REEL-C 6568

A.T.I 165759

UNCLASSIFIED

DEPARTMENT OF THE ARMY

STUDIES IN VISUAL ACUITY

PRS REPORT NO. 742

PREPARED BY THE
STAFF, PERSONNEL RESEARCH SECTION
THE ADJUTANT GENERAL'S OFFICE

1948

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1948

For sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.
Price $1.00 Buckram

FOREWORD

The purpose of the studies reported herein was to examine various aspects of visual acuity through a factor analysis of correlations among tests of visual acuity; to select the most representative tests; and to investigate item difficulty and scoring methods. The studies were conducted by the Personnel Research Section, AGO, Department of the Army.

Members and former members of the Personnel Research Section with major responsibility for the research include the following: Dr. Edwin R. Henry, Dr. Robert J. Wherry, Dr. Douglas H. Fryer, Dr. Horace H. Corbin, Mr. Calvin Taylor, Mr. Lawrence Karlin. Technicians of this Section who were responsible for the analysis of the data were: Dr. E. K. Taylor, Mr. Richard H. Gaylord, Mr. Kenneth Wood, Miss Claire Tajen, and Mr. Arthur Fifer. Drs. Robert J. Wherry and Horace H. Corbin were responsible for the preparation of the report.

Members of the Subcommittee on Visual Testing of the Army-Navy-NRC Vision Committee acted as consultants, including Capt. Charles W. Shilling (MC), USN, Chairman, and Dr. Donald G. Marquis, Executive Secretary.

Additional members of this committee who deserve special mention for their advice and participation in constructing the tests are: Lt. Comdr. Dean Farnsworth H(S), USNR, Lt. Comdr. (DR.) Henry Imus H(S), USNR, Lt. Comdr. Elsworth B. Cook H(S), USNR, and Dr. H. Richard Blackwell.

The test charts and chart changers used in the Fort Dix, N. J. study were produced, at their expense, by the Bausch & Lomb Optical Co., Rochester, N. Y. Members of this firm who devoted considerable time and effort in this task were: Mr. Fred Jobe, Mr. Ray Stegeman, and Mr. E. O. Bergmann.

The data derived from the administration of the three commercial devices and other tests at the Submarine Base, New London, Conn., were made available by and analyzed with the helpful advice of Lt. Comdr. Elsworth B. Cook H (S), USNR, Lt. Comdr. John H. Sulzman (MC), USNR, and Lt. (j.g.) Neil R. Bartlett H(S), USNR.

CONTENTS

List of Tables

List of Figures

Section I. PROBLEM

1. Origin of the Problem

During the last few years it has become evident to all concerned that present methods of testing of visual acuity have given rise to great inconsistency in the physical standards for vision which have been adopted for military service. Current devices and methods used in the measurement of this important ability have been, at best, of questionable reliability and validity. The method of testing visual acuity has varied greatly among the services, within the services, and frequently from day to day at the same installation. This variability can be traced in most instances to both the lack of standardization of testing conditions and the use of different test forms which have led to inconsistent interpretations. Such inconsistency in testing a basic function has undoubtedly worked great hardship on the individual, interfered with training schedules, led to faulty classification, and consequently cost the government a considerable amount of time and money.

Visual acuity has been measured at two principal points in the process of Army selection, training, and assignment. The first such point was at the armed forces induction station where all prospective service men were examined by means of the Snellen Test. Acceptance for or rejection from one, the other, or all the services was accomplished at this time. The second point was at various training centers and classification and assignment stations. Here a second examination was often given, as it certainly was in the selection of aviation cadets and many other specialists. If such a reexamination were not made, the record of the induction station examination was interpreted for such assignments as were known, or thought to require specific visual ability. A consistant attempt was made to make maximum use of the manpower available on the basis of visual abilities as well as many other indices.

The extent to which the use of vision tests of doubtful reliability and validity may have been costly is partially revealed by the results of medical screening for induction into the armed services. An analysis of DDS Form 221, the Worksheet used at Induction Stations for processing, assembled for the year April 1942–March 1943, reveals that 7.8 percent of those rejected from service in that period were so treated because of visual deficiencies. This percentage is not large when taken by itself, but in comparison with other causes for rejection it assumes great importance. Of some 30 causes listed in this report, the major causes of rejection and percentages attributable to each are:

Cause of rejection	*Percentage of those rejected*
Mental diseases	12.5
Syphilis	9.4
Musculoskeletal deficiencies	9.3
Cardiovascular deficiencies	8.8
Eye deficiencies	7.8
Hernia	7.8
Educational deficiency	7.7
Neurological deficiencies	5.5
Ear deficiencies	5.0
Tuberculosis	3.7
Mental deficiency	3.0

The percentage for other causes are below 3.0 percent in each case. With the exception of rejections based on mental disease, syphilis, and musculoskeletal and cardiovascular deficiencies no other group stands higher on the list than that rejected for faulty vision.

If the figures quoted above can be accepted as fairly representative of operations over a longer period, we may estimate the number of men affected. Of 10,810,169 men examined over the period of 1 November 1940 to 31 December 1943 it is reported that 3,509,248 were rejected. These figures are for combined white and colored groups. On the basis of the percentages quoted above, 7.8 percent, or roughly 273,000 of these rejectees were eliminated on the basis of visual defects. Disregarding classification and assignment variables, this number of men represents, approximately, the enlisted men for *20 infantry divisions*. From the same basic data it may be noted that educational deficiency accounted for 7.7 percent or approximately 270,000 of the rejectees. It appears that an unstandardized methodology for evaluating vision accounted for as many or more rejections from the services as did a much studied technique for evaluating educational achievement.

Undoubtedly, not all of the 273,000 men eliminated for defective vision were rejected on the sole basis of visual acuity tests. Many other considerations were involved. However, there is no question that these tests served as the initial screening device and were the primary means of identifying rejectable men. Again, there is no question but that many of the men eliminated by these unstandardized tests would have been treated similarly had the tests undergone as much study as is presently contemplated. Yet it is virtually certain that many misclassifications were made in not only the group rejected but among those selected for service.

We have treated only the induction screening of prospective military personnel as an instance of the wasteful operation of the tests and methods in vision testing. After induction, minimum vision standards for specialized training and duty exercised further influence on the selection of service men. The number of cases in which men were selected and assigned to special training on the basis, among other things, of visual acuity tests given at one station only to be found wanting in vision when they reached the school, have been reported in great numbers. Needless to say, researches to improve this general situation have long been necessary.

2. Research Approaches Accomplished

During World War II many steps were taken toward solution of the more basic problems of testing visual acuity. The coordinating agency for most of the work accomplished was the Subcommittee on Visual Examinations of the Army-Navy—NRC Vision Committee. The parent organization was formed in 1944 to coordinate research of all types in the field of vision after it was found that many widespread agencies were occupied in the same researches and might profitably get together in a joint organization. The committee was, and is, composed of members of both the Army and Navy and several scientists from civil institutions. The Subcommittee on Vision Examination first prepared a manual on the testing of visual acuity which specified standard optimal conditions under which the tests should be given. They also prepared several test charts, mainly letter charts, which were subjected to considerable study at several military installations. It was hoped that as a result of the subcommittee's work, a standard procedure and standard charts for testing vision would be available to all units of the armed forces. This would naturally enhance the value of any such tests in that the variability of test results brought about by changes in the physical set-up and method of testing from establishment to establishment would be reduced to a minimum.

A second source of concern in this process of standardization was indecision on the part of many investigators as to what visual acuity tests were actually measuring. Questions most frequently asked were: What is visual acuity? What does a score on a visual acuity test mean? These questions arose as a natural result of the use of many different test-objects (letters, checkerboards, contrasting grays, and forms) in tests currently employed. The answers to these questions were considered available, in part, through further research in the form of field studies in which large groups of examinees would be given many types of tests and comparison made of performances of the same group on all of the forms involved. It was found that many agencies were interested in seeing this research accomplished. The Air Surgeon initiated a request that a study be made, "to determine: (*a*) The reliability of * * * visual acuity tests * * *. (*b*) Intercorrelations of * * * visual acuity tests." (Reference, Memorandum for: Assistant Chief of Staff, G–1, Subject: Visual Examination, dated 21 September 1945.)

The Surgeon General concurred in this proposal (1st Ind. D/F), from ASF, SGO to the Assistant Chief of Staff, G–1, file No. SPMDP 702, dated 10 October 1945). As a result, The Adjutant General was directed to make "a study * * * to establish standard procedure and standard charts for determination of visual acuity and uniformity in testing the various visual criteria." Reference, Disposition Form, From: G–1, WDGS To: TAG, file WDGAP 201.5 (10 Oct. 45), Subject: Visual Examination, dated 22 Oct. 45.) The Personnel Research Section, Personnel Research and Procedures Branch of The Adjutant General's Office was given this task.

3. Present Problem

With the Subcommittee on Visual Examination of the Army-Navy-NRC Vision Committee, officials of the Personnel Research Section planned the initial research project (PR–4075–01) which is covered by this report. The problem outlined for the project is threefold:

a. To determine under standard conditions the various aspects of visual acuity through a factor analysis of correlations among tests of visual acuity recommended by the Army-Navy-NRC Vision Committee.

b. To select the most representative (pure and reliable) test of each such aspect.

c. To make preliminary investigations of item difficulty and scoring methods on the test studied.

Section II. POPULATION

1. Number of Subjects

Seven hundred and ninety-two enlisted men at Fort Dix, N. J. Of these 261 were tested and retested with 24 hours between the sessions. The remaining 531 men were tested only once.

2. When and Where Tested

The men were tested between 30 September and 5 November 1946 at Reception Center, Fort Dix, N. J.

3. Basis of Selection

Initially, men were selected for testing on the basis of age and Army Snellen visual acuity test score for the *left eye only.* This information was obtained from WD AGO Form 63 at the physical profiling station of the Reception Center processing flow. Criteria for selection were:

a. Equal numbers in each age group, by year, from 18–29 years of age.

b. Distribution of each age group by Army Snellen left-eye visual acuity scores according to the following:

	Percent
Over 20/40	5
20/40	10
20/30	20
20/20	50
Below 20/20	15

This method of selection proved to be unsatisfactory on two counts:

(1) Unreliability of Army Snellen scores obtained from Form 63, and

(2) Constriction of the flow of personnel through the reception center to a point where it was impossible to sample the age groups on an equal basis.

Unreliability of the Army Snellen scores was first of all a result of the fautly test practice of stopping successful examinees at the end of the 20/20 line. It soon became obvious to the research staff at Fort Dix that almost no men were appearing with previously determined Snellen scores below 20/20, i. e., 20/15. Unreliability of these indices was evidenced, too, by differences noted between the Snellen scores of record and those obtained in the current research.

In spite of these deficiencies an attempt was continued throughout the study to select men on the basis of a wide variety of ages and a wide spread of left-eye scores on the basis of previously obtained Snellen scores. Table I shows in terms of percent, the desired and obtained distributions of Army Snellen score and age. As to previous Army Snellen, the distribution of obtained cases matches the desired fairly well at the over 20/40, 20/40, and 20/30 categories but fails at 20/20 and below. As mentioned, the small percent (0.82) for below 20/20 is probably due to failure to test men in this part of the chart. The huge number of cases piled up at 20/20 (69.36 percent), most likely includes many men whose scores would be somewhere below 20/20 if the test had been extended to their "real" threshold. Meanwhile, age of the group piles up at 18 and 19 years; we expected no cases of 17-year-olds. Down the years the table shows fewer and fewer cases with a slight upward jump at 28 years for unaccountable reasons. In general, it can be said that a spread of ages was obtained, but that, short of waiting on cases for months, equal numbers at each age were unattainable.

TABLE I.—*Percentage of cases in previous Army Snellen and age categories for 731 and 792 cases respectively showing the desired and obtained distributions*

Previous Army Snellen			Age		
Score	Desired percent	Obtained percent	Years	Desired percent	Obtained percent
Over 20/40	5. 00	7. 93	17	0. 00	0. 63
20/40	10. 00	6. 02	18	8. 33	20. 08
20/30	20. 00	15. 87	19	8. 33	21. 97
20/20	50. 00	69. 36	20	8. 33	9. 97
Below 20/20	15. 00	. 82	21	8. 33	6. 82
			22	8. 33	5. 30
Total	100. 00		23	8. 33	4. 42
			24	8. 33	3. 91
			25	8. 33	3. 66
			26	8. 33	3. 41
			27	8. 33	4. 80
			28	8. 33	9. 85
			29	8. 33	5. 18
			Total	99. 96	100. 00

4. Typicality of the Sample

The men tested were judged typical of the general population coming into the Army as of October–November 1946. They included a mass of enlistees who sought Army service in order to come under the benefits of the GI bill of rights for which an October 5 deadline had been set. A few men were reenlistees, but the majority were typical of the postwar draft group. No reason has been found to suspect a typical quality in the visual or other characteristics of those tested.

Section III. VARIABLES AND CONTROLS

Fourteen tests of visual acuity were constructed for the study. There were designed for administration under identical conditions of lighting, test room layout, and examining procedure. Description of the test variables and the conditions under which the tests were administered follows:

1. Test Charts

The *basic design* of all the tests was the chart type universally used to measure for acuity, such as Army Snellen. Into this test-form several types of visual objects were reproduced, each type representing what experts in the visual examination field felt were the basic aspects of visual acuity as measured by the respective tests. The types of objects employed were:

a. LETTERS. These have been used universally in the past as the clinical material for scaling visual acuity. Of the four letter tests used in the study three were modifications of the more familiar Snellen test, designed to give more reliable scores. Extensive research went into the development of these tests at Naval Medical Research Laboratory, New London, Conn., and the AAF School of Aviation Medicine, Randolph Field, Tex.

(1) *Army Snellen* (see fig. 1, Test 2). The letters in each row are of equal size and equal width of stroke and space between strokes. The size of the letters and the width of stroke decreases for succeeding rows from top to bottom of the chart. The number of letters per row increases by one from row to row beginning with one at the top and reaching nine at the bottom.

(2) *New London Letter* (see fig. 2, Test 5). The letters in each row are equated for difficulty and are of equal size and width of stroke. The size and width of stroke decrease from row to row from the top to the bottom of the chart. The number of letters per row increases by one from five in the first row to eight in the fourth and remains at eight in the rows which follow.

(3) *AAF Letter* (see fig. 3, Test 8). The same six letters are employed in each row, and within the rows they are equated for difficulty, size, and width of stroke. The size and width of stroke decreases from row to row from the top to the bottom of the chart.

(4) *AAF Constant Decrement* (see fig. 4, Test 11). The size of the letters and width of stroke decrease continuously by one Snellen

L

O B

T C L

P L O E

E D T O L

P L C T D E

D T O E C L P

C L D T O E C L

P E Z O L C F T D

FIGURE 1.—*Army Snellen—Test 2.*

G N V H Y

V C E K H Y

O X Z V G E K

N C Y Z G X K H

O N C E H N K X

Z G H Z O H C Z

Y E V Y E K C N

O Z H V O X Y C

Z K X V G E N O

C Y V N K E G V

N G Z X K N C E

FIGURE 2.—*New London Letter—Test 5.*

791030—48——2

K N G S Y P

N K Y P G S

G P K Y S N

S G N K P Y

Y S P G N K

P Y S G K Z

K G Y N S P

N Y G K P S

G K S P Y N

S N P G K Y

Y P N S G K

FIGURE 3.—*AAF Letter—Test 8.*

V N C

E O Z N V O

Z C V E Z O N C

V E N E Z O N C Z

C V E O Z C N E V O

FIGURE 4.—*AAF Constant Decrement—Test 11.*

step from letter to letter. This decrement runs across each row and from top to bottom of the four rows of the chart.

b. RESOLUTION OBJECTS. There were four tests which employed objects such as dots and lines in an attempt to measure resolution as an aspect of visual acuity. The letter tests, described above, have been thought to measure resolution but are complicated by the inclusion of familiarity of the letters. Each of the present tests required the examinee to report for each item in which of four possible quadrants of a diamond-shaped space the test object appeared. This mode of response required the examinee to report "place where" the object occurred, whereas, in the letter tests the examinee reported the actual letters. It is essential to the success of these tests that the examinees continue to respond even when they believe it is guesswork. The tests in this group were—

(1) *Dot Variable Size* (see fig. 5, Test 3). A dot of constant blackness was located in one of four corners of a constant-sized white diamond which was surrounded by a field of uniform gray. The size of the dot was constant within each row of four diamonds but decreased from row to row from the top of the chart down. The dots were located in corners of the diamonds in a chance order.

(2) *Line Resolution* (see fig. 6, Test 9). A line of constant length and constant blackness was located in one of four corners of a constant-

sized white diamond surrounded by a field of uniform gray. In each item the line was placed to radiate from the diamond center to one of the four corners. In each row of the test chart there are four diamonds, each containing a line of equal width which pointed to a different corner of each diamond in a chance order. The width of the lines decreased from row to row down the chart.

