by apparatus tests to the coefficient of multiple determination. In order that validity might be maximized, two courses of action were suggested: the introduction of new tests to aid in the identification and description of other potential factors in the pilot criterion, and the purification of tests with respect to factors highly weighted in the criterion.

The application of traditional multiple-regression techniques to the study of validity indicated that pencil-and-paper tests individually furnished as much variance to the total predicted variance as the psychomotor tests, or perhaps even more variance than these apparatus tests. Particularly important was the finding that, for the West Point groups, the spatial relations factor could be better described by a form of pencil-and-paper test than by any one of the psychomotor tests.

In general, the more factorially complex tests furnished the largest amounts of variance to the total predicted variance. A form of a test in instrument comprehension, which was highly complex factorially for the Negro group, yielded more than half the variance to the coefficient of multiple determination. However, for the West Point Cadets, two relatively pure tests—a different form of the test in instrument comprehension, and an adaptation of the familiar rotary pursuit task—made the second and third greatest contributions of variance, .0752 and .0589 respectively.

For the West Point Cadets, one test contained a substantial amount of unique variance (variance not common to any other test) in the factor of pilot interest. The identification of the factor as pilot interest is somewhat tentative, inasmuch as a single test, a biographical data blank, largely defined it. However, the presence of a high weight in the criterion for the same factor in which the test was saturated is indicative of a high degree of validity for the biographical data blank. As expected, the multiple-regression approach showed that its contribution to the total predicted variance was .0510. For the Negroes the same factor appeared in this test, but the small loading of the factor in the pilot criterion limited its predictive value (.0115 to the total predicted variance). A second test, involving a rudder control task, contributed unique variance to the validity of the battery administered to Negroes. The high weights in the factor, identified as kinesthesis, in both the test and the criterion, accounted for its variance contribution (.0277) to the coefficient of multiple determination.

B. Conclusions

Certain limitations in the interpretation of the results have been indicated previously. The lack of complete identity of the tests in the two batteries naturally restricted, to a degree, the comparisons which might be made in factor loadings of tests which were identical in the two batteries. The lack of a statistical test for the standard error of a factor loading prevented calculation of the probability of a reliable difference between two or more factor weights. In the multiple-regression procedures, tests for linearity were not made, although superficial checks indicated that the assumptions of homoscedasticity and of rectilinearity were fulfilled to a satisfactory degree.

Sweeping generalizations of the results obtained for the two samples to the description of characteristics of racial or socio-economic groups is definitely not warranted. Nor can any conclusions be made, from the results obtained, concerning the indication of possible genetic differences in the two samples.

The following five conclusions have been formulated from the results of the investigation:

(1) In general, the factor loadings of tests of the two batteries administered to the two groups of West Point Cadets and Negroes and of the same tests given to a representative white aviation-cadet population were comparable, although some differences did appear, notably with respect to factors weighted less than .400 in a test for any one of the three populations.

(2) The factorial composition of the pilot criteria was markedly different for the three populations. Part of the observed differences may be due to the unreliability of the criteria. Some similarities were apparent:

(a) The factor of spatial relations was valid for all three populations in the prediction of pilot success.

(b) The intellectual factors of reasoning, number, and verbality were not valid for the prediction of pilot success.

(3) A new factor identified as kinesthesis appeared for the Negro population only. This factor was the most valid one for prediction of pilot success for Negroes.

(4) The unique contribution of a biographical data blank to the validity of a test battery was demonstrated for two populations of West Point Cadets and Negroes. A large amount of the variance of this test was identified to be that of pilot interest.

(5) 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 measure-

ment of human abilities frequently effected by more cumbersome devices.

C. Suggestions for Future Research

As mentioned several times previously, the development of tests relatively pure with respect to factors present in the pilot criterion is necessary for raising the validity of a test composite. Now that a substantial portion of the area of the pilot criterion is defined, individual tests which describe one factor at a time might well be administered on an experimental basis to pilot cadets in the peacetime Army. Factor analyses of the results would be likely to indicate that further revisions in tests would be desirable. Moreover, several tests need to be introduced on a trial-and-error basis, to test hypotheses concerning the roles of other factors in the pilot criterion.

For reasons of economy, particular attention should be directed toward development of pencil-and-paper tests which can duplicate the functions of the more cumbersome apparatus tests. For the space factor, marked progress in this direction already has been made.

Separate batteries for selection of personnel in different air-crew positions seem to be necessary if differential prediction is to be maximized. Tests measuring intellectual factors undoubtedly need to be included in batteries administered to prospective navigators. However, the presence of such tests in a composite administered to certain groups of pilot candidates may detract from the validity of the battery, if there has been initial selection on a general qualifying examination (which measures necessary minimum intellectual requirements for pilot training).