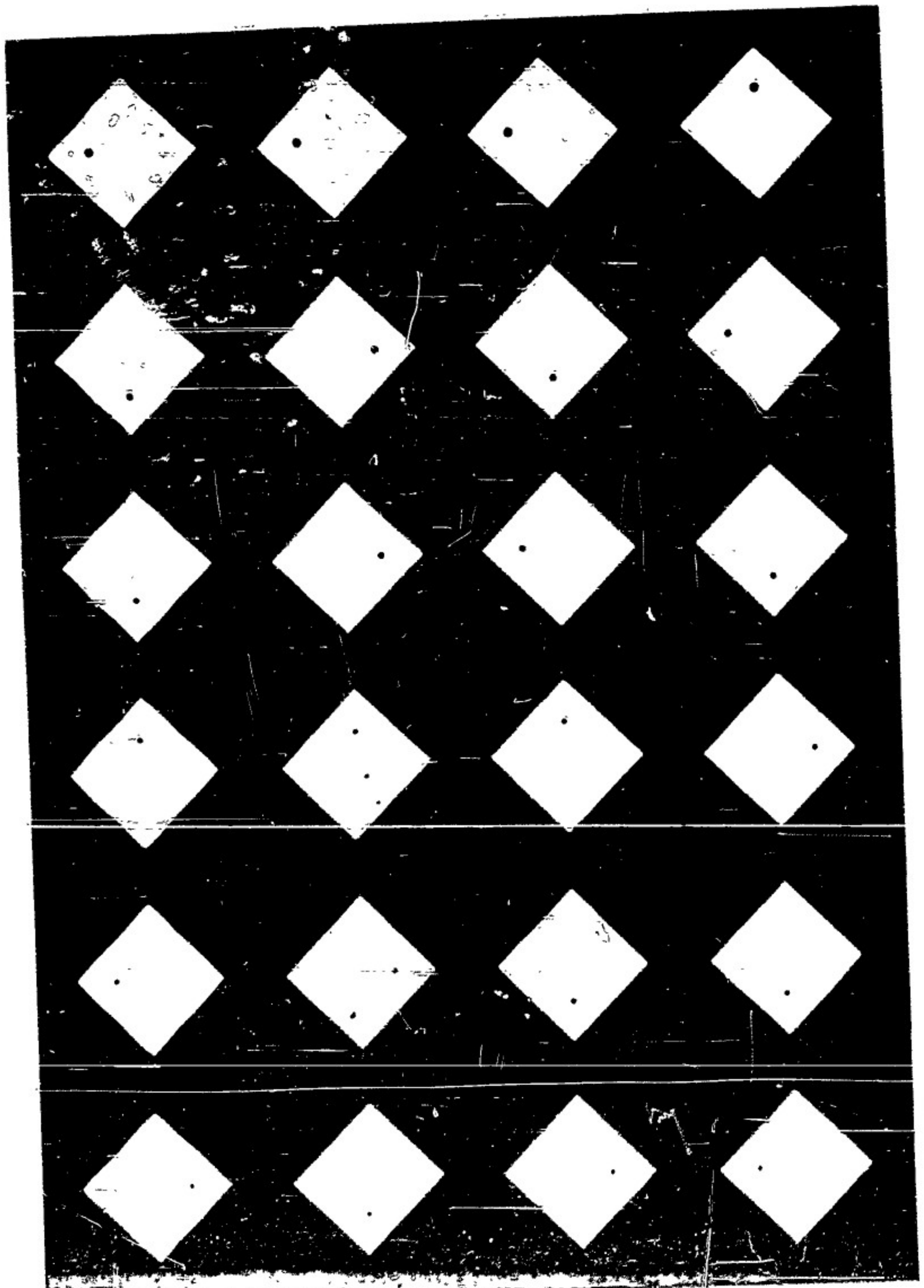


FIGURE 5.—*Dot Variable Size—Test 3.*

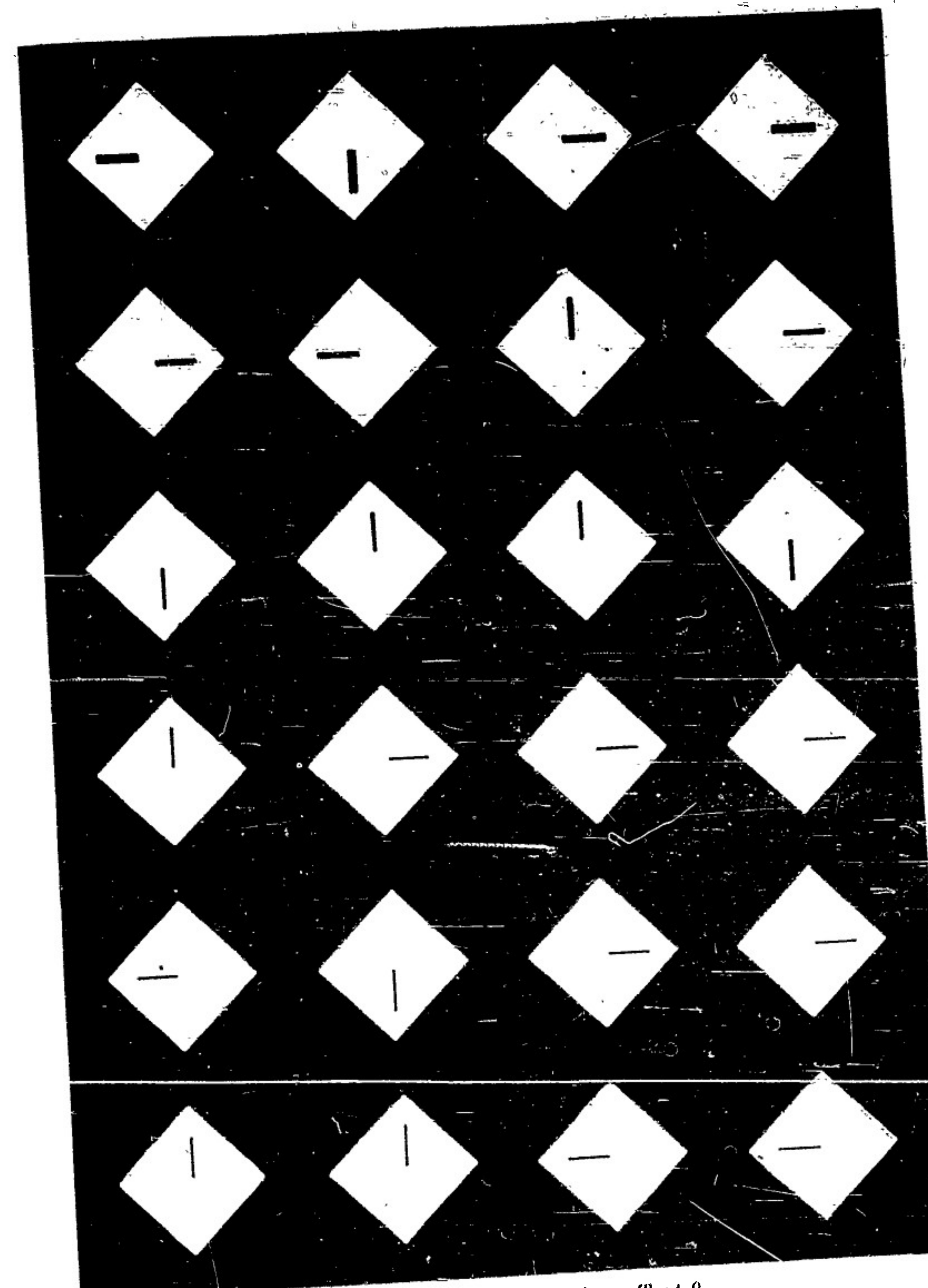


FIGURE 6.—*Line Resolution—Test 9.*

(3) *Bausch and Lomb Checkerboard* (see fig. 7, Test 7). This was the same test object used in the Bausch and Lomb Orthorater increased in size for use at 20 feet. Each test item consisted of a diamond grouping of four squares, one of which contained a grid (checkerboard) coarser than the other three. The coarser grid was a checkerboard type made up of rows of alternating black and white

squares, while the grids of the other three squares in the diamond consisted of rows of black dots. Viewed from 20 feet, the three squares consisting of black dots were not resolved and appeared to examinees as squares of gray. The size of each diamond group remained the same for each row of four items, but decreased from row to row down the chart. As the whole group decreased in size, the size of checkerboard dotted squares decreased in proportion as in photographic reduction. The position of the checkerboard square varied from item to item in chance order.

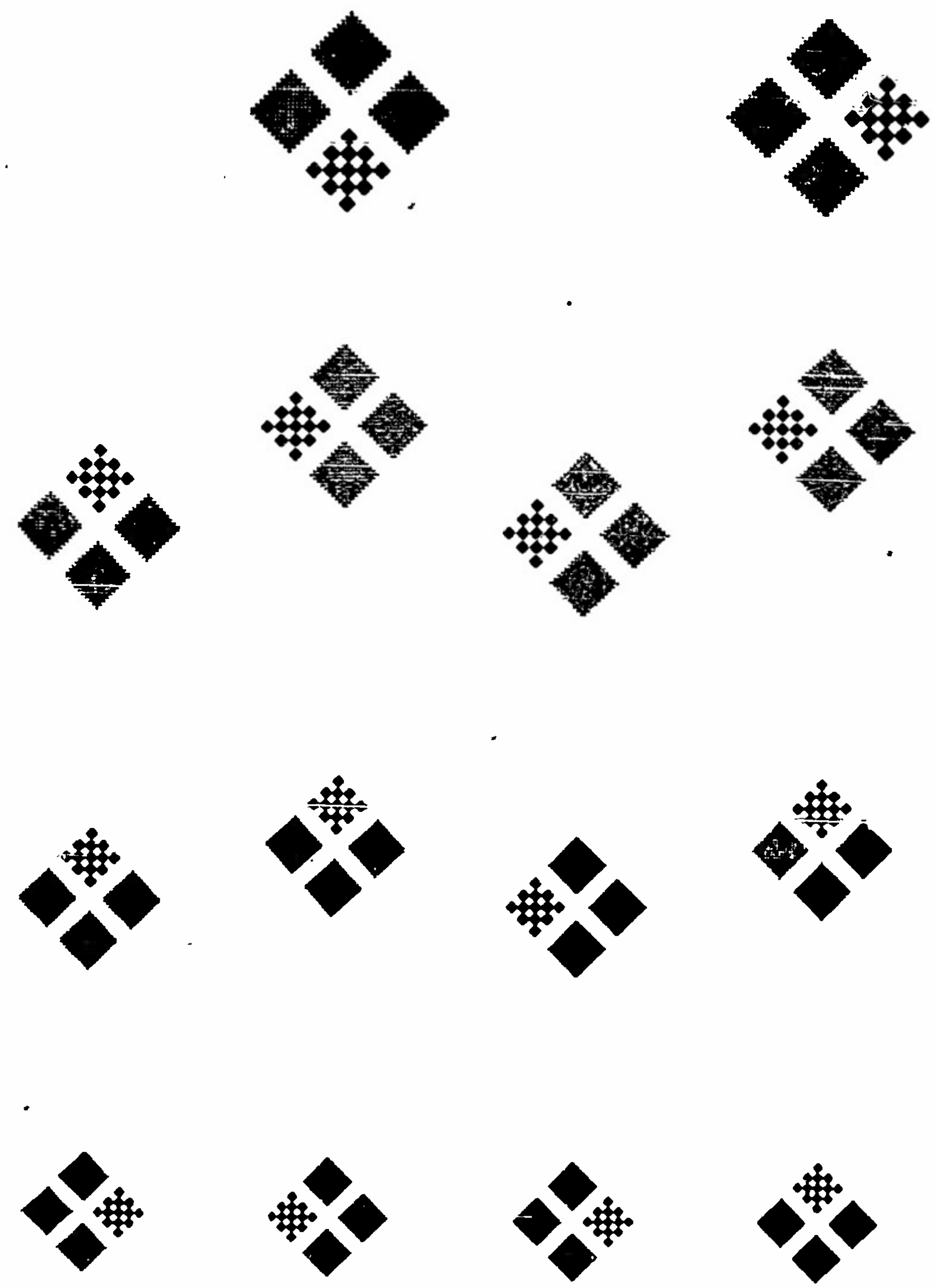


FIGURE 7.—*Bausch and Lomb Checkerboard—Test 7.*

(4) *Checkerboard Variable Grid* (see fig. 8, Test 13). This test used the same test object as the Bausch and Lomb Checkerboard test, except that the figure size remained constant while the checkerboard square of the diamonds of successive rows incorporated an increasing number of black and white spaces with a resultant decrement in the size of the spaces. Four items were used in each row of this test, and once again the location of the square containing the checkerboard grid is distributed in chance order among all the items.

c. CONTRAST SENSITIVITY OBJECTS. Two tests were included which employed brightness contrast between the objects and background. The objects varied in grayness from that of the background in both tests, being always darker in shade. Both tests employed the same response device that was used in the resolution tests described above. The tests were:

(1) *Quadrant Variable Contrast* (see fig. 9, Test 4). Each test object consisted of three quadrants of a constant-sized square similar in shape to an arrowhead. The brightness of the arrowhead varied from a darker than the background to a shade which almost matched the background. Contrast between the figure and background decreased from left to right within each row of items and from the top row down. The direction the arrowhead pointed in each item was either up, down, right, or left, determined on a chance basis. In figure 9 we present only a single item of this test. This item is not a facsimile of any item of the actual test but shows the essential nature of the test object. Photo reproduction of the actual brightness items proved impossible and for that reason this fictitious item is presented.

(2) *Dot Variable Contrast* (see fig. 10, Test 10). The test object was a constant-sized dot about 0.25 inch in diameter. The brightness of the dot varied from darker than the background to a shade which almost matched the background. The dot was located in one of the four corners of a constant-sized diamond which was the background mentioned above. The diamonds in turn were surrounded by a field of uniform gray. Contrast between the dot and ground decreased from left to right and within each row of items and from the top row down. The location of the dots in the corners of the diamond was determined on a chance basis. Figure 10 presents a single test item for the reasons stated under (1) above. Again, this sample is more illustrative of the nature of the item than could be achieved by photo reproduction of the entire test chart.

d. FORM DISCRIMINATION. Two tests were included which used a type of form discrimination as the basis of composition of the test objects. In both cases this was a discrimination between rectilinear and curvilinear sides of the objects employed. Both tests depended

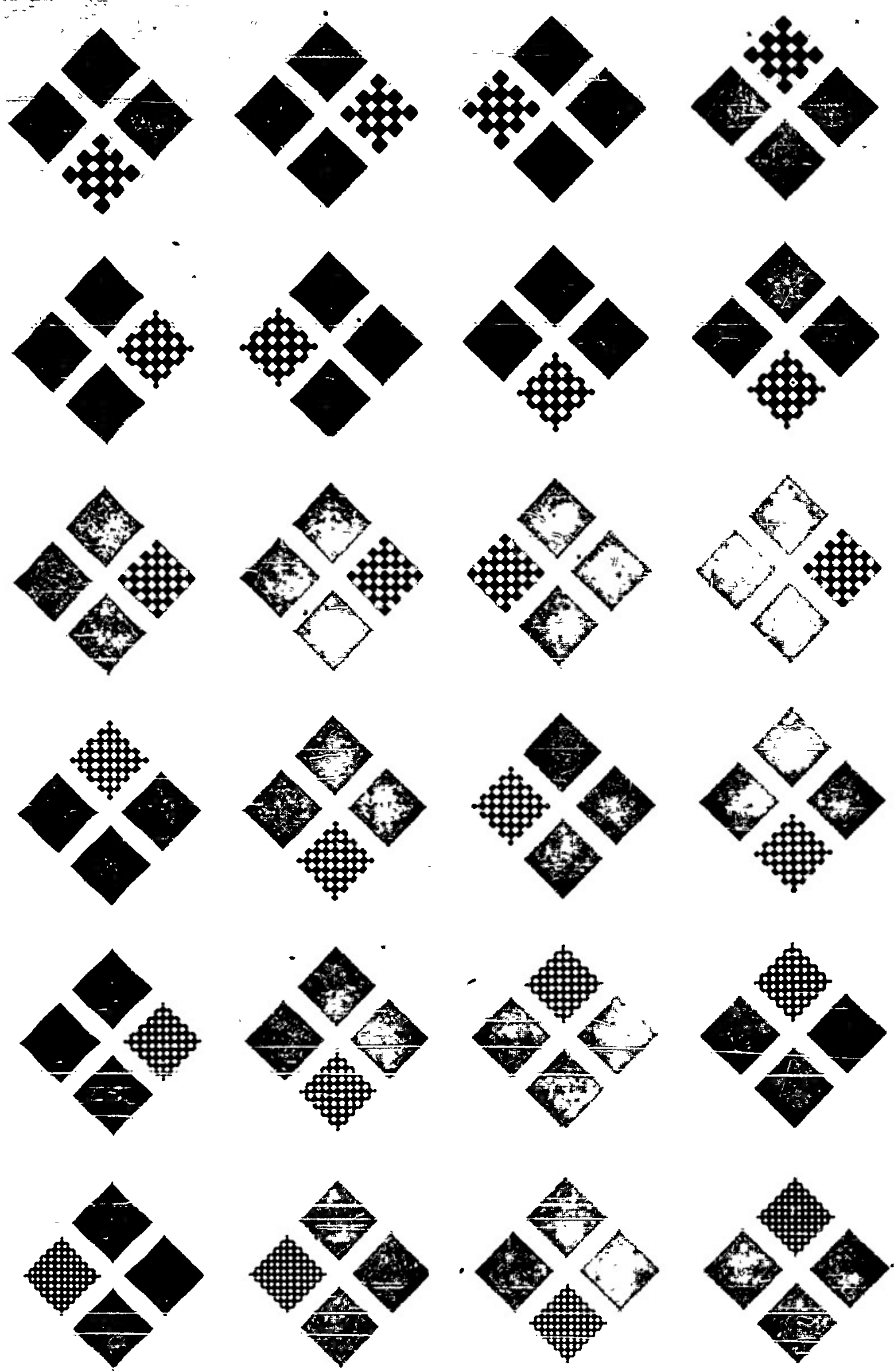


FIGURE 8.—*Checkerboard Variable Grid—Test 13.*

FIGURE 9.—*Quadrant Variable Contrast—Test 4.*

on success in presenting curvilinear forms with rectilinear ones without variations in area among them. All items were constructed in a manner which held such areas constant.

(1) *Triangle Discrimination* (see fig. 11, Test 6). Each item looked something like a Maltese cross. Four triangle shapes were placed about a circular dot so their apexes touched the dot at 90° intervals, and their bases were at the top, bottom, right, and left of

FIGURE 10.—*Dot Variable Contrast—Test 10.*

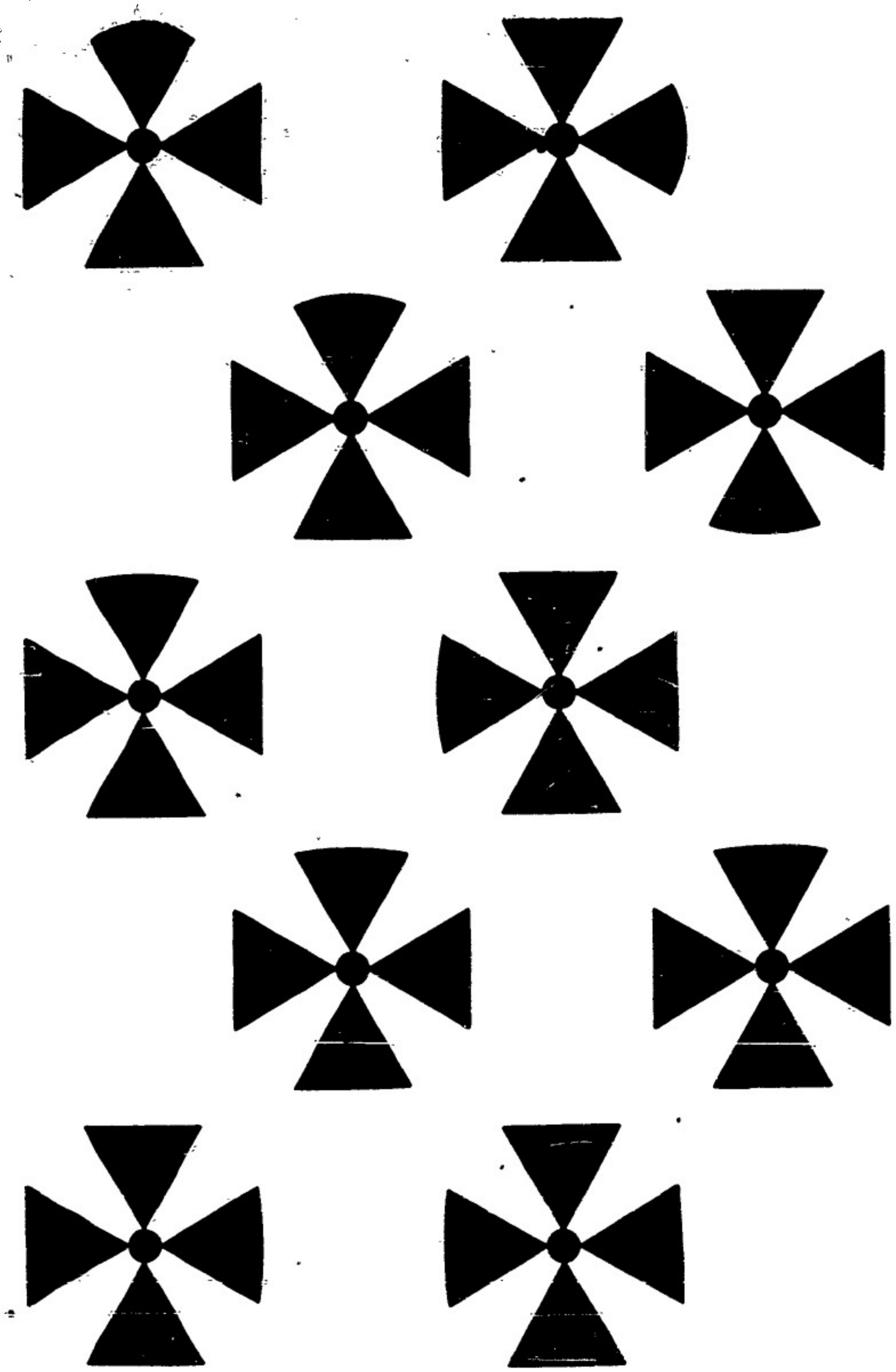


FIGURE 11.—*Triangle Discrimination—Test 6.*

the figure. Three of the triangles were equilateral, while the fourth was identical with the others in all respects, except the side, or base, opposite the apex was bowed convexly. The convexly bowed side was constructed in such a way that this triangle was equal in area to each

of the other three. The degree of convexity was constant for the two items in any row, but decreased from row to row from top to bottom of the chart. The position of the curved-sided triangle was distributed among the items on a chance basis.

(2) *Square Discrimination* (see fig. 12, Test 12). Six square-shaped objects were placed side by side in a row, making one item of the test in each case. Of the six objects in each row, one was not a true square while the other five were. The sides of the odd figure were convexly bowed, but were so drawn as to make the area of this figure equal to that of the other five (1 square inch). The odd figure was distributed in chance order among the middle four figures in each row. The degree of convexity of the sides of the non-square object decreased from row to row from the top of the chart down.

e. VERNIER OR ALINEMENT. (1) One test was included that used this type of object which has been thought to measure a distinct aspect of visual acuity. The design of the items was again in a form which offered the examinees one of four possible choices of the correct response.

(2) *Vernier Acuity* (see fig. 13, Test 14). Each item was an equal-armed cross placed in the upright position. One half of one of the four arms of each was displaced laterally in a clockwise direction. The degree of displacement decreased by groups of items. The first two items had equal displacements, while the rest of the displacement decreased after each group of four items. The position of the arm which was displaced varied among the items on the basis of chance.

f. PRACTICE. It was essential to the value of the tests that the examinees continued to respond even though they reached a point where their reports were guesswork. This was especially true of those tests in which they reported in which of four possible positions the test object was seen. To train the examinee to continue to the point of "playing their hunches" a practice test was given at the beginning of each examination session. This test had the added effect of instructing the men on the mechanics of taking the tests.

(1) *Practice* (see fig. 14, Test 1). The test object was a gap in a modified form of a Landolt ring. The square C was placed in the diamond position so that the position of gap was distributed among the items on a chance basis. The size of the squares and the width of a stroke and gap decreased by groups of items. This test object may be considered either as a letter type on the basis of its C characteristic or as a purer resolution type since the same shape is used repeatedly in the test.

(2) *Scaling of items.* The test charts were constructed on two principles of item scaling, i. e., the physical steps of decrement in size,

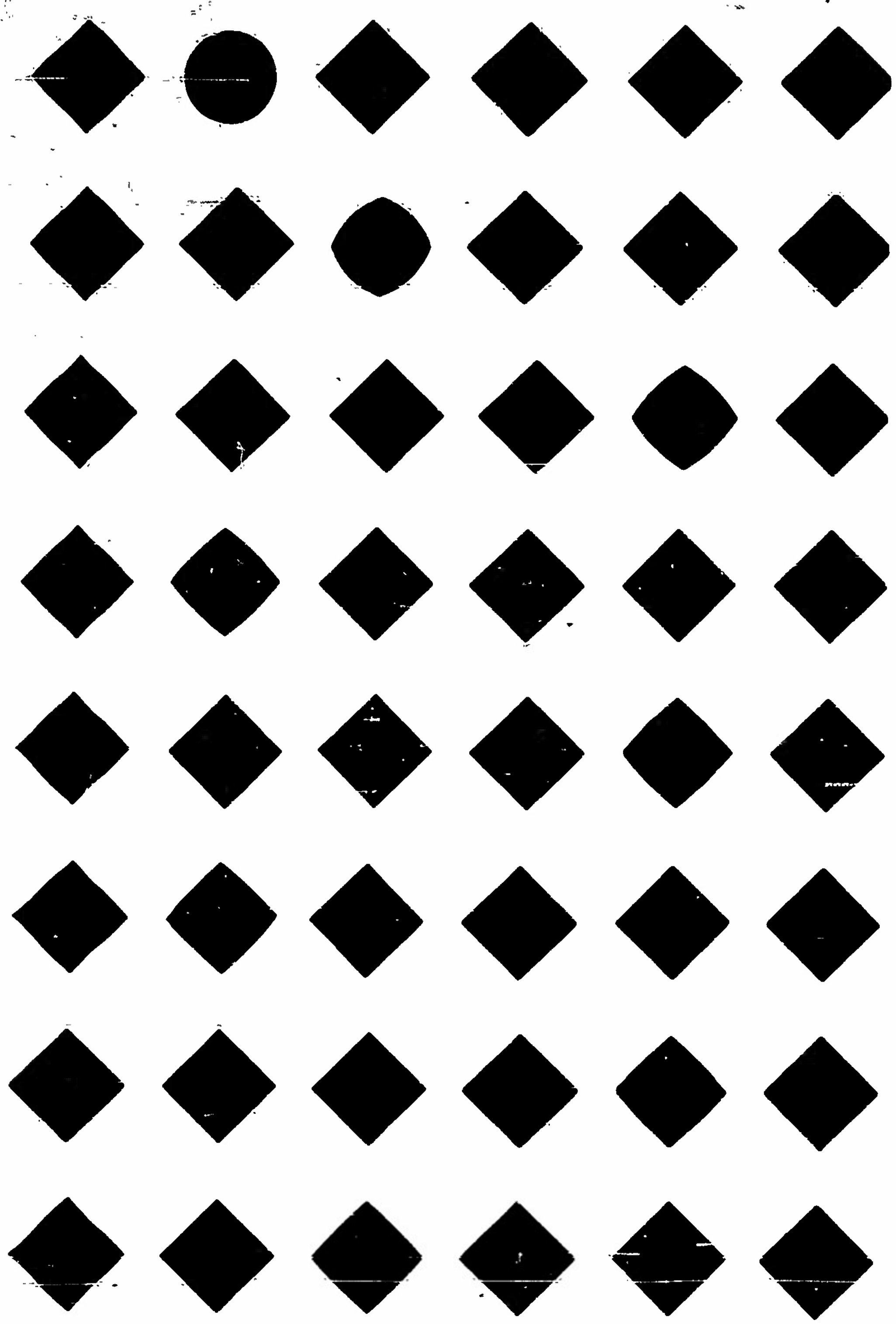


FIGURE 12.—*Square Discrimination—Test 12.*

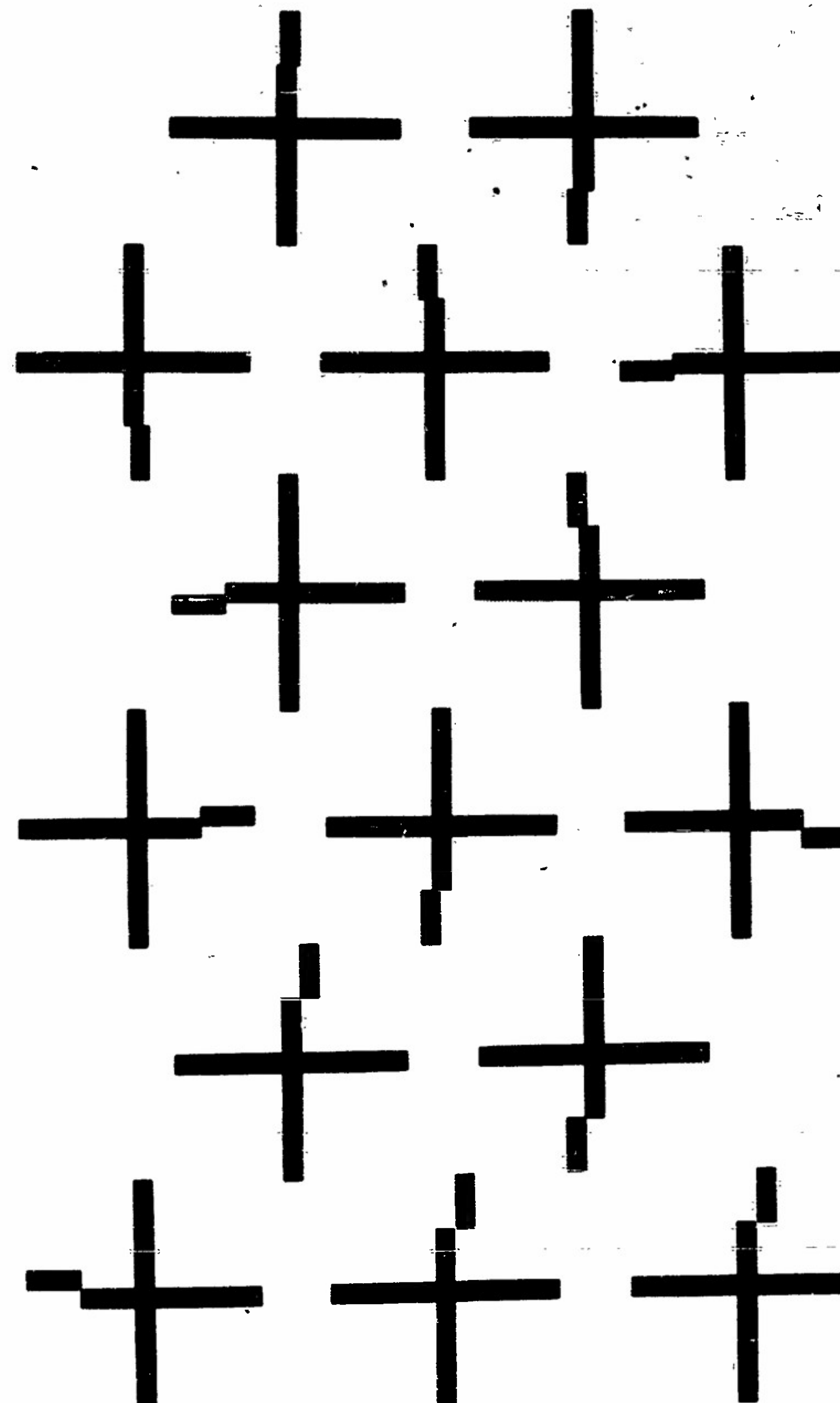


FIGURE 13.—*Vernier Acuity—Test 14.*

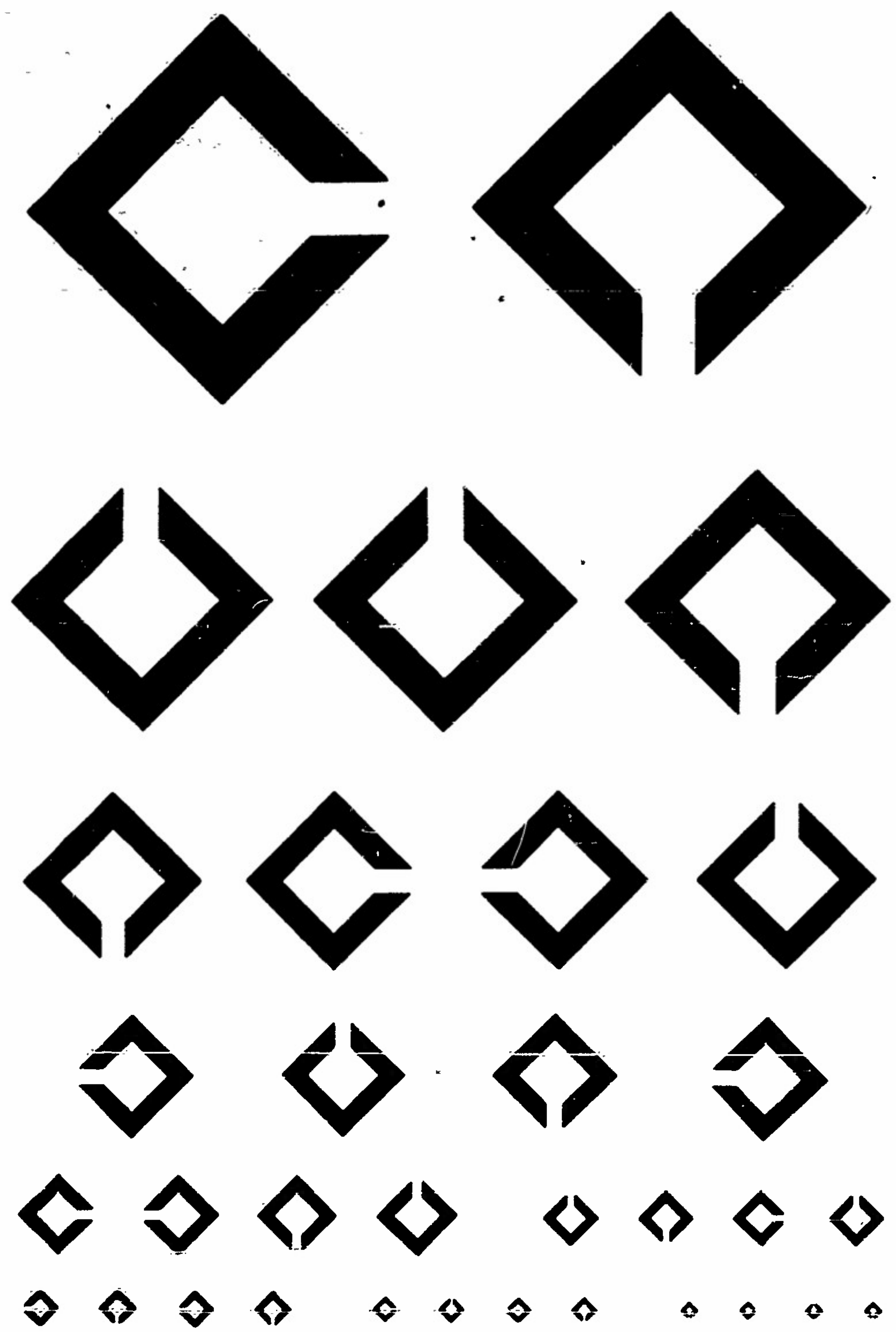


FIGURE 14.—*Practice Test (Square C)—Test 1.*

contrast, convexity, and alinement that were used. The two principles were Snellen-type units and logarithmic units.

(*a*) *The Snellen unit* bears a relationship to normal acuity which, theoretically, has been set as correct resolution of two points separated by 1 minute of visual angle at a distance of 20 feet from the examinee.

Hence, correct readings of letters whose critical spacings and width of stroke subtend 1 minute of visual angle at 20 feet are said to represent normal vision. Of course, populations have been tested where the average of the populations read correctly below this size of letter, but the Snellen scaling still depends on the concept. In the familiar fraction 20/20, the numerator indicates the distance of 20 feet from which the letters are observed while the denominator gives the size of the letters in terms of the distance at which the width of stroke and sizes of the gaps in said letters subtend 1 minute of visual angle. When the letters of a test are larger than the 20/20 ones, their size is determined as that which will subtend the same 1 minute of angle at greater distance. Thus the 20/40 Snellen letter, while observed from a distance of 20 feet, given by the numerator of the fraction, is increased in size over the 20/20 letter to a degree where its parts would subtend 1 minute at 40 feet, given by the denominator. Likewise, smaller letter sizes are based on a shortened observation distance at which the parts of the letter would subtend 1 minute of angle. The 20/15 letter size is that subtending 1 minute at an observation point 15 feet away. It readily can be seen that this unit depends on an arbitrary definition in terms of observation distance and depends on the assumption that so-called normal acuity is constant at any distance.

(*b*) *The logarithmic unit* employed in a majority of the tests used in this study is a function only of the visual angle. The test objects decrease in size in steps which bear a logarithmic relationship to the visual angle. This scaling unit assumes that perceivable increments in discrimination bear a constant relationship to the point of change. With this constant arbitrarily set at two, an object subtending 4 minutes is reduced by ½ to 1 subtending 2 minutes of visual angle in a decrement of one unit. The next step in the scale becomes 1 minute and 0.500, 0.250, and 0.125 minute of angle follow in that order. This scale can be extended beyond the range given here and can be made more discriminative by computing the intermediate angles.

(*c*) *A detailed tabulation* of the units employed in both the Snellen type and logarithmically scaled tests with the physical specifications of the items of all tests will be found in appendix A.

(3) *Decrements between items and between groups of items.* In addition to the two scaling principles employed, two methods of presentation of items were used: First, a single item was used at each scale value, and second, several items were used at each value. The AAF Constant Decrement test is an example of the first method in which each successive letter was reduced in size by one Snellen step. On the other hand, in the AAF Letter test six letters were presented at each value of the logarithmic scale. We have named the two methods (*a*) item to item decrement, and (*b*) group to group decrement. In most cases the group to group decrement was accom-

plished from row to row of the test chart. This did not hold for the Practice and Vernier Acuity Tests where the size of the test objects and the necessary spacing on the chart caused the decrement to occur within some of the rows.

(4) *Number of charts per test.* Limitations in the size of the charts (19 by 13 inches) led to the use of two charts for certain tests where the complete range of the items could not be included on a single chart. The second chart in each of these cases was a continuation of the first beginning in the top row with the scale value following that of the last row of the first chart. It might have been more desirable to present all the items of a single test on one chart since the items of the second chart confronted the examinee with a more difficult range which could conceivably influence the response. However, this would have entailed redesign of the entire apparatus and procedure and, at the time, was considered of slight influence on the present study.

TABLE II.—*Gross characteristics of tests employed in a study of the factors in visual acuity, PR 4075-01*

Test No.	Title	Type of item	Scale	Decrement	Number of items	Number of charts
1	Practice	Resolution	Log	Group	33	1
2	Army Snellen	Letter	Snell	Group	45	1
3	Dot Variable Size	Resolution	Log	Group	48	2
4	Quadrant Variable Contrast	Brightness	Log	Item	10	1
5	New London Letter	Letter	Log	Group	82	1
6	Triangle Discrimination	Form	Log	Group	20	2
7	Bausch and Lomb Checkerboard	Resolution	Log	Group	42	2
8	AAF Letter	Letter	Log	Group	66	1
9	Line Resolution	Resolution	Log	Group	48	2
10	Dot Variable Contrast	Brightness	Log	Item	10	1
11	AAF Constant Decrement	Letter	Snell	Item	36	1
12	Square Discrimination	Form	Log	Item	16	2
13	Checkerboard Variable Grid	Resolution	Log	Group	48	2
14	Vernier Acuity	Alignment	Log	Group	30	2

(5) *Controlled variables in the checkerboard tests.* The Bausch and Lomb Checkerboard and Checkerboard Variable Grid Tests presented a unique problem in variable control. From a study of the initial versions of these tests which were printed in toto it was apparent that two modes of response were possible in the case of some items: (*a*) A response based on seeing the checkerboard square of the item as a checkerboard, and (*b*) a response based on seeing it in terms of brightness. By shifting the response, after looking for actual checkerboards failed, it was possible to go on through a considerable portion of the tests on the basis of which square of the diamond looked darker. To control this an extensive, if rough, test was made of all the items used in the two tests in question. First, the items were produced individually. A sample of each size was placed at a distance from which resolution of the checkerboard was no longer possible, according to the report of special examinees used in this trial. By turning the items so that the correct response varied in position on a chance basis and recording responses of several men, it was possible to determine items on which all four alternatives obtained about the same number of responses and were therefore least subject to correct or incorrect responses on the basis of brightness judgments. These selected standard items were then used to pick others of the same size by a matching technique. This method was rough, but later results with the charts and remarks of examinees indicate that brightness judgments were to a large extent eliminated.

(6) *Summary of test chart characteristics.* Table II summarizes the salient characteristics of the tests in the order in which they were presented in the visual examination. More detailed information may be consulted in appendix A.

2. Test Rooms

The visual examinations were administered in six test booths which were close to identical in all features of construction and illumination. Such variations as existed was a function of the available materials and minor differences in construction. Prior to actual testing, the examinees were oriented and adapted in rooms specially prepared for these purposes. The entire installation was included in a mess hall at the Reception Center, Fort Dix, N. J., which was modified for the purpose of this experiment. A floor plan of this building giving the essential layout of facilities is shown in figure 15. This plan shows a central waiting room, used for orientation and interview of examinees, from which passages lead to two adaptation rooms (A1 and A2). From these, other passages lead to three test rooms in either wing of the building (E1, E2, E3 from A1 and E4, E5, E6 from A2). Specifications for all these rooms were determined through consultation with members of the Vision Committee and extensive trial construction with the materials provided.

a. WAITING ROOM. This room was supplied with indirect lighting but otherwise contained no special features. It accommodated 30 men at a time and was equipped with tables for interviewing purposes.

b. ADAPTATION ROOMS. Experimental controls were first exerted at the adaptation rooms where, while awaiting the examinations, examinees were exposed to controlled visual conditions.