Of course, persistent attempts at a better description of the factorial structure of the criteria corresponding to various air-crew positions are paramount in importance. As the structure of the criterion in question becomes better known, the introduction of new tests and the revision of others will be necessary.

Not to be minimized, however, is the need for improved criteria. It is quite probable that the gross pass-fail criterion could advantageously be replaced by many independent and relatively pure criteria. The use of several independent criteria, each one of which can be differentiated into a number of levels of performance (instead of merely pass-or-fail), would appear to be a promising approach.

The application of these techniques in the field of vocational guidance affords an unlimited area of research. A comprehensive battery of relatively pure tests measuring the more important abilities in different occupations would permit a profile of one's scores in different factors to be made. If cut-off points could be established for factor scores in different occupations, a counselor after studying an examinee's profile of scores would be in a better position to give vocational advice. Applications in the field of educational guidance are immediately apparent.

Finally, the extension of factor-analysis procedures to the fields of personality and temperament would permit the clinical psychologist the use of the profiles suggested. Moreover, the inclusion in the same battery of pure tests designed to measure both aptitudes and personality factors would serve to narrow the gap between these two areas of testing and to allow joint use of these factor scores in vocational guidance, provided of course that the pattern of the temperament and aptitude factors in a given occupation had been ascertained.

In short, the potentialities of the factor-analysis procedures barely have been realized. It is only through continuous research that the horizon of man's knowledge of the primary abilities and of the components of temperament may be broadened. In combination with other procedures of demonstrated merit, the techniques of factor analysis afford at the present time a most encouraging means for the solution of fundamental problems in psychological and educational testing and in experimental psychology.

BIBLIOGRAPHY

A. Books

1. Ezekiel, M. *Methods of correlation analysis,* Second edition. New York: John Wiley & Sons, 1941. 531 pp.
2. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill Book Co., 1942. 333 pp.
3. ———. *Psychometric methods.* New York: McGraw-Hill Book Co., 1936. 513 pp.
4. Hull, C. *Principles of behavior.* New York: D. Appleton Century, 1943. 422 pp.
5. Thurstone, L. L. *Multiple factor analysis.* Chicago: Univ. Chicago Press, 1947. 535 pp.
6. ———. *A factorial study of perception.* Chicago: Univ. Chicago Press, 1944. 148 pp.

B. Periodicals

7. Balinski, B. An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psych. Monographs,* 1941, **33**, 191-234.
8. Clark, M. P. Changes in primary mental abilities with age. *Archives Psych.,* 1944, **291**, 1-30.
9. Guilford, J. P. Human abilities. *Psychological Rev.,* 1940, **47**, 367-394.
10. Guilford, J. P. New standards for test evaluation. *Educational & psych. Measur.,* 1946, **6**. 427-438.
11. Guilford, J. P. & Zimmerman; W. S. Some A.A.F. findings concerning aptitude factors. *Occup.,* 1947, 26: 154-159.
12. Langfeld, H. (Ed.) Symposium on operationism. *Psychological Rev.,* 1945, **52**, 241-294.
13. Richards, T. W. Genetic emergence of factor specificity. *Psychometrika,* 1941, **6**, 37-42.
14. Spence, K. The nature of theory construction in contemporary psychology. *Psychological Rev.,* 1944, **51**, 47-68.
15. Thomson, G. The influence of univariate selection on the factorial analysis of ability. *British Jour. Psych.,* 1938, **28**, 451-459.
16. Thomson, G. & Lederman, W. The influence of multivariate selection on the factorial analysis of ability. *British Jour. Psych.,* 1939, **29**, 288-306.
17. Thurstone, L. L. Primary mental abilities. *Psychometric Monog.,* 1938, **1**, 1-128.
18. ———. The effects of selection in factor analysis. *Psychometrika,* 1945, **10**, 165-198.
19. Woodrow, H. The common factors in fifty-two mental tests. *Psychometrika,* 1939, **4**, 99-108.
20. Zimmerman, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika,* 1946, **11**, 51-55.

C. Government Publications

21. Du Bois, P. (Ed.) *The classification program.* Report No. 2, Army Air Forces Aviation Psychology Program. Washington, D.C.: U. S. Government Printing Office, 1947, 394 pp.
22. Guilford, J. P. (Ed.) *Annual report of Psychological Research Unit No. 3.* Santa Ana, Calif.: Santa Ana Army Air Base, 1944.
23. ———. (Ed.) *Printed classification tests.* Report No. 5, Army Air Forces Aviation Psychology Program. Washington, D.C.: U. S. Government Printing Office, 1947, 919 pp.
24. *Composite factor analysis summary.* San Antonio, Texas: Department of Records and Analyses, AAF School of Aviation Medicine, 1045th AAF Base Unit, September, 1945.
25. *Research notes 45-1.* Psychological Section, Office of the Surgeon, Headquarters, AAF, Training Command, January 13, 1945.