(1) The rooms were about 10 feet long, 5 feet wide, and 8 feet high.

791639—48——3

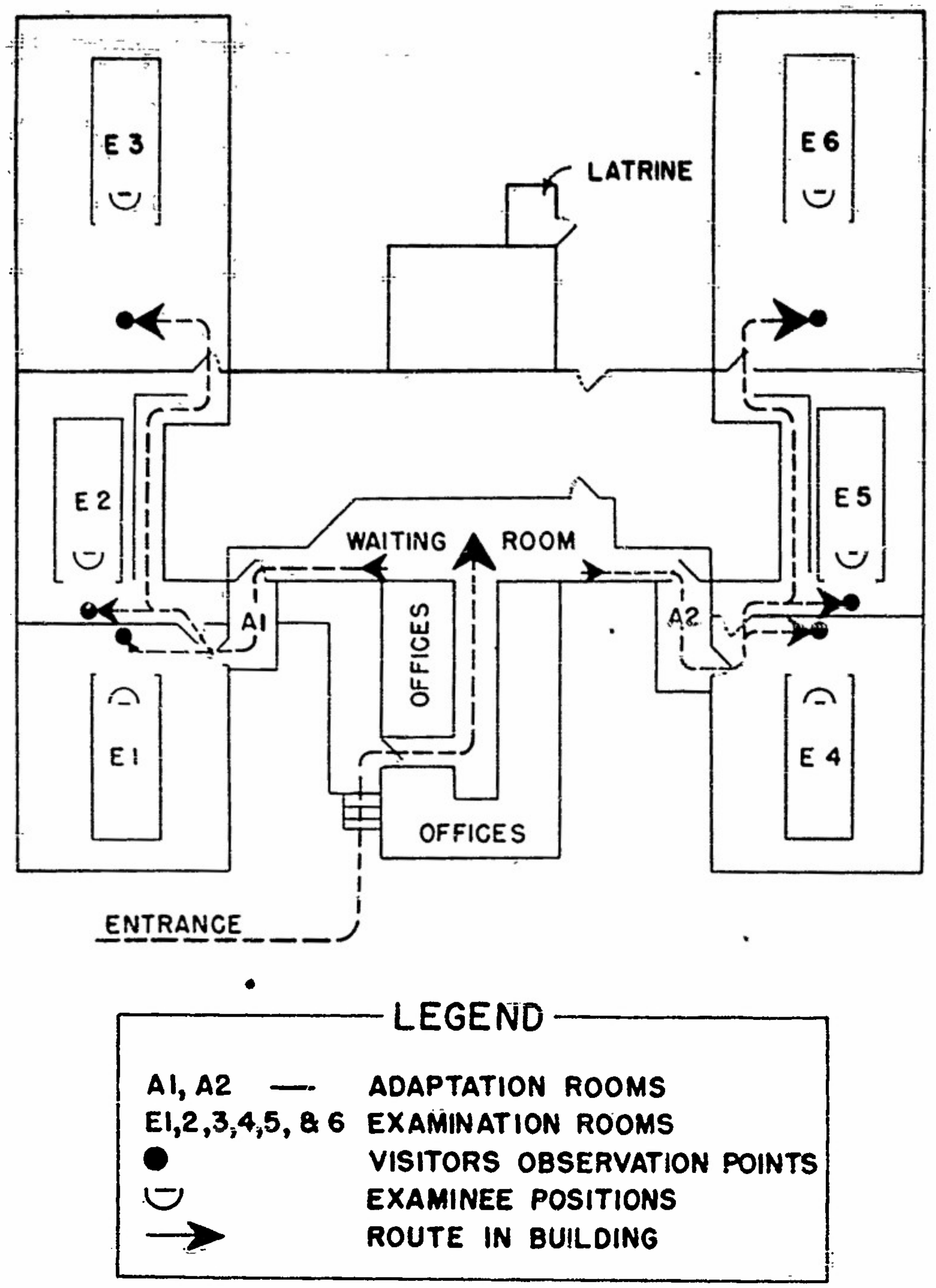


FIGURE 15.—*Schematic diagram, Vision Examination Building, Fort Dix, N. J.*

With both doors closed, they were virtually lightproof with respect to outside illumination.

(2) The walls were painted flat yellow. All sources of glare and visual microstructure were eliminated.

(3) A centrally located overhead lamp (100-watt) was shielded from direct vision of the examinees by a special reflector. This was the only source of illumination in the room.

(4) The walls at a 4-foot height from the floor measured approximately 4 foot-lamberts as measured by the Macbeth Illuminometer calibrated to measure brightness. This brightness complied with the plan to have it fall between 2 and 5 foot-lamberts.

(5) Three chairs were installed for examinees awaiting examination.

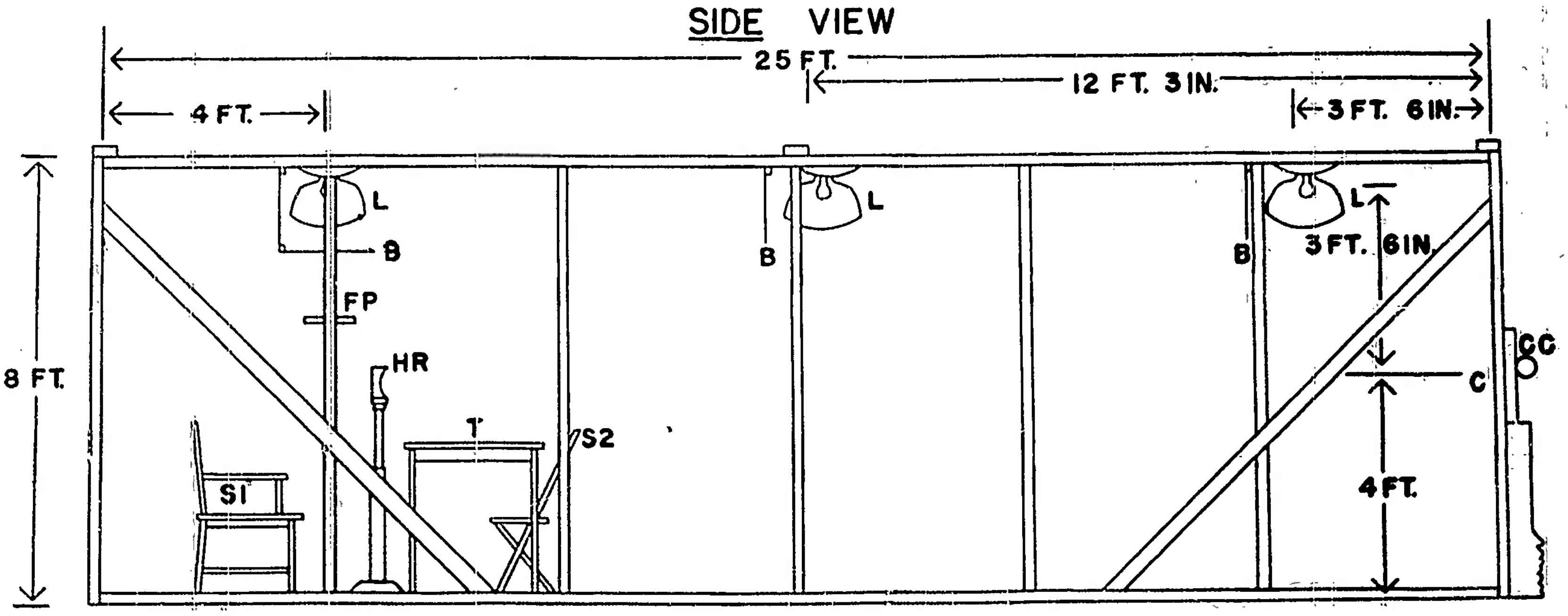
(6) Air circulation was accomplished by a fan mounted on a corner pedestal while the rooms were closed. Between test sessions, the doors to the waiting room were opened.

c. TEST BOOTHS. It was at the booths that the most exacting physical conditions were established and maintained. The booths were constructed of 2 by 4 lumber according to the plan shown in figure 16. This construction was accomplished within six sections of the building separated by sheetrock partitions. The windows and doors of these outer rooms were either walled off completely or shielded so that outside light would not fall directly on the walls of the test booths. The shields were movable plasterboard and plywood baffles placed as close to the windows as was required to avoid incidence of daylight on the booth walls. They did not prevent flow of outside air into the rooms. The following points cover the specifications of construction, furnishing, and illumination used in the booths proper:

(1) *Dimensions.* To inside dimensions, the booths were 25 feet long, 10 feet wide, and 8 feet high.

(2) *Wall covering.* The inside walls were covered with osnaburg material of a special quartermaster lot located for this purpose. A yellow-white in color, the cloth was originally specified as having a reflectance value of about 40 percent. This value was found variable depending for the most part on the degree to which the cloth was stretched in application to the walls and variations in the weave of different pieces. These provided no serious obstacle to achieving standard lighting conditions.

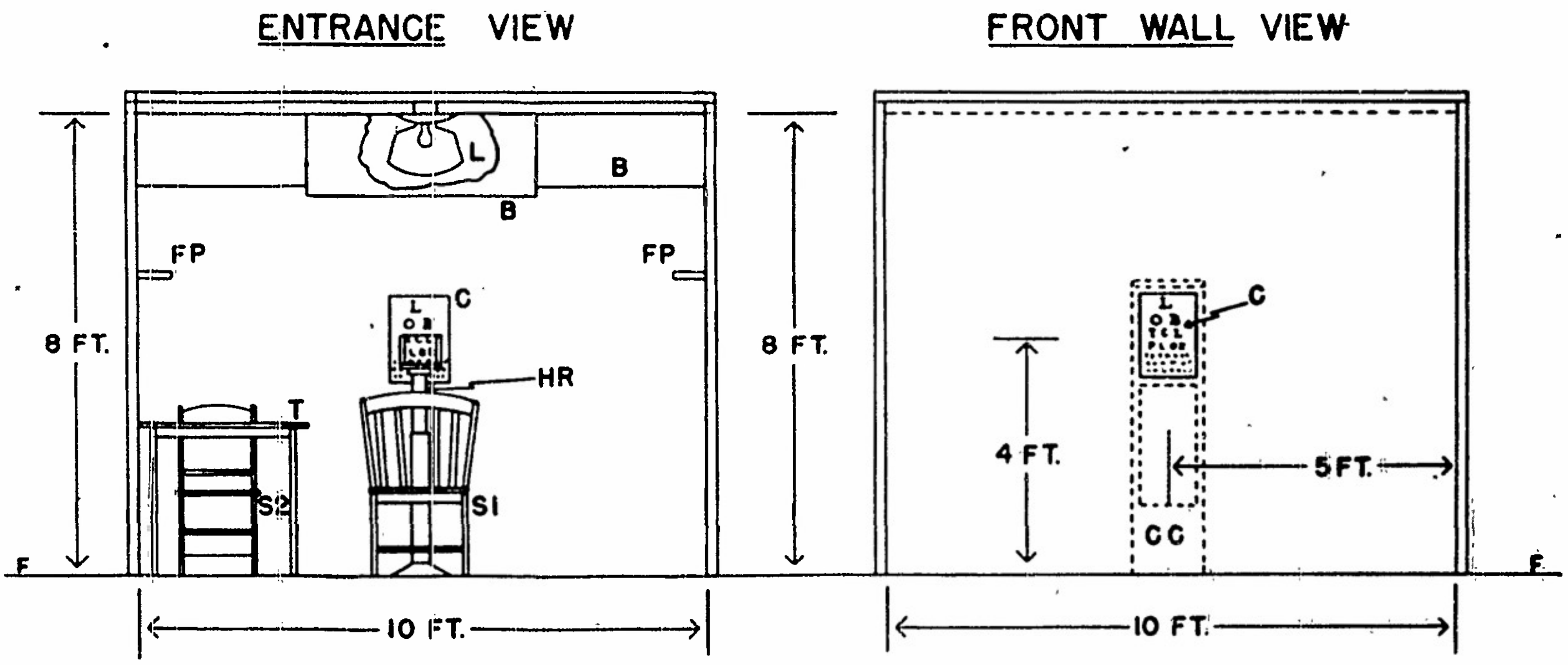
(3) *Lighting.* Placement of the lights in the booths and the dimensions of booth walls were inter-related in obtaining the desired brightness of the critical visible areas. For the booth dimensions used, placement of three 200-watt lamps at specified points on the center line of the ceiling accomplished the desired effects. The specific points where lamps were installed measured 3.5, 10.25, and 21.0 feet from the front wall as is shown in figure 16. Each lamp was enclosed by a standard 14-inch diffusing globe. Each of the two lights nearest the front wall was shielded from the direct vision of the examinee by drops from the ceiling placed about 2 inches behind the diffusing globes and extending the width of the booth. The baffles were made of several thicknesses of cloth and dropped to a point where direct vision of the globes was impossible. They also served as reflectors and raised the brightness of the front wall to the desired height. The light directly over the examinee was similarly shielded from his vision while entering the test booths. In addition, a distracting glare from overhead was



LEGEND

C - TEST CHART
HR - HEAD REST
B - LIGHT BAFFLE
FP - FAN PEDESTAL
L - LIGHT
T - TABLE
SI - EXAMINEE'S CHAIR
S2 - EXAMINER'S CHAIR
CC - CHART CHANGER

FIGURE 16.—*Plan of booth construction for experimental study of visual acuity tests.*



LEGEND

C - TEST CHART
HR - HEAD REST
B - LIGHT BAFFLE
FP - FAN PEDESTAL
L - LIGHT
T - TABLE
SI - EXAMINEE'S CHAIR
S2 - EXAMINER'S CHAIR
CC - CHART CHANGER

FIGURE 16.—*Plan of booth construction for experimental study of visual acuity tests.*—Continued.

eliminated by turning this baffle toward the front wall and stretching it between the diffusing globe and the examinee's head.

(4) *Lighting conditions obtained.* The booth construction described above produced brightness of all critical areas which approximated the standards adopted by the Subcommittee on Procedures and Standards for Visual Examinations of the Army-Navy-NRC Vision Committee. The subcommittee's standards proposed, "the brightness of the walls of the testing room at head height shall be not less than 3 foot-lamberts nor greater than the brightness of the test charts", and "the brightness of the chart shall average 12 foot-lamberts and shall not be less than 10 nor more than 15 foot-lamberts. Under no conditions shall there be shadows or reflections visible on the chart."

The degree to which these specifications were approximated may be noted from the measurements given in table III. The table gives the range of brightness found for the top and bottom sections of a plain white chart in position for testing in each of the test booths. It also gives the mean brightnesses for the top and bottom of the chart and the means for the total chart. All the values given are in foot-lamberts.

TABLE III.—*Ranges and means of brightnesses in foot-lamberts found at the top and bottom of a plain white test chart in each of the test booths*

Booth No.	Chart top			Chart bottom			Total chart
	Range		Mean	Range		Mean	Mean
	High	Low		High	Low		
1	14.0	10.0	12.4	12.4	9.8	11.4	11.7
2	14.0	10.5	12.5	12.3	9.5	10.9	11.8
3	13.8	11.0	12.1	11.7	9.2	10.4	11.3
4	13.9	10.5	12.5	12.2	9.9	11.1	11.8
5	13.0	10.4	12.2	12.0	9.6	10.9	11.7
6	13.0	9.9	11.9	11.7	8.9	10.4	11.2

The table shows that on the average the desired chart standard (12.0 foot-lamberts) was approximated. However, considerable variation occurred, especially from the top to the bottom of the charts, with the bottom always lower in brightness. This was a function of the overhead lighting and was much more acute when the booths were first erected. The difference from top to bottom was partially corrected by laying cloth on the floor from the front wall to a point about half way to the headrest. This arrangement succeeded in raising the brightness at the bottom to a greater extent than at the top of the chart. In addition, a black shape was painted on the diffusing globe of the light closest to the front wall so as to reduce the brightness of the top of the wall to a greater extent than the bottom. Although these devices were never as successful as desired, the correlation was as complete as could be obtained without an entirely new design for the booths. In practice, the walls of the booths closely approximated the desired specifications. At no point visible to the examinees did the brightness of booth walls exceed that of the test charts or fall below the prescribed minimum value of 3.0 foot-lamberts. There was a constant gradation of wall brightness from about 8.0 foot-lamberts at the area next to the charts to about 4.0 foot-lamberts at the rear of each booth. Behind the mid-booth baffle (light shield) a noticeable shadow appeared measuring 3.0 foot-lamberts in the corners formed by the baffle, ceiling, and side walls. This shadow, falling as it did well outside of a 10° visual angle around the center of the test charts, was discounted as effecting the examinations.

(5) *Chart changers.* The Bausch & Lomb Optical Co. produced an electrically operated chart changer which held the charts in position for testing and permitted the examiner to change them by remote control. The 1/8-inch stock of the front of the changer was sewed and glued to the cloth of the front wall of the booth so that the opening through which the charts were viewed cast a minimum of shadow on the charts. The changer had a capacity of 8 charts; thus, with 23 charts in the battery, 3 separate loadings of the changer were necessary. Because of operating difficulties connected with the chart changer, and the need for strict control in the order of presentation, a booth attendant whose main job was to oversee and service the chart changer was assigned to each examination unit.

(6) *Furniture.* Each booth was equipped with a chair and headrest for the examinee and a table and chair for the examiner. The headrest was designed to fix the position of the chin and forehead of the examinee so that the eyes were 20 feet from the charts and the head was erect and could not turn from side to side. Mounted in a pipe base fixed to the floor, the wooden rest could be raised and lowered to fit the individual examinee. The examinee's chair was of standard height (seat 18 inches from the floor), and during the test was pushed against the base of the headrest. In this position, making use of the arms of the chair for support, the examinee could maintain an erect posture with his head in proper position without undue strain. Meanwhile, the examiner sat at the table, a standard Army field table, in a position to the left and slightly forward of the examinee. From this position the examiner could conveniently give instructions and observe the activity of the examinee's left eye during the test. The headrest was cut away to facilitate observation of the left eye.

(7) *Ventilation.* The entrance to each booth was partially open at all times. Two fans, mounted on pedestals on either side of the examinee and directed toward the front of the booth, provided adequate circulation without billowing the cloth of the walls and baffles.

3. Other Examining Equipment

a. SAMPLE ITEMS. Each examiner was supplied with a set of 14 cards on which 2 or more sample items were presented for each of the experimental tests. These were exposed to examinees along with the instructions for each test and were designed to facilitate getting the test task across with a minimum of verbal explanation. The sample items did not duplicate the actual test objects or spatial arrangements of them and provided no observable cues to correct responses.

b. OCCLUDERS. A flesh-colored plastic occluder, held in place by elastic string circling the head, was used to cover the right eye during examinations. It fitted the face with comfort and was concave away from the lids and lashes. Its most notable characteristic was admission of general light to an open right eye while eliminating test objects from the visual field of that eye. As a result it was possible to present relatively equal amounts of light to both open eyes and to avoid touching the lashes of the occluded one and other variations existing in the use of various hand-held devices.

Section IV. PROCEDURES

The procedures adopted for testing and data collation were designed to routinize standard treatment of the hundreds of cases collected. These procedures were focused on administration of the 13 experimental tests and the practice test, to each examinee, either once or twice, maintaining suitable experimental controls. A staff was selected and trained to carry out the following sequence of operations:

Procurement of examinees.
Scheduling groups to be examined.
Scheduling examination procedures.
Preparing the test booths.
Orienting the examinees.
Interviewing.
Collection of personal data from individual records.
Light adapting the examinees.
Examining.
Scoring.
Collating the data.
Coding all data.

1. Selection and Training of the Staff

a. SELECTION. Twenty-five enlisted men were selected for the experimental staff. It was decided that the group should have had no previous experience in vision testing. Recognizing the possible existence of poor testing methods on the part of men with previous vision-testing experience, we preferred to train men from scratch. Initial selection was made on the basis of AGCT score, education, voice quality, a rough estimate of maturity, and interest to the point of voluntary choice of this assignment. The men selected on this rough screening were classified further by sampling their ability to read test instructions, to perform clerical or mechanical operations, and on new statements of their interests. In this manner five different jobs were filled with some subsequent reassignment and rotation of personnel to accomplish unforeseen activities of the study.

(1) Examiners, 14 men.
(2) Orienter, 1 man.
(3) Interviewers, orienter and 2 examiners.
(4) Booth supervisors, 6 men.
(5) Data clerks, 4 men.

b. TRAINING. One full week was devoted to training his staff. Training consisted of lectures, supervised practice, and, finally, 2 days of "dry-run" for the whole procedure. Greatest concentration of training went into the development of a professional attitude on the part of the staff. It was hoped that all personnel would appreciate the experimental nature of the work and maintain the controls by desire rather than by compulsion. Then too, we sought to develop ability and confidence to deal with people. The examiners and orienter were coached in reading instructions and the series of events in the test procedure. The booth supervisors were trained to operate and maintain all booth equipment and to make illumination surveys. Meanwhile, the clerks laid out the processes for scheduling all events and handling the data. As usual, plans made in the office had "bugs" when applied in the field. During the conferences of the training period, all procedures were whipped into more workable form. Instructions to examinees were reworded, mechanical failures turned up and had to be met by expedients, and clerical operations were simplified. In simplifying the job, while maintaining the basic design of the study, the enlisted staff proved invaluable. In the sections which follow the procedures described are those finally adopted.

2. Operations

a. PROCUREMENT OF EXAMINEES. Examinees were drawn from the processing stream of the reception center. The men were spotted at the physical profile section by reviewing age and previous left-eye Snellen scores recorded on WD AGO Form 63, "Report of Physical Examination." Selection of those to be tested was made in an attempt to fulfill the distributions of age and Snellen score previously described in section II of this report.

b. SCHEDULING GROUPS TO BE EXAMINED. Examinations were scheduled between the hours of 0730 and 1700. Men were scheduled to arrive in groups of 30 each, one group for the morning and one for the afternoon. The AM groups were tested only once and made up the sample of 531 cases we shall refer to as the TEST group. The PM groups, totaling 261 cases, were tested and retested in the same order on the following afternoon, approximately 24 hours after the first examination. This group we shall call the TEST–RETEST group. With 6 test booths operating and an AM or PM load of 30 men to be tested, it can be seen that 5 testing sessions were required in each part of the day. This schedule was based on an estimated 50 minutes for each session which turned out to be adequate. However, as experience was gained, the staff shortened this time to roughly 40 minutes for the longest examining sessions while some examinations of men with very poor vision required as little as 20 minutes. Nevertheless, to preserve order each session was started simultaneously by all examining units so the over-all time remained pretty much the same.

c. SCHEDULING EXAMINATION PROCEDURES. (1) *TEST group.* As we have seen, each examinee was given the practice test, test 1, followed by the 13 experimental tests. In this procedure an important control was employed. The order in which the tests were administered was systematically randomized among the men tested to equate the effect of fatigue upon each test. The tests were always presented in the order of their listing in table II (see sec. III) but the number of the first test was varied systematically. For men examined in the TEST group we had a running schedule for test order similar to the following sample:

AM TEST

Test sessions	*Test order*
1	2–3–4–5–6–7–8–9–10–11–12–13–14
2	3–4–5–6–7–8–9–10–11–12–13–14–2
3	4–5–6–7–8–9–10–11–12–13–14–2–3
4	5–6–7–8–9–10–11–12–13–14–2–3–4
5	6–7–8–9–10–11–12–13–14–2–3–4–5

As indicated, tests were presented in numerical order, but as each new examinee was tested the chart which had been first for the previous session was placed at the end of the stock. For the first session of the AM TEST on the day following our sample, the test order would be: 7–8–9–10–11–12–13–1–2–3–4–5–6, and so on. The same test-order schedule was preserved booth by booth, and over the entire experiment. This resulted in complete equation of the number of times each test was given first, second, and up to thirteenth.

(2) *TEST–RETEST group.* An identical schedule was kept for the TEST-RETEST group. In this treatment, however, another control was used. Men were retested with the same order of test presentation as that of first test session. For this purpose the same schedule was retained for pairs of afternoons: Monday—Tuesday, Wednesday—Thursday, and Friday—Saturday.

(3) *Examiner rotation.* The examiners were changed from session to session, two being assigned to each of the six booths. This alternation of examiners was scheduled with special care for the TEST-RETEST sessions. A different examiner administered the RETEST in every case. This procedure was adopted on the assumption that a great source of unreliability in this form of testing is introduced by variation of examiners.

d. PREPARING THE TEST BOOTHS. Each test booth, described in section III, was prepared for the examination by a booth supervisor. The booth supervisor and the examiner in each booth worked as a team. In preparing the test booth for each session the supervisor checked the following:

(1) *Test charts.* The condition of the test charts offered the greatest mechanical problem of the study. After much trial and checking of different methods of storing the charts, one was adopted which gave

the least overnight warping. This storage became a continuous task for the supervisors. In addition, they constantly checked on the cleanliness and general physical condition of the charts, substituting new ones when necessary.

(2) *Chart changer.* The supervisor made certain that the changer would function as well as possible. Daily oiling and repair of malfunctions was backed up by constant supervision while the tests were in progress. The chart changer never worked in a completely automatic fashion. As a chart jammed or failed to fall all the way out of the exposure aperture, the supervisor hand operated the machine to expose the next chart. This was carried out with dispatch whenever it occurred and, while regrettable, did not give rise to excitement or serious hold-ups in test procedure.

(3) *Loading the changer.* Before each session, the changer was loaded with the first six or seven tests in the order prescribed by schedule. As testing progressed, charts for later tests were inserted behind those already loaded.

(4) *Examiner's materials.* The supervisor made certain of the examiner's materials prior to each session. This included the supply of occluders and pencils, arrangement of the sample cards, and instructions for the tests in the order scheduled for the ensuing session.

(5) *Spot check of illumination.* Supervisors and examiners were constantly alert to possible failures of critical areas. Such failures included lightbulbs, cloth rips and separations, and any visible shifts in the general level of brightness which might occur during the testing. If suspicions arose they called for more accurate meter survey before the testing was continued.

e. Orienting Examinees. Upon reception at the vision examination building the men were given brief orientation:

> You will take several eye tests this morning (afternoon). Some are like the tests you took when you entered the Army. Others are new. Some of these tests may be better than others. The only way to find out is to try them all and compare the results. That is why you have been asked to take these tests. The scores you make will not go on your record or affect your job in the Army.

f. Interview. Certain personal data items were obtained in an interview following the orientation. All such data were originally assembled on the IBM data form designed for the study. A copy is included in appendix B. Other sources of data were individual records, the examination procedure, illumination surveys, and purely administrative items such as dates. These will receive brief attention as we come to them. The data, other than vision test scores, included—

(1) Name of examinee.
(2) Army serial number.
(3) Age on last birthday.

(4) AGCT standard score.
(5) Number of school grades completed.
(6) Station code (all Fort Dix).
(7) Previous Snellen score (left eye only).
(8) Statement of the examinee as to whether or not he wears glasses (other than sun glasses) and, if so, whether habitually or for reading only.
(9) Estimated number of hours of rest in bed the examinee had the previous night.
(10) General health of the examinee estimated by himself as "excellent," "good," "average," or "fair."
(11) Date and hour of the test.
(12) Indication of whether first test or retest.
(13) Question of examinee's having taken drugs in the last 24 hours. The term "drug" as used here did not include standard medication such as aspirin, but referred to opiates, morphine, marijuana, belladonna, etc., and to heavy drinking. If the examinee had taken any of these drugs or had been drinking heavily within the last 24 hours he was dismissed without examination.
(14) Number of test taken first.
(15) Foot-lambert reading at center blank white chart taken from periodic measurement.
(16) Examiner's name.
(17) Indication of astigmatism (not taken until the study had been started hence the total sample did not get this treatment).
(18) Test liked most (not recorded on the data form but coded from information gathered by examiners).
(19) Test liked least (not recorded on the data form but coded).
(20) Irregularities (coded from information supplied by examiners and including: none, power failures, illness, wrong test order, wrong spatial orientation of chart, occluder not fitting, outside disturbances, head cold, other).
(21) Question of whether or not the examinee's eyes were tired after the tests were completed (coded from information supplied by examiners including: no, left eye, right eye, or both).
(22) Race (all white in this study).
(23) Booth number (coded from administrative schedules).

g. Collection of Personal Data From Individual Records. As mentioned, some items of personal data in the list cited above were obtained from WD AGO Form 20, Soldier's Qualification Card, and WD AGO Form 63, Report of Physical Examination. These records were inspected by the men of our staff who selected and scheduled the men to be tested. This information was sent by name and serial number of the data section where it was collated with the other items for each examinee.

h. Light Adaptation. After the interview, the six men to be tested in any session were divided into two groups of three, one for tests in either end of the building. Each group was led to the respective adaptation rooms, seated, oriented, and remained there for 10 minutes, at which time the examiners picked up their examinees, introduced themselves, and proceeded immediately and directly to the test booths. Examinees and examiners who had completed the previous session waited on the test booth side of the adaptation room and left the area as the new groups moved in for testing. The new orientation received by examinees at the adaptation rooms was as follows:

> You are asked to spend a few minutes in this room, which is lighted much the same as the test room, in order to get used to the change in light from the outside. While here, it is best not to read or stare at anything. Please do not smoke, as it is difficult to air this room which is used all the time. You will be told about the tests in the testing room. We thank you for your cooperation in this program to secure a better eye test for the Army. Please do the best you can in these tests.

i. Examination. Administration of the tests consisted of adjusting the examinee to the situation, giving general instructions, giving specific instructions for each succeeding test, exposing the tests, recording the responses, and, after all tests were given, questioning the examinee to gain his impressions.

(1) *Adjusting the examinee.* The examinee was seated and headrest adjusted to a position in which the plane of his face was roughly perpendicular to the floor. Examinees were urged not to hang themselves by the chin on the rest but to use it as a guide for positioning their heads. The occluder was then introduced and tried over the right eye. In the adjustment of the occluder examinees were told to be sure their lashes did not touch the inside during a blink. Accompanying these activities the examiner gave running instruction as follows:

> You will take a series of 14 eye tests. Please do your best in each test. The left eye will be the only one tested. This cover will be worn over the right eye for all the tests (give examinee occluder). Please keep it over your right eye during testing. Keep both eyes open at all times. Whenever you are being tested, you will also place your head against this headrest (explain as necessary). If at any time during the test, your eye feels tired or "waters," tell me, and we will stop for a rest. There will be a 2-minute rest period between tests.

(2) *General instruction.* With the examinee seated the examiner gave general instructions:

> Here is the way you take the tests.
>
> For each test, charts will appear at the other end of the room. In some tests letters will appear on the charts, and in other tests figures will be shown. Before each test, I will show you a sample letter or figure such as you will see on the chart when the test begins.

> You will be told what to look for in each test and how to tell me what you see. Always start each test with the top row of letters or figures. In one test, there is only one letter in the top row. In other tests, the top row will have as many as six letters or figures. Start with the letter or figure at the left end of this row and go from left to right across the chart. When you have finished the top row, begin at the left of the second row and again go from left to right. Continue in this way down the chart, always going from left to right across the rows as you do in reading.
>
> Each test gets harder and harder as you go through it. Always keep on trying. Guess if you are not sure. You will be surprised at the number of times a guess is correct. Continue to call out your answers to any test until I say "That is enough." Some tests have two charts to be shown one after the other. You will be told how many charts are coming before each test begins. If two charts are used in a test, the second will appear immediately after you have finished the first one. Go on immediately at the left end of the top row of the second chart and continue until I say "That is enough."

(3) *Astigmatic index.* With the examinee in position for testing, the astigmatism chart was exposed to the left eye. Examinees were asked to describe what they saw. An attempt was made to avoid suggesting the appearance of the radial lines on the chart by the form of questioning. Examinees were asked to report any differences that appeared, and, if so, how the radial lines differed. Finally, those who perceived differences were asked at what position of the clock (1, 2, 3, 4 o'clock) the lines appeared "darker," "more distinct," or the other qualities adopted by the individual to describe their experience. The nature of the data and the errors of measurement, e. g., head position, attendant to this rough index have led us to doubt the value of these data.

(4) *Specific instruction.* Instruction for each visual acuity test began by showing the sample items for the test and continued as in this quotation of the instruction for the practice test:

> Here is a sample of the figures you will see on the chart. Notice that one corner of each figure is missing. You are to tell me which corner of each figure is missing by saying: top, bottom, right, or left. The missing corner, gets harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row of figures. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."
>
> Are there any questions? Get ready. Keep both eyes open. Go ahead.

The specific instructions for all tests are shown in appendix D.

(5) *Exposing the charts.* Before the signal, "go ahead," was given, the examiner exposed the chart for the test. If the chart changer failed to function at this time, the "go ahead" was delayed while the booth supervisor hand-operated the equipment to make the proper exposure. This was usually accomplished in a matter of seconds and was followed instantly by the test. Similar delays were encountered

throughout the procedure and were handled with dispatch by the examiner-supervisor team. Such delays were infrequent and had no observable influence on the data. However, we cannot advocate such difficulties as conducive to standard practice.

(6) *Recording.* Recording responses was a three-part job, marking the IBM data sheet, periodic observing of the examinee, and supplying motivation.

(*a*) *Marking.* The examiner marked correct responses only on the section of the date sheet for each test. Electrographic pencils were used. Each test was continued until the examinee made three consecutive errors. No attempt was made to record the nature of the errors in the calls of the examinees.

(*b*) *Observing.* Examiners periodically observed the position of the examinee in the headrest correcting twists of the head as they occurred. More important, the left eye behavior was carefully watched for squinting, excessive blinking, watering, and other signs of fatigue. Tests were stopped when these signs were noted and suitable corrective rest or reinstruction were applied.

(*c*) *Motivating.* The test scores from this group of tests depended on the examinee's perseveration in the task through the "threshold." That is to say, the examinees had to continue to call some response even when the visual situation made it sheer guesswork. This presented a problem of urging the examinees to continue the task, but in a way that would be standard for all. In addition to the initial instruction to "keep trying" and "guess if you are not sure," standard methods were adopted in the test period. When an examinee halted on an item, 10 seconds were allowed for the response. If no response was elicited in that time, the examiner repeated, "keep trying, please, guess if you are not sure." Another 10 seconds were allowed for the response. If none was obtained, the prodding phrase was repeated. If in the 10 seconds following this second statement to the examinee he failed to respond, the test was stopped. It was assumed in treating the data of such men that the next three items, including the one on which the examinee halted, would be misses.

(7) *Rest periods.* A 2-minute rest period was allowed between tests. Examinees were instructed to lean back, remove the occluder to a position on the forehead, and not to look at the charts during this period. During this period examiners answered questions about the procedure without reference to specific test items.

(8) *Questions at the end of the session.* The examinee's responses to the following questions were recorded at the end of each session:

(*a*) Which test did you like most?

(*b*) Which test did you like least?

(*c*) Are your eyes tired? (If the answer was "yes," the examinee was asked if this applied to the left eye, right eye, or both.)

j. Collating the Data. The data from interviews, testing, and personal records were collated by the data clerks. The visual acuity tests were scored in two ways at this time: *number right to three consecutive misses* and *number of attempts.* In later treatments of the data other scoring formulae were investigated. These will be discussed below. All of the accumulated data were checked before the final field operation, coding.

k. Coding. All of the data were coded in the field for punching on IBM cards. The items so coded included the 23 points mentioned under *f* above, and the test scores. A sample code sheet which was used for this purpose is shown in appendix C.

l. Retest. All of the foregoing procedures were applied in the retest sessions except as noted. Another important variation in the retest sessions was the abbreviation of the orientation and test instruction. For these sessions, all but the most essential parts of the instruction was eliminated.

791639- 48 — 4

Section V. RESULTS AND DISCUSSION

Three major studies were made of the data collected at Fort Dix:

Determination of test-retest reliability by seven different scoring methods and examination of the relationship between methods.

Factorial analyses of three intercorrelation matrices to discover the aspects of acuity measured by the 14 wall chart tests.

Item analyses and study of frequency distributions.

Because of its implicit value in throwing light on the factors measured by combined wall chart and commercial devices as well as phoria and stereopsis tests, we also present factorial analyses of data collected at Submarine Base, New London, Conn. These analyses will follow the factor analysis of the 14 wall chart tests mentioned above.

Minor studies were made of the Fort Dix data to determine the influence of certain variables in testing procedure, the intercorrelation of AGCT and letter test scores, and the opinions of examinees about the tests.

1. Test-Retest Reliability by Seven Scoring Methods

The first step in analysis of the 14 wall chart tests was a determination of their reliability. Moreover, it was essential to find out which of several feasible scoring methods would yield the highest reliability. In the absence of other independent criteria we decided to adopt the most reliable and consistent scoring techniques as the basis for further analytic study.

From the 261 cases who were tested and retested 24 hours later reliability coefficients were computed in the formal Pearson product-moment r's. These are presented in table IV for each of seven methods of scoring studied. The methods, described briefly at the foot of table IV, were founded on considerations of the forms of the tests, methods of testing, and previous methods of scoring such devices.

The means and standard deviations for each test and retest by each of the scoring methods are given in table V.

Most of the tests show adequate reliability if a correlation of 0.80 or higher is accepted as adequate. From table IV we see that correlations differ considerably depending on the method of scoring employed. However, all but four tests have reliability coefficients of 0.80 or higher by one or another scoring method. The four tests which have unsatisfactory reliability in their present form are:

TEST 4. Quadrant Variable Contrast, with reliability coefficients ranging from 0.41 to 0.58 by the different methods of scorning.

TEST 6. Triangle Discrimination, with coefficients ranging from 0.65 to 0.75.

TABLE IV.—*Reliabilities of 13 vision tests and the practice test with 6 methods of scoring (N=261)*

[Showing test-retest correlations (Pearson product-moment) for 13 vision tests and the practice test by 7 methods of scoring. The tests were given 24 hours apart]

Test	Scoring method							Method adopted
	A	B	C	D	E	F	G	
1. Practice test	0.60	0.76	0.80	0.76	0.75	0.74	0.80	C
2. Army Snellen	.69	.84	.82	.78	.83	.80	.88	C
3. Dot variable size	.84	.86	.88	.85	.92	.85	.85	C
4. Quadrant variable contrast	.58	.54	.59	.53	.41	-----	.57	A
5. New London letter	.69	.84	.84	.81	.83	.85	.88	C
6. Triangle discrimination	.68	.73	.72	.65	.66	-----	.69	A
7. Bausch and Lomb checkerboard	.79	.80	.81	.76	.75	.76	.79	C
8. AAF letter	.78	.87	.89	.85	.88	.87	.89	C
9. Line resolution	.75	.84	.86	.83	.83	.83	.85	C
10. Dot variable contrast	.49	.45	.40	.32	.30	-----	.43	A
11. AAF constant decrement	.72	.83	.84	.82	.84	-----	.87	C
12. Square discrimination	.42	.42	.43	.37	.33	-----	.44	A
13. Checkerboard variable grid	.84	.82	.85	.81	.76	.83	.81	C
14. Vernier acuity	.74	.80	.76	.69	.73	-----	.80	C

Scoring methods

A. Number of rights to the first miss.
B. Number of rights to the first miss plus number of attempts.
C. Number of rights to 2 consecutive misses.
D. Number of attempts to 2 consecutive misses.
E. Number of attempts up to the first point following which 3 of 4 items are missed.
F. Line number after which at least 50 percent of the items are missed. Not used for tests 4, 6, 10, 11, 12 and 13.
G. Number of rights to 3 consecutive misses.

TEST 10. Dot Variable Contrast, with coefficients ranging from 0.30 to 0.49.

TEST 12. Square Discrimination, with coefficients ranging from 0.33 to 0.44.

These four tests are, interestingly enough, those with the fewest number of items (10, 20, 10, and 16, respectively). In addition, two of them, tests 6 and 12, give rise to distinct configurational effects as presently displayed on the charts. The items are undoubtedly too close together. In terms of the number of items, some improvement in reliability can be expected by increasing the test length, i. e., adding a number of similar items to each test. By means of the Spearman-

TABLE V.—*Means and standard deviations of scores on 13 vision tests and the practice test for test and retest and for 7 scoring methods (N=261)*

Test		A M	A SD	B M	B SD	C M	C SD	D M	D SD	E M	E SD	F M	F SD	G M	G SD
1. Practice test	Test	20. 53	6. 91	45. 59	12. 11	23. 23	6. 14	24. 11	6. 33	23. 76	6. 66	5. 07	. 94	23. 70	6. 05
	Retest	21. 82	6. 64	47. 25	12. 34	23. 61	6. 19	24. 40	6. 35	23. 88	6. 60	5. 12	. 96	24. 10	6. 08
2. Army Snellen	Test	23. 19	9. 86	55. 19	18. 49	28. 22	9. 50	29. 87	9. 71	29. 37	9. 68	7. 01	1. 74	29. 28	9. 38
	Retest	24. 01	10. 01	56. 69	17. 73	28. 96	9. 22	30. 67	9. 34	30. 32	9. 18	7. 12	1. 67	30. 08	8. 87
3. Dot Variable size	Test	17. 98	8. 88	40. 79	18. 29	20. 14	9. 25	20. 97	9. 59	20. 83	9. 70	4. 97	2. 37	21. 13	9. 44
	Retest	18. 27	9. 30	42. 12	19. 28	20. 77	9. 52	21. 74	9. 91	21. 47	9. 87	5. 08	2. 43	21. 83	9. 80
4. Quadrant Variable Contrast	Test	4. 12	1. 86	9. 59	3. 97	4. 55	1. 97	4. 85	2. 21	4. 25	2. 56	-----	-----	4. 77	2. 00
	Retest	4. 19	1. 80	9. 81	3. 73	4. 63	1. 93	4. 95	2. 16	4. 39	2. 61	-----	-----	4. 90	1. 88
5. New London Letter	Test	15. 59	15. 91	51. 64	29. 72	27. 82	15. 64	30. 08	16. 42	31. 97	16. 52	4. 67	2. 28	31. 38	15. 89
	Retest	18. 05	16. 28	56. 29	30. 06	30. 24	15. 90	32. 61	16. 66	34. 31	16. 65	4. 98	2. 29	33. 57	15. 98
6. Triangle Discrimination	Test	8. 82	4. 88	22. 73	9. 41	11. 07	5. 17	12. 07	5. 56	12. 13	6. 04	-----	-----	11. 71	5. 12
	Retest	9. 10	4. 85	23. 65	9. 49	11. 54	5. 29	12. 69	5. 72	12. 82	6. 18	-----	-----	12. 46	5. 00
7. Bausch and Lomb Checkerboard	Test	10. 43	5. 57	24. 69	11. 40	11. 77	5. 59	12. 45	5. 85	11. 90	5. 96	3. 28	1. 54	12. 61	5. 78
	Retest	10. 96	5. 66	25. 71	11. 47	12. 34	5. 64	13. 02	5. 90	12. 44	5. 74	3. 38	1. 55	13. 12	5. 77
8. AAF Letter	Test	22. 48	13. 34	56. 91	23. 81	28. 23	13. 42	29. 78	13. 78	30. 39	13. 78	4. 97	2. 33	30. 02	13. 50
	Retest	25. 12	13. 48	59. 96	26. 60	29. 95	13. 56	31. 33	13. 92	32. 02	13. 91	5. 26	2. 31	31. 85	13. 56
9. Line Resolution	Test	30. 81	12. 13	67. 84	22. 67	33. 97	11. 81	34. 95	12. 03	35. 11	12. 24	8. 43	2. 96	35. 04	11. 35
	Retest	31. 07	11. 98	69. 03	22. 14	34. 92	11. 37	36. 00	11. 60	36. 35	11. 96	8. 67	2. 90	35. 86	11. 16
10. Dot Variable Contrast	Test	2. 82	1. 58	7. 37	3. 58	3. 45	1. 85	3. 85	2. 22	3. 26	2. 63	-----	-----	3. 80	1. 90
	Retest	2. 97	1. 59	7. 59	3. 61	3. 51	1. 73	3. 89	2. 04	3. 22	2. 50	-----	-----	3. 81	1. 84
11. AAF Constant Decrement	Test	17. 82	9. 72	41. 88	17. 16	21. 17	9. 06	22. 18	9. 24	22. 11	8. 59	-----	-----	22. 22	8. 50
	Retest	18. 11	9. 76	42. 03	16. 98	21. 74	8. 74	22. 82	8. 95	22. 46	8. 44	-----	-----	22. 60	8. 28
12. Square Discrimination	Test	4. 85	2. 31	13. 17	5. 92	6. 03	3. 09	6. 63	3. 71	6. 52	4. 44	-----	-----	6. 87	3. 24
	Retest	4. 75	2. 54	13. 24	6. 41	6. 12	3. 34	6. 82	3. 97	6. 66	4. 68	-----	-----	6. 97	3. 54
13. Checkerboard Variable Grid	Test	12. 90	6. 11	29. 30	12. 86	14. 30	6. 43	14. 89	6. 75	14. 29	6. 89	3. 46	1. 58	15. 10	6. 70
	Retest	14. 06	6. 22	31. 44	13. 22	15. 21	6. 43	15. 82	6. 68	15. 33	6. 85	3. 64	1. 62	16. 00	6. 71
14. Vernier Acuity	Test	18. 52	5. 77	41. 47	10. 84	20. 43	5. 54	21. 40	5. 83	21. 04	5. 99	-----	-----	21. 08	5. 42
	Retest	18. 74	5. 97	41. 93	11. 34	20. 49	5. 91	21. 41	6. 21	21. 26	6. 44	-----	-----	21. 36	5. 59

Brown prophecy formula we can estimate the reliabilities of longer tests from the known reliabilities of the present versions. These estimates are given in table VI for tests 4, 6, 10, and 12. In this case we have estimated reliabilities for tests doubled in length and have used the highest reliability obtained for the present versions as given in table VI.

TABLE VI.—*Present test-retest reliabilities for 4 vision tests and estimated reliabilities for comparable tests of twice the present length, by Spearman-Brown prophecy formula*

Test	Present number of items	Present r's	Estimated r's
4. Quadrant Variable Contrast	10	0. 59	0. 74
6. Triangle Discrimination	20	. 72	. 84
10. Dot Variable Contrast	10	. 49	. 66
12. Square Discrimination	16	. 44	. 61

From the estimated reliability coefficients in the last column of table VI we see that doubling the length of test 6, Triangle Discrimination, raises its reliability to 0.84. This is more satisfactory than the present finding. As for doubling the other tests, some improvement is predicted but not enough to attain the arbitrary level we have set for satisfactory reliability, namely 0.80. However, evaluation of reliability solely in terms of the size of the coefficients omits several considerations of highly practical importance. Until validity studies, factorial analysis, and the practicability of constructing and administering revised tests are considered, the reliability of the tests cannot be thoroughly evaluated.

In terms of validity, some of the tests used in the current study may be expected to be more valid for given prediction purposes than others. If such proves to be the case, such tests may be useful even though a wide range of reliability is found for the present forms. Of course, an attempt should be made to improve tests with low reliability. Assuming that all present tests are equally valid, those with initially low reliability may be expected to show greater improvement in validity following an increase in their reliability than those whose initial reliability is high. For example let us assume validity coefficents of about 0.50 for all of the present tests. If the Quadrant Variable Contrast, with present reliability at 0.59, were doubled in length, we have seen that the new test would be expected to have a reliability of about 0.74. In turn we may prophesy the improvement in validity by formula. In the case of our example a validity of about 0.56 may be expected. Compare this with what may be expected in the way of improvement for a test with initial reliability of 0.80 and validity of

0.50. The validity expected in this case will be of the order of 0.53 while reliability shoots to 0.89. From this example, it is apparent that unreliable tests gain proportionately more when lengthened than do those with high reliability. It is apparent that validity should be considered along with test reliability.

A second consideration in evaluating reliability is the degree to which any test may be found to measure different factors or aspects of, in this case, visual ability. As will be seen in the discussion of the factorial analysis in the sections below, tests may be found which measure several aspects of visual ability as well as other unknown factors. If the goal of measurement is to measure such aspects in pure and reliable form to predict performance on the job, it may become important to revise selected tests to improve their prediction for certain purposes. Until we know for which aspect we need improved tests, an arbitrary size of reliability set at 0.80 is not too meaningful.

It is one thing to speak of improving tests and quite another to construct the new items or add similar ones to the scale. In light of the difficulties encountered in constructing and using the present forms a word of caution seems in order. Only by the employment of new and more exact techniques for scaling brightness discrimination items will mass production of such tests be accomplished. Large scale production of checkerboard items seems equally unfeasible by the methods presently employed. Finally, the present chart forms are felt to be a block to improvements in reliability because of limitations in size. Certain items such as the triangles, squares and vernier crosses were too close together. Increasing the space between items will lead naturally to problems of chart size. These points will be treated further in the discussion of item analyses below.

What of the selection of a scoring method for further analysis of the data? From table IV there seems little to choose between methods G and C with respect to highest reliability. Letter tests are slightly higher by method G, but other individual tests are slightly lower in reliability by this method. Method A gives the highest value for test 10 Dot Variable Contrast and about the same value as method C for test 4 Quadrant Contrast. In the case of these two tests as well as the square and triangle tests there is some logic in choosing method A because the items were presented one for each stimulus value. After much consideration, and perhaps for tenuous reasons, we adopted the methods shown in the last column of the table.

At the time of adoption method G had been overlooked. As a result, the two factorial analyses of Test and Retest intercorrelations were completed on the basis of the methods adopted as shown in table IV. The final factorial analysis of 531 cases who were tested only once was made of scores based solely on method G. As we shall see this made

little difference in the results of the analyses. Finally, item analyses and other reported studies were also based on the G method.

Some light is thrown on the comparative value of the scoring methods by an analysis of the intercorrelations between them. Although these are available in detail, they are considered too extensive for inclusion in this report. In general, scores by the various methods are highly intercorrelated. This is especially true between first-test scores by the different scoring methods and the retest scores by those methods. First-test against retest correlations were, of course, lower, representing error in measurement from test to retest.

2. Factor Analysis of 14 Wall Chart Tests

Three separate analyses were made of the intercorrelations (Pearson product moment r's) or the test scores. The distinction of the three analyses parallels that of the three experimental treatments: Test, Retest, and Test-only. Two intercorrelation matrices, Test and Retest, were computed from scores attained by 261 cases so treated. The third was computed from scores attained by 531 cases who were tested but not retested. Each original matrix is presented in tabular form with the factor loadings and residual intercorrelations when the factors were extracted. The three tables are:

Table VII. Intercorrelations (lower left of the matrix) Residuals (upper right), and Factor Loadings from analyses of 261 *first test* cases of enlisted men tested twice at Fort Dix, N. J.

Table VIII. Intercorrelations (lower left of the matrix), Residuals (upper right), and Factor Loadings from analyses of 261 retest cases of enlisted men tested twice at Fort Dix, N. J.

Table IX. Intercorrelations (lower left of the matrix), Residuals (upper right), and Factor Loadings from analyses of 531 cases tested once at Fort Dix, N. J.

Table X. A summary of the factor loadings of all three analyses bringing them together for discussion purposes.

In factor analyzing these data a group factor method was used to derive the aspects which make meaningful the relationships represented by the original intercorrelations. In this process a minimum of rotation was required (at most two rotations) to reduce the loadings on orthogonal axes to simple meaningful structure. In all 3 analyses 4 factors were discovered which account in varying degrees for the variance in test scores on all 14 tests (practice having been included) other than that contributed by specifics, which are unanalyzable in the current residual matrices, and error or lack of reliability. Each factor will be discussed in turn.

a. Factor I—Retinal Resolution. Scores on all tests are dependent in general on this factor. Loadings on this aspect of visual acuity range in size from 0.34 to 0.93 for all three analyses and variously account for from 11.6 to 86.5 percent (square of the loadings) of the variance in test scores. The tests having been rearranged for purposes of segregating those which cluster with highest loadings on each factor, it can be seen that the first eight tests (7, 13, 5, 8, 2, 11, 3, and 9) have the highest loadings on factor I. Included in this group we find that two *checkerboard* tests, all four *letter* tests and the *dot* and *line resolution* tests. In these tests, with only slight differences among the three analyses, factor I accounts for from 67.2 to 86.5 percent of the variance in score. Relatively high loadings are also found on the practice (1) and vernier (14) tests (51.8 to 67.2 percent of the variance.) The other tests, the two contrast and square and triangle discrimination tests (4, 10, 6, 12) also show up with significant loadings on this factor. The amount of variance accounted for is lower in these cases (11.6 to 46.2 percent, but in only two instances, test 12 for the first test and retest samples, does the variance attributable to any other factor exceed that found in factor I.

What is this aspect that so generally accounts for variance in score on all the present tests? It looks like a general ability to discriminate no matter what is being seen. It may be called ability to discriminate between a differentiated and an undifferentiated area of the visual field. Since the tests for which this factor explains the greatest amount of variance predominantly involve what has been called retinal resolution, this name is suggested. However, before adopting the title, it is wise to examine another rationale. The two checkerboard tests (7 and 13) were standardized before their use in the present study to eliminate brightness differences of one or another part of the items as cues to correct responses. Rough as the equation for brightness may have been, it seems to have worked pretty well. For, without any rotation of the orthogonal axes, these tests were highly loaded on factor I while showing near zero or slightly negative loadings on factor II (brightness). This is additional evidence that this factor is one involving discrimination of spatial pattern and it is called retinal resolution.

(1) *Test purity*. The degree to which any tests measure retinal resolution, more or less free of other factors, requires a look at the loadings on other factors. We have found that the checkerboard tests (7 and 13) are free of the brightness aspect. It also appears that these two tests show near zero loadings on the other two factors discovered. All other tests show some contamination by other modes or aspects of response. From these analyses it is concluded that these two checkerboard tests measure retinal resolution in greatest purity.

Practically speaking, the present analyses offer some interesting suggestions on how to test retinal resolution in pure form if such is desired. First, it is obvious that just any old chart will not do. Some tests which have been construed to measure this function have load-

TABLE VII.—*Intercorrelations table, residual matrix and factor loadings of 14 visual acuity tests given twice to 261 EM at Fort Dix, N. J.—First test.*

RESIDUALS (above diagonal) / INTERCORRELATIONS (below diagonal)

Tests	7	13	5	8	2	11	3	9	4	10	14	6	12	1	I Reso-lution	II Bright-ness	III Letter Form	IV
7 Bausch and Lomb Checkerboard		-01	00	-02	03	-03	00	01	-01	01	-03	02	02	-02	90	-01	-03	-01
13 Checkerboard Variable Grid	80		-04	-03	02	-03	-02	-02	01	01	00	-01	02	-04	90	01	-03	-09
5 New London Letter	75	70		00	01	-01	01	00	04	-01	-01	06	-02	00	84	11	21	08
8 AAF Letter	77	76	82		-01	00	00	-01	02	-01	00	01	03	-02	89	09	29	01
2 Army Snellen	78	78	81	84		01	-02	00	00	00	02	01	-01	02	85	28	24	01
11 AAF Constant Decrement	74	75	78	85	85		-01	03	01	-01	00	02	00	00	87	15	24	-01
3 Dot Variable Size	78	76	79	83	83	81		00	01	-02	01	-02	01	-02	87	31	08	02
9 Line Resolution	74	73	73	76	82	80	85		-01	01	01	-02	01	03	82	44	-01	-05
4 Quadrant Contrast	48	52	54	54	61	56	65	69		02	04	02	-02	01	55	55	-06	-08
10 Dot Contrast	49	48	52	53	59	54	59	63	53		-04	01	00	-03	54	41	06	18
14 Vernier	61	64	65	69	73	69	73	73	59	48		-01	06	05	72	29	08	04
6 Triangle Discrimination	58	53	66	62	65	63	63	60	49	52	54		00	02	63	28	09	29
12 Square Discrimination	37	34	37	43	41	39	43	40	26	38	43	45		02	40	19	08	45
1 Practice (Square C)	69	68	73	76	80	77	75	78	56	49	70	59	38		80	21	18	-03

FACTOR LOADINGS: columns I–IV.

TABLE VIII.—*Intercorrelations table, residual matrices and factor loadings of 14 visual acuity tests given twice to 261 EM at Fort Dix, N. J.—Retest.*

INTERCORRELATIONS (below diagonal); RESIDUALS (above diagonal); FACTOR-LOADINGS

Tests	7	13	5	8	2	11	3	9	4	10	14	6	12	1	Resolution I	Brightness II	Letter Form III	IV
7 Bausch and Lomb Checkerboard		00	03	-02	-01	-02	-01	01	01	-01	02	01	01	01	93	-07	-05	-01
13 Checkerboard Variable Grid	82		00	-01	00	-01	00	00	00	-01	00	-01	-01	-02	89	02	04	01
5 New London Letter	80	77		-01	-03	-01	-01	-02	-02	00	-03	02	-03	-03	85	05	39	-05
8 AAF Letter	80	80	86		-02	00	-01	00	00	00	00	00	-04	-04	90	-01	28	02
2 Army Snellen	76	77	79	82		02	-03	01	01	-02	01	00	-01	00	85	07	26	11
11 AAF Constant Decrement	77	78	83	85	84		00	01	00	-01	03	-03	-01	02	87	12	25	04
3 Dot Variable Size	80	81	81	83	79	84		00	-02	02	05	01	00	-06	90	22	12	07
9 Line Resolution	77	77	77	80	79	82	85		01	01	02	01	-01	01	85	31	13	06
4 Quadrant Contrast	52	53	50	52	54	56	61	65		00	03	01	-03	-02	58	48	00	02
10 Dot Contrast	57	57	59	59	55	59	64	62	47		-03	00	02	01	64	20	08	-09
14 Vernier	67	66	67	70	70	74	69	75	58	50		00	-02	01	73	27	17	02
6 Triangle Discrimination	60	59	60	62	64	60	68	65	48	42	55		-01	-01	66	16	07	42
12 Square Discrimination	30	31	33	38	40	38	40	39	26	25	32	45		-01	34	17	21	45
1 Practice (Square C)	71	70	73	75	75	79	74	80	58	58	70	60	38		79	29	20	06

TABLE IX.—*Intercorrelations table, residual matrix and factor loadings of 14 visual acuity tests given once to 531 EM at Fort Dix, N. J.*

INTERCORRELATIONS (below diagonal); RESIDUALS (above diagonal); FACTOR LOADINGS

Tests	7	13	5	8	2	11	3	9	4	10	14	6	12	1	Resolution I	Brightness II	Letter Form III	IV
7 Bausch and Lomb Checkerboard		01	00	01	01	00	01	01	-01	00	01	03	-01	01	90	02	00	02
13 Checkerboard Variable Grid	82		-02	02	00	02	00	00	01	01	02	-02	00	-01	90	00	-01	-01
5 New London Letter	80	78		02	-01	01	02	-01	00	02	00	02	-02	-02	89	00	25	-01
8 AAF Letter	82	82	89		-02	00	00	-03	00	-01	01	-01	00	-01	89	11	32	-01
2 Army Snellen	82	81	87	90		01	01	00	00	01	02	00	-02	01	90	09	32	-02
11 AAF Constant Decrement	80	82	87	90	92		01	02	01	01	01	00	-02	00	89	13	29	02
3 Dot Variable Size	82	81	85	85	87	87		02	-02	00	01	02	-02	-01	90	20	11	00
9 Line Resolution	81	79	81	84	87	88	89		-01	00	-01	01	00	00	88	31	16	02
4 Quadrant Contrast	58	59	58	62	62	64	64	68		-01	01	00	-02	-01	65	37	02	01
10 Dot Contrast	54	54	56	57	59	60	62	65	52		00	00	01	-01	59	40	04	-05
14 Vernier	74	74	77	80	84	84	82	84	66	60		00	00	-01	80	32	24	19
6 Triangle Discrimination	65	59	64	63	65	66	68	69	52	46	68		00	-02	68	19	08	32
12 Square Discrimination	49	49	48	50	48	49	48	50	35	32	51	48		01	55	01	04	32
1 Practice (Square C)	75	73	75	80	82	81	79	82	61	56	78	63	51		82	21	18	11

TABLE X.—*Factor loadings from analysis of first test, retest, and test only matrices*

TESTS	FACTOR LOADINGS FIRST TEST: Resolution I	Brightness II	Letter III	Form IV	FACTOR LOADINGS RETEST: Resolution I	Brightness II	Letter III	Form IV	FACTOR LOADINGS TEST ONLY: Resolution I	Brightness II	Letter III	Form IV
7 Bausch and Lomb Checkerboard	90	-01	-03	-01	93	-07	-05	-01	90	02	00	02
13 Checkerboard Variable Grid	90	01	-03	-09	89	02	04	01	90	00	-01	-01
5 New London Letter	84	11	21	08	85	05	39	-05	89	00	25	-01
8 AAF Letter	89	09	29	01	90	-01	28	02	89	11	32	-01
2 Army Snellen	85	28	24	01	85	07	26	11	90	09	32	-02
11 AAF Constant Decrement	87	15	24	-01	87	12	25	04	89	13	29	02
3 Dot Variable Size	87	31	08	02	90	22	12	07	90	20	11	00
9 Line Resolution	82	44	-01	-05	85	31	13	06	88	31	16	02
4 Quadrant Contrast	55	55	-06	-08	58	48	00	02	65	37	02	01
10 Dot Contrast	54	41	06	18	64	20	17	02	59	40	04	-05
14 Vernier	72	29	06	04	73	27	17	02	80	32	24	19
6 Triangle Discrimination	63	28	09	29	66	16	07	42	68	19	08	32
12 Square Discrimination	40	19	08	46	34	17	21	45	55	01	04	32
1 Practice (Square C)	80	21	18	-03	79	29	20	06	82	21	18	11

ings, to be sure, but relatively smaller amounts of score variance are accounted for by retinal resolution. Such tests are contrast and form discrimination tests (4, 10, 6, and 12). To use these tests to measure factor I is to invite troubles. A second point is the advisability to pick easily constructed tests to measure this ability. Why slave over letter tests which offer such problems in equating letter difficulty, when the checkerboards so patently avoid the problems? Something could be said too for tests utilizing simple lines or dots. But, as we shall see, in their present form these tests seem to measure other factors as well as retinal resolution.

(2) *Consistency of loadings over the three samples.* The first factor, retinal resolution, has quite consistent loadings for all the tests as we move from sample to sample. Loadings on this factor tend to be slightly higher for the retest and test-only samples than for the test sample. The larger sample in the case of test-only (531 cases) and greater intra-individual stability in the retest group may account for the generally higher intercorrelations obtained in these analyses.

b. FACTOR II—BRIGHTNESS DISCRIMINATION. The highest loadings on this factor in all analyses were found on the two brightness contrast tests (4 and 10). From 4.0 to 30.3 percent of the variance in test score on these two tests is accounted for by this factor. It is therefore named brightness discrimination.

The surprising fact is the significant or near significant loadings found for other tests on this factor. In three of these tests (3, 9, and 14) the aspect accounts for from 4.0 to 19.4 percent of test score variance. The letter tests apparently do not measure brightness discrimination having loadings ranging from zero to insignificant positive values with one glaring exception. The Army Snellen (2) has a rather significant loading (0.29) in the analysis of first-test data. For some reason 7.8 percent of score variance on this test is accounted for by the brightness factor in this analysis. Apparently brightness cues were present in the test. This is conceivable and may be explained by the amount of previous investigation and revision that went into the Snellen chart which was less than that to which the new London and AAF letter tests were subjected. Other tests for which the factor has barely significant but rather consistent loadings are 6, 12, and 1. In these tests up to about 7.8 percent of the variance is accounted for by brightness discrimination. The loadings for the checkerboard tests (7, 13) cluster around zero.

The explanation for the loadings of this aspect of vision on the line resolution, dot variable size, vernier, and practice tests (9, 3, 14, 1) seems rather simple. In all these as well as the contrast tests correct response is a matter of seeing "place where" a darker or lighter area appears. Thus in the dot variable size test the place where a slightly darker blur appears may be correctly called after all clarity of the dot appearance has vanished. If this mode of responding is accomplished by eye movements, such a response should be enhanced by bringing into play other populations of end organs. This may account for the consistently higher loadings on the line resolution test (9). In this case the line (always of constant length) may be considered a series of dots which when moving about as a retinal image would by change cover greater populations of end organs than in the case of a single dot. Similar explanations may be offered for loadings on the vernier and practice tests (14 and 1) and in the more visible ranges of the discrimination tests (6, 12). In all these cases correct responses do not always depend on clearly seeing the object as presented but may be made on the basis of where the items appear darker or lighter.

This is in direct contrast to what is required in both letter and checkerboard tests. Once the examinee's eye shifts or the objects become blurred for any reason whatsoever the test is over unless by some chance brightness cues are present in certain unstandardized items.

(1) *Test purity.* This might better have been labeled test impurity, for no tests were found which measured the brightness discrimination factor alone. In fact, for all of the instruments on which this factor has significant loadings, the amount of variance in score accounted for is always the same or less than that accounted for by retinal resolution. If we select the best test of the brightness aspect in the present battery, quadrant contrast (4), retinal resolution accounts for 30.3, 33.6, and 42.3 percent of variance in the test scores of the three samples. Meanwhile, brightness discrimination accounts for only 30.3, 23.0, and 13.7 percent of test score variance in the respective analyses. It is apparent that in their present form the contrast tests measure resolution to a greater extent than they do the brightness discrimination factor. This indicates that such tests require a great deal of spatial localization in the retinal resolution sense. In the case of the quadrant contrast test the response called for was in terms of "which way the arrow is pointing;" while in the dot contrast response was made in terms of where the dot appeared in a diamond surround. This suggests a way of developing a purer brightness discrimination test: simply make the response independent of spatial pattern or as it has been called "edge" discrimination. This was obviously not accomplished in the present tests. "Present" or "absent" in large unstructured areas rather than "present here" or as an "arrow pointing" would be more suitable modes of response. This improvement is common to laboratory practices, and it remains to invoke such a device for clinical measurement of brightness discrimination ability if such is justified as having sufficient validity.

(2) *Consistency of loadings.* The brightness discrimination factor loadings are generally lower for the retest analysis than those for the first-test analysis. The rather significant loading on Army Snellen

(0.28) diminishes to insignificance (0.07) on retest. This indicates that first-test performances included more responses actively associated with brightness discrimination (eye movements or deviation from trying to see the exact nature of the objects) than in the case of retest. For the larger sample of test-only (531 cases) the loadings on brightness discrimination are of about the same order of magnitude as those found for the retest sample. Because of the greater number of cases with higher initial intercorrelations the results of the larger test-only analyses are probably closer to the "true" values than are those of the first-test analysis.

c. Factor III—Form (Letter) Perception. This factor has consistent, rather significant, loadings on all the letter tests (5, 8, 2, 11). The amount of letter test score variance accounted for is not high, ranging from 4.4 to 15.2 percent. The practice test (1) also has consistent loadings in the factor but less significant ones (3.8 to 4.0 of score variance). Since this test employed a modified C as the test object, small loadings are not surprising and lead one to believe that this aspect is a "letter perception" factor. Before jumping on this name, however, let's look at some more evidence. With few exceptions the factor has near zero or insignificant positive loadings on all other tests. Exceptions are loadings of 0.17 and 0.24 on the Vernier Acuity (14) retest and test-only analyses, a loading of 0.21 on Square Discrimination in the retest analysis and other scattered loadings accounting for only about 2.0 percent of score variance in any case.

This aspect is definitely a type of form perception dependent on the specific shape of the objects presented in the tests. In this study it seems to be tied up mostly with letter objects, but it likely will be found in any test in which form naming is involved. Some evidence for this is found in the analyses of data from commercial devices given in the next section of the present report. In these analyses the Telebinocular for depth, or stereopsis, test is found to have loadings on a form (letter) factor. Examination of the slide reveals that responses are made in terms of the identity of whole forms namely, hearts, squares, stars, circles, and crosses. This, as in the case of letters requires a response in terms of differential naming of the objects. All of which reveals that this factor is not confined to letter objects but may include other whole patterns as well.

If such is the case, the influence of past experience with definitive visual objects may play a role. As we shall see in a later section, the letter test scores bear close to zero relationship with AGCT scores. The AGCT measures a large verbal facility (vocabulary) factor, but this is not the same as sheer letter facility which seems closer to the problem in the present analysis. All of which may explain why the zero relationship with AGCT is found in this study. Another cogent explanation is to be found in the low factor loadings found on this factor. Too small an amount of test score variance may have been accounted for by the form (letter) factor to show appreciable correlation between the letter tests and AGCT.

(1) *Test purity.* It is doubtful that this perceptual factor will ever be measured in a pure form. Retinal resolution will always be required to a large extent in all such tests. In the present battery the form (letter) factor accounts for only 4.4 to 15.2 percent of score variance on tests 5, 8, 2, and 11 whereas retinal resolution accounts for 70.6 to 81 percent of the variance on the same tests. However, when the eyes can be corrected and resolution more or less mechanically partialed out, certain tests should do a better job of measuring this factor. They should not depend on diminishing size of the objects but rather introduce levels of difficulty in terms of distortions or omissions of parts of the objects. Of course, such tests might be subject to training which should be considered in any evaluation of them.

(2) *Consistency of loadings.* Loading of the form (letter) perception factor on the tests are rather consistent across the three analyses. There is a slight tendency for higher loadings to appear on retests than on first tests a fact which may or may not be confirmed by analyses of results of several days of repeat testing. Such a study has been made with tests being given to the same group on 5 consecutive days. The data has not been analyzed. It will be interesting to note the effects of such repetition on this letter factor as well as the other three.

d. Factor IV—Simple Form Perception. This factor has consistent significant loadings on the Triangle and Square Discrimination tests (6, 12) and near zero or insignificant loadings on all other tests. The loadings are not high accounting for from 8.4 to 21.2 percent of test score variance. The existence of this factor as distinct from the form (letter) perception factor is interesting since initial assumptions lumped the two together.

Factor IV, having significant loadings on the two tests (6 and 12) in which curved sides are discriminated among rectilinear ones, is named Simple Form Perception. It differs from factor III in the sense that the form (letter) perception aspect is associated with tests of identification of whole objects whereas the present factor has highest loadings on tests in which a simple visual judgment (curved sides) is made about a constant object (triangle or square). In the case of the Square Discrimination test (12) both modes of response are possible, at least in the early stages of the test. An examinee may tell which object has "curved sides" (factor IV), or he may report which is a circle rather than a square (factor III). With practice such seems to be the case. In the retest analysis test 12 measures factor III to some extent (4.4 percent of variance) as well as factor IV (20.6 percent of variance).

Factor IV accounts for from 8.4 to 21.3 percent of test-score variance

on tests 6 and 12. On other tests the loadings are insignificant or so inconsistent as to defy explanation.

(1) *Test purity.* As in the case of the letter perception factor, this factor is impurely measured by the tests employed. This is unavoidable since the tests are so obviously dependent on factor I, retinal resolution. Then too, some infusion of factor II, brightness, is noted. Apparently, there is some possibility of response on the basis of dark or bright areas of the triangles and squares. Finally, as we have mentioned, there is a possibility of impurity arising when the task becomes one of discriminating whole objects (factor III).

(2) *Consistency of loadings.* As indicated above, the loadings for all tests except 6, 12, and 14 are consistently low over all three samples. The loading for test 14 (Vernier) is low in the "First Test" and "Retest" samples, but unaccountably rises in the "Test-only" sample. However, this loading is still low. Tests 6 and 12 have significant and fairly consistent loadings over all three samples. It is therefore in terms of the loadings on these two tests, as well as the consistently low loadings for all the others, that this factor is defined.

e. Summary. Four factors are independently measured by the 14 tests employed in this study. They are:

I. Retinal Resolution.
II. Brightness Discrimination.
III. Form (Letter) Perception.
IV. Simple Form Perception.

Factor I, Retinal Resolution, accounts for most of the score variance on all the present tests. This factor is measured quite purely by tests 7 and 13. These two checkerboard tests are the best instruments for measuring far (20 feet) retinal resolution. Letter, dot, and line tests (5, 8, 2, 11, 3, and 9) measure this factor to about the same percent of score variance, but measure other factors as well, namely, brightness discrimination and form (letter) perception. Turning to the other factors, we find they are measured to a smaller extent and always with impurity since the tests always measure factor I as well. Brightness discrimination is measured best by the two brightness tests (4 and 10) but on other tests, such as line (9) and vernier (14), the factor has significant loadings. This suggests that this factor is measured in all tests in which judgments of "place where" darker or lighter areas appear lead to correct responses.

Form (Letter) Perception, Factor III we find associated with all tests in which whole objects are named. In this sense, the factor is misnamed, letters being but one type of whole object which may be considered as measuring this aspect. In the present battery the factor has highest loading on the letter tests.

Factor IV, Simple Form Perception, is an aspect of visual ability

701639—48——5

dealing with more basic line form discriminations within the whole object but independent of the name of said objects.

Factors I and II seem to be definitely associated with retinal and lensatic phenomena. In the case of retinal resolution a definite spatial pattern seems to be involved, whereas brightness discrimination seems less structured. Factors III and IV are perceptual involving cortical activities for such responses as whole form naming and discrimination of linear qualities. These perceptual aspects are more closely related to organization of visual cues and have barely scratched the surface of this important ability area.

f. WARNING. If tests such as the ones used in the present study are to be utilized for screening and classification purposes, one must be careful to consider what aspects are measured and with what degree of purity. For determining retinal resolution as an indicator of the adequacy of lensatic and retinal mechanisms of the eye, two tests of the present group will serve, namely the checkerboards. The other tests measure one or more of the other three factors as well as retinal resolution. It is interesting to examine certain representative tests with respect to the way in which variance in test score is accounted for by the information provided by the present study. The test-only analysis is used. In each example the percent of variance accounted for by the aspect under consideration is shown in italic.

Retinal Resolution Tests

1. Bausch and Lomb Checkerboard (7):

	Percent
Retinal Resolution	*81.0*
Brightness Discrimination	0.0
Letter Perception	0.0
Line Form Perception	0.0
Specific and Error	19.0
Total	100.0

The specific and error variance is attributable, respectively, to the portion of intercorrelation remaining in the residual table after the present analysis and lack of reliability in the present measure. (NOTE.—The Checkerboard Variable Grid test performs in an identical manner in this analysis.)

2. AAF Letter Test (8):

	Percent
Retinal Resolution	*79.2*
Brightness Discrimination	01.2
Letter Perception	10.2
Line Form Perception	0.0
Specific and Error	9.4
Total	100.0

In addition to resolution this test measures some ability to deal with letters which may or may not prove valid and must be remembered when using the tests for any given purpose.

3. Line Resolution Test (9):

	Percent
Retinal Resolution	*77.4*
Brightness Discrimination	9.6
Letter Perception	2.6
Simple Form Perception	0.0
Specific and Error	10.4
Total	100.0

In addition to resolution, this test measures some ability to utilize brightness cues which destroys its purity in measuring the resolution aspect. Form (letter) perception is also apparent to a slight degree.

Brightness Discrimination Tests

1. Quadrant Variable Contrast (4):

	Percent
Retinal Resolution	42.3
Brightness Discrimination	*13.7*
Letter Perception	0.0
Line Form Perception	0.0
Specific and Error	44.0
Total	100.0

In addition to measuring more resolution than brightness this test is quite unreliable in its present form.

2. Dot Variable Contrast (10):

	Percent
Retinal Resolution	34.8
Brightness Discrimination	*16.0*
Letter Perception	0.2
Line Form Perception	0.3
Specific and Error	48.7
Total	100.0

This test, showing the same pattern, is even less reliable than the preceding one.

3. Vernier Acuity (14):

	Percent
Retinal Resolution	64.0
Brightness Discrimination	*10.2*
Letter Perception	5.8
Line Form Perception	3.6
Specific and Error	16.4
Total	100.0

This test, while never considered as predominantly measuring brightness discrimination ability, seemingly measures this aspect to a greater degree than previously expected. Among the tests it seems the one most prone to measure, to a greater or lesser extent, all the aspects discovered. Thus, it affords an examinee opportunity to attempt the task in many alternative ways. There is virtually no specific left in this test after the present analyses making it very improbable that inclusion of more vernier tests in a battery will produce an additional specific vernier factor.

Form (Letter) Perception Tests

New London Letter (5):	Percent
Retinal Resolution	79.2
Brightness Discrimination	0.0
Letter Perception	*6.3*
Line Form Perception	0.0
Specific and Error	17.0
Total	100.0

The form (letter) perception aspect, as can be seen from this summary, is measured to a slight but significant extent. This is true of all such tests in the present battery.

Line Form Perception

Triangle Discrimination Test (6):	Percent
Retinal Resolution	46.2
Brightness Discrimination	3.6
Letter Perception	0.1
Line Form Perception	*10.2*
Specific and Error	39.9
Total	100.0

As a test of line form perception this test as well as the Square Discrimination test (12) is far from pure (resolution 46.2 percent of variance) and is rather unreliable in its present structure.

g. Future Areas of Factorial Search. The present battery of 14 tests leaves certain aspects of visual ability virtually untouched and others are only measured to a limited extent. The predominant role of the retinal resolution factor (I) is undoubtedly responsible for the latter. The checkerboard tests have licked the problem of measuring retinal resolution (I), and if such is desired, tests of brightness discrimination (II) can be developed along lines eliminating resolution in the sense of spatial pattern of the test objects. Beyond these factors, which are correlated with physical and physiological properties of the eye, lies an area concerning the utilization of the patterns made available by the eye. In terms of variance in responses to everyday visual situations, analysis and test development in this perceptual area may develop important sources of individual differences once the eye

has been corrected. It remains for validity studies to point out the way for future efforts in both areas.

3. Factor Analysis of Visual Tests by Three Commercial Devices

a. Data. Data in this paragraph of the report were collected by United States Navy personnel at the Medical Research Department, United States Submarine Base, New London, Conn. A report "Visual Acuity Measurements with Three Commercial Screening Devices," Bureau of Medicine and Surgery, Research Project No. X–493 (Av–263–p), describes the instruments, tests, and the three screening devices. The tests were administered to 128 persons, mostly naval personnel and mostly male. Distributions and reliabilities are presented in the report mentioned above as are controls for position of test in the battery.

The primary purpose of the study was to accomplish a comparison of wall charts with the screening devices as measures of acuity and of Maddox Rods with the three screening devices as measures of phoria. Depth tests were added to the binocular acuity study, and interpupillary distance (IPD) was related to phoria measurement as a minor problem.

The three screening devices were the Bausch and Lomb *Orthorater* (two machines), American Optical Co. *Sightscreener*, and Keystone *Telebinocular*.

The present report extends the analysis of these data by means of factorial analysis. For purposes of convenience the following subgroups of tests were treated independently:

(1) *Table XI, acuity tests, left eye.* This study includes both test and retest correlations for all wall and machine charts and the average for one near and one far wall chart. The tests included are:

- 2 wall charts, far.
- 1 wall chart, near.
- 4 machine charts, far.
- 4 machine charts, near.

(2) *Table XII, acuity tests, right eye.* This study includes the same data as for left eye.

(3) *Table XIII, acuity and depth tests binocular.* In this study, one far acuity machine test is dropped and one near acuity machine test is added to those mentioned above while several depth variables are also added. The tests are:

- 2 wall charts, far.
- 1 wall chart, near.
- 3 machine charts, far.
- 5 machine charts, near.
- 4 machine charts, far depth.
- 1 machine chart, near depth.

The above tables give the original intercorrelations and the residual correlations based upon factor loading reported in table XIV for all three studies to facilitate comparison. The variables, which are listed only by number in table XI, XII, and XIV, are named in table XIV.

(4) *Table XV, phoria tests, orthorater versus maddox rod tests, with age as a control variable.* Tests and the control variable included in this study are:

- 4 far vertical phoria tests; 2 from Orthorater (plus one retest) and Maddox Rod.
- 4 near vertical tests, same.
- 6 far lateral phoria tests, 1st Orthorater (with special far excursion measurements, test and retest), 2d Orthorater (plus retest), and Maddox Rod.
- 6 near lateral phoria tests, same as far lateral and, age in years.

(5) *Table XVI, phoria tests, three devices, with interpupillary distance as a control variable.* The tests and control are:

- 6 far vertical phoria tests, 3 machines (plus retests).
- 4 near vertical phoria tests, 2 machines (plus retests).
- 6 far lateral phoria tests, same as far vertical.
- 4 near lateral phoria tests, same as near vertical and, interpupillary distance.

Tables XV and XVI again present only intercorrelations and residuals based upon factor loadings. These are brought together in table XVII for comparison and supplementation. The variables are named in table XVII.

And finally, to permit the establishment of the degree of overlaps of factors on certain selected tests, we have—

(6) *Table XVIII, factor integration matrix.* This study includes—

- 3 far acuity tests, 1 wall and 2 machine charts.
- 3 near acuity tests, same.
- 2 far depth tests, machine.
- 1 near depth test, machine.
- 4 near lateral phoria tests, 3 machine plus Maddox Rod.
- 3 near vertical phoria tests, 2 machine plus Maddox Rod.
- 2 far lateral phoria tests, 1 machine plus Maddox Rod.
- 2 far vertical phoria tests, 1 machine plus Maddox Rod.

Again, table XVIII gives only intercorrelations and residuals while factor loadings are reported in Table XIX.

All factors are orthogonal and have been rotated to as near simple structure as the orthogonal framework permits.

b. Factors in Acuity Tests. Table XIV shows that the correlations among the acuity tests (1–26) can be explained by one general and three common factors.

(1) *Retinal resolution.* The general factor, A, has significant loadings on all acuity tests (1–26) and on all depth tests (27–36) as well. The loadings are highest for the far tests (1–13) and slightly lower for the depth tests (27–36) than for the near tests (14–26). This

Table XI.—*Acuity tests, left eye, intercorrelations*

[Circled values are remainders of reliability or part-whole correlations]

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		83	83	79	82	68	69	71	70	76	78	60	59	53	61	60	63	62	58	59	61	61	48	53
2	(03)		80	80	82	66	70	69	69	79	82	53	49	56	62	63	61	62	59	60	60	65	48	51
3	03	02		89	96	71	69	69	69	74	77	57	55	56	65	65	63	61	60	61	62	63	48	53
4	00	03	(09)		96	68	69	68	67	76	75	58	56	54	64	61	62	63	61	63	62	63	46	53
5	-04	-02	(11)	(12)		70	69	69	68	77	78	58	56	56	65	64	63	63	62	62	63	64	48	55
6	-01	-01	02	01	-05		86	82	82	66	69	60	57	54	53	58	74	68	64	67	64	52	43	56
7	01	03	01	04	-05	(03)		87	88	67	72	57	56	51	51	55	69	71	65	70	63	55	42	52
8	03	01	01	02	-07	00	00		87	64	71	61	60	49	50	53	68	64	60	67	64	56	44	53
9	02	01	00	01	-07	00	02	(03)		71	72	58	57	54	55	57	68	68	65	70	61	59	41	50
10	00	04	-01	02	-05	-05	-05	-08	00		88	58	53	51	60	57	66	68	64	71	62	73	48	49
11	-02	03	03	-03	-11	-07	-05	-06	-05	(08)		62	57	54	58	59	69	72	71	72	62	70	50	57
12	05	00	01	07	-02	03	-03	01	00	03	03		86	21	26	25	47	40	36	41	37	41	30	37
13	06	-02	-02	07	-02	01	-04	00	-01	01	00	(38)		16	24	20	38	36	33	36	32	33	22	28
14	00	02	00	00	-01	01	00	00	02	-04	-02	-08	-10		73	92	62	65	65	63	69	58	51	60
15	-01	00	-03	01	00	-02	00	00	02	-01	-04	-05	-03	(09)		90	67	72	65	66	67	65	52	60
16	-01	01	-02	-03	-01	00	00	00	00	-05	-04	-05	-04	(26)	(17)		68	73	67	69	72	63	55	63
17	03	-01	-01	01	-03	04	-02	-01	-02	-02	-01	05	00	-03	00	-02		85	80	84	72	66	57	65
18	02	00	01	02	-03	00	02	-03	00	-01	02	00	00	-01	03	01	(02)		82	86	76	72	60	67
19	00	00	00	02	-02	00	00	-03	00	-01	05	-04	-09	03	00	00	04	03		82	68	67	55	59
20	00	00	00	04	-03	-01	00	-01	01	04	03	-02	-02	00	01	-04	04	03	(06)		75	70	54	62
21	05	03	03	05	-02	03	-02	05	00	01	00	-02	01	08	02	05	-01	01	-01	03		73	57	61
22	-02	02	-02	-01	-04	-05	00	02	04	08	04	03	00	-01	-01	-03	-02	01	02	03	(10)		51	59
23	02	02	00	01	-02	-01	-01	-03	-02	00	01	04	00	02	-02	00	01	02	01	-01	04	-01		71
24	01	-01	00	00	-01	02	-02	01	-04	-06	00	05	00	04	00	01	00	-01	-03	-02	00	01	(23)	

Residuals

TABLE XII.—*Acuity tests, right eye, intercorrelations*

[Circled values are remainders of reliability or part-whole correlations]

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		77	80	81	83	65	70	66	60	64	69	61	54	56	44	53	51	46	51	49	50	50	39	20
2	-03		80	84	85	69	69	66	64	65	72	56	47	53	53	58	51	50	53	52	51	58	41	29
3	00	01		87	96	73	72	72	67	66	75	59	51	56	59	60	55	53	57	56	54	54	41	29
4	-02	03	05		96	73	75	74	65	73	78	55	48	64	60	62	59	56	63	59	58	65	42	27
5	-06	-02	08	06		75	75	76	67	72	78	59	52	62	62	65	58	56	63	59	58	60	42	30
6	03	01	02	02	00		83	83	77	68	69	64	61	48	39	46	60	60	55	58	53	56	48	37
7	00	-04	-02	01	01	00		77	84	70	71	61	55	47	40	45	59	55	54	57	51	55	40	35
8	02	01	04	06	04	10	00		76	74	68	59	51	52	50	53	57	58	58	60	52	57	45	38
9	-02	00	00	-02	-01	01	00	05		68	66	57	45	47	46	49	62	63	54	60	56	60	44	39
10	01	01	-01	05	-02	-03	00	08	02		81	59	46	48	52	54	56	52	53	56	58	64	42	36
11	05	07	07	08	03	-03	-01	00	-02	13		58	48	58	57	60	60	57	60	61	59	65	52	38
12	04	-01	-01	-04	-07	02	00	02	01	02	01		78	32	25	33	39	33	41	40	34	36	26	25
13	05	-03	-01	-03	-04	06	00	00	-06	-04	-02	29		23	14	19	33	24	41	31	27	24	21	15
14	01	-01	01	04	-01	00	00	06	01	02	05	-01	01		78	92	66	71	69	69	64	70	56	45
15	-07	03	08	03	02	-05	-01	07	05	05	05	-05	-04	14		92	65	68	64	69	70	73	56	50
16	-03	02	02	-02	-02	-05	-03	03	00	00	00	00	-01	24	23		67	71	67	73	70	75	59	48
17	01	-01	01	00	00	00	-04	00	01	-01	-03	03	07	-01	-02	-05		85	81	81	75	76	60	48
18	-03	-01	00	-02	01	02	-07	03	03	-02	-04	00	01	04	00	-01	01		81	82	72	77	59	53
19	-01	00	02	03	02	00	-01	05	-01	-02	00	05	16	03	-03	-04	04	04		78	70	72	55	48
20	-04	-02	00	-02	-02	00	-03	05	02	00	-01	04	05	02	02	02	01	01	03		71	72	56	53
21	00	00	01	00	00	-01	-05	00	02	05	01	-01	02	01	06	02	00	-03	00	-02		81	58	48
22	-01	06	-01	05	00	-01	-04	02	02	08	04	-01	-03	06	08	06	-01	00	00	-03	10		60	48
23	-01	00	-01	-04	-04	04	-06	03	00	00	05	-01	01	05	05	04	-01	-03	-02	-03	03	03		72
24	-07	00	-02	-07	-02	00	-04	03	01	01	-01	04	00	04	08	04	-01	-01	-01	02	01	-01	33	

Residuals

TABLE XIII.—*Acuity and depth tests, binocular, intercorrelations*

[Circled values are remainders of reliability or part-whole correlations]

Test	1	2	3	4	5	6	7	8	9	10	11	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1		81	75	77	77	71	68	68	66	59	70	50	60	58	52	52	52	50	52	57	48	47	46	46	42	38	30	32	42	45	33	40	38	36
2	13		79	78	81	72	69	69	67	66	72	48	62	53	58	58	57	55	57	57	49	46	46	52	51	43	31	34	47	52	45	53	40	37
3	01	03		83	94	76	70	72	69	64	76	56	65	65	57	53	61	56	58	61	56	49	44	54	42	36	38	41	45	50	48	47	35	34
4	02	02	02		95	76	74	72	72	69	79	50	62	59	56	58	59	55	53	57	50	48	42	47	46	40	37	40	44	52	41	44	38	40
5	00	01	07	09		79	74	75	73	68	79	56	66	65	59	59	62	58	59	61	52	48	44	53	46	41	38	41	47	53	45	46	38	39
6	03	-02	04	-01	-01		88	88	86	71	76	50	55	56	65	65	57	62	59	57	52	54	41	46	48	42	50	53	45	51	46	46	42	42
7	03	-02	01	00	-03	-05		86	86	73	76	49	54	55	61	66	57	61	53	58	51	58	43	48	41	44	53	54	42	53	45	48	40	45
8	05	-01	05	00	-01	-04	-05		87	71	74	48	51	52	58	62	54	59	55	56	48	55	42	46	40	44	47	49	47	55	46	46	44	45
9	03	-02	01	00	-02	00	00	03		68	76	49	55	54	62	62	55	61	56	54	52	58	41	44	39	41	48	50	40	49	42	44	38	37
10	01	02	00	04	-01	-02	01	00	-01		73	50	46	49	54	56	52	54	61	64	51	48	35	35	40	38	32	37	36	45	35	40	30	31
11	00	-02	-01	01	-02	-01	01	01	03	05		59	65	63	61	62	63	65	62	60	57	54	49	46	37	38	34	43	39	51	41	43	36	33
14	-02	-05	00	-03	-01	-04	-03	-03	-02	03	01		67	87	61	66	69	61	59	65	47	46	48	41	29	28	25	32	25	26	19	18	38	37
15	00	02	00	00	01	-05	-02	-03	-01	-05	00	09		92	66	68	71	68	60	64	57	57	51	59	31	30	21	24	29	34	28	29	37	33
16	01	-06	01	00	00	-01	00	00	-01	-02	-01	25	23		70	72	76	69	64	69	55	56	55	57	33	31	27	30	28	33	24	26	40	38
17	01	01	02	03	05	-02	-06	-06	-02	-03	-04	-04	-05	00		84	85	84	71	71	67	64	56	56	34	33	31	38	29	40	29	31	43	39
18	-02	-01	-06	01	01	-02	00	-01	-02	-01	-04	02	-02	02	-01		82	82	68	72	56	64	52	52	31	30	29	35	30	40	25	28	40	42
19	-02	00	02	04	03	-03	-02	-02	-03	00	-01	01	-03	01	00	01		79	72	72	64	62	54	52	32	29	25	34	25	33	28	23	43	33
20	00	01	02	03	05	-03	-04	-03	-02	00	03	00	02	02	00	02	-01		63	68	57	60	52	55	36	33	28	37	29	38	23	31	42	41
21	02	01	01	00	-01	-01	-06	-03	-02	-02	00	01	-01	00	00	00	01	-03		70	58	58	47	49	28	32	19	31	27	37	29	25	40	29
22	04	-02	-01	01	-01	00	01	01	-04	09	-04	06	01	03	01	04	00	02	05		55	53	52	48	35	32	24	29	33	43	32	31	39	35
23	03	-02	03	03	01	04	04	02	03	03	02	-04	02	-04	02	-07	-01	-04	00	-05		72	55	48	26	26	22	32	36	35	23	31	25	21
24	02	-03	00	00	-01	-03	02	01	03	-01	-02	-04	02	00	04	-01	-02	-04	02	-03	21		49	45	23	29	18	23	26	32	21	22	28	30
25	04	01	-04	-01	-02	00	04	04	02	-03	01	00	-02	01	-01	-03	-05	00	-02	-01	07	06		65	25	33	27	30	32	40	31	29	28	36
26	00	05	03	00	04	02	06	06	02	-04	-04	-06	07	06	03	00	-03	06	03	-01	04	04	23		23	31	27	26	34	35	27	28	30	32
27	05	10	00	07	02	01	-03	-05	-01	07	-01	-01	-02	-02	03	00	00	07	-01	05	01	00	-05	-05		57	34	38	54	55	45	55	62	51
28	05	04	-03	06	00	-03	02	-01	03	06	02	-02	-03	01	-01	-02	-05	02	02	00	-03	04	-01	02	03		38	31	57	63	37	37	70	78
29	02	-05	07	05	05	01	01	-06	02	-04	-01	01	-04	04	-07	-06	-04	-06	-09	-02	-02	-10	04	06	-02	00		79	41	43	37	50	46	50
30	02	-06	-05	04	02	-01	00	-06	00	-04	03	08	-01	05	00	-01	05	02	-01	-02	03	16	07	05	02	-06	18		43	45	39	54	52	46
31	02	-03	-05	00	-01	-05	-05	-02	-04	-04	-06	-03	-03	-03	-05	-04	-06	-01	-03	05	00	00	-03	02	04	09	-01	01		75	71	75	54	61
32	-02	00	-01	07	05	-02	02	02	-04	00	03	-05	-02	-03	-04	-02	-04	-01	00	03	-02	-01	-01	00	01	04	-01	-05	06		64	76	57	62
33	00	-01	04	05	05	02	03	02	-02	-02	01	-02	03	-01	-03	-06	-03	-04	02	-01	-03	-02	-02	-02	02	-04	01	-04	04	04		72	48	43
34	02	04	-02	01	00	03	01	-03	-05	01	00	-05	01	00	00	-03	-03	05	-01	00	02	-02	-03	-01	03	06	04	06	-02	07	01		52	61
35	-03	02	-04	05	-02	04	-03	-02	-01	01	02	03	-02	05	03	03	01	06	07	03	-03	03	00	-01	02	01	-04	05	-11	-01	-02	-10		76
36	04	00	-01	09	02	07	-01	-03	-03	00	01	01	-05	03	-06	01	-12	00	-05	00	-08	02	-02	01	-07	11	-01	-01	00	09	-04	03	-07	

Residuals

Table XIV.—*Summary of acuity and depth factor loadings*

Test	Name	A Retinal Resolution			B Accommodation			C Form (letter)			D Resistance to Interference			E Depth	F Sight Screener (Depth)
		B*	L	R	B	L	R	B	L	R	B	L	R		
1	Far, Snellen	70	86	76	11	-03	00	41	26	30	10	01	06	02	04
2	Far, Snellen, Retest	73	84	77	12	02	00	28	24	42	08	03	11	13	-04
3	Far, New London	79	83	81	14	01	01	41	34	37	00	05	08	10	-02
4	Far, New London, Retest	80	83	81	03	03	09	42	31	41	13	00	09	01	-01
5	Far, New London, Average	90	94	90	06	-01	05	33	21	43	-01	01	-07	90	09
6	Far, Orthorater, 1	86	77	86	04	04	-01	06	11	-01	42	44	23	10	10
7	Far, Orthorater, 1, Retest	85	78	83	03	03	-07	01	01	08	44	52	50	08	07
8	Far, Orthorater, 2	85	79	80	01	-01	01	-03	00	03	40	48	20	11	09
9	Far, Orthorater, 2, Retest	82	77	79	07	03	01	03	07	-01	35	49	37	06	01
10	Far, Sight Screener	75	86	81	11	11	07	05	08	02	18	05	06	04	-07
11	Far, Sight Screener, Retest	80	90	82	20	06	15	26	10	03	15	13	09	04	-06
12	Far, Telebinocular	--	65	72	--	-15	-15	--	-07	05	--	18	-01	--	--
13	Far, Telebinocular, Retest	--	62	63	--	-24	-25	--	-04	02	--	23	04	--	--
14	Near, New London	53	54	54	48	42	51	19	31	28	11	15	08	-01	11
15	Near, New London, Retest	57	61	52	50	43	58	31	40	23	03	04	00	01	09
16	Near, New London, Average	59	61	60	54	55	66	23	39	23	05	11	02	-10	04
17	Near, Orthorater, 1	56	69	63	66	48	59	02	08	01	38	31	29	10	-05
18	Near, Orthorater, 1, Retest	58	70	59	59	56	62	10	04	04	33	25	35	07	-06
19	Near, Orthorater, 2	55	66	61	70	47	57	14	09	10	20	24	16	03	05
20	Near, Orthorater, 2, Retest	54	69	62	60	49	57	04	02	09	37	29	24	05	-03
21	Near, Sight Screener	64	61	59	49	47	53	-01	17	09	07	23	20	-01	00
22	Near, Sight Screener, Retest	63	69	62	50	39	53	10	17	06	02	00	22	07	-02
23	Near, Telebinocular (Circle)	53	49	47	50	39	44	05	18	06	02	06	19	07	-10
24	Near, Telebinocular (Circle), Retest	52	56	38	43	42	40	03	17	-06	26	17	21	01	-10
25	Near, Telebinocular (Letter)	41	--	--	47	--	--	19	--	--	02	--	--	18	03
26	Near, Telebinocular (Letter), Retest	43	--	--	37	--	--	26	--	--	07	--	--	11	00
27	Far Depth, Sight Screener	42	--	--	04	--	--	11	--	--	06	--	--	43	31
28	Far Depth, Sight Screener, Retest	40	--	--	10	--	--	03	--	--	03	--	--	44	33
29	Far Depth, Telebinocular	41	--	--	00	--	--	-13	--	--	30	--	--	32	13
30	Far Depth, Telebinocular, Retest	49	--	--	-01	--	--	-17	--	--	21	--	--	35	10
31	Far Depth, Orthorater, 1	49	--	--	00	--	--	11	--	--	01	--	--	61	04
32	Far Depth, Orthorater, 1, Retest	52	--	--	09	--	--	04	--	--	03	--	--	69	-04
33	Far Depth, Orthorater, 2	44	--	--	-01	--	--	04	--	--	-02	--	--	66	-13
34	Far Depth, Orthorater, 2, Retest	48	--	--	-06	--	--	10	--	--	01	--	--	64	-04
35	Near Depth, Sight Screener	36	--	--	20	--	--	04	--	--	09	--	--	63	52
36	Near Depth, Sight Screener, Retest	34	--	--	23	--	--	01	--	--	19	--	--	60	50

*B - binocular
L - left
R - right

factor is undoubtedly the basic measure of visual acuity, and is accordingly named *Retinal Resolution*. The high loadings of this factor for near and depth tests indicate that these tests are greatly influenced by the resolving power of the retina. Near and depth test scores should either be corrected on the basis of far test scores or should be determined only after proper lenses to bring far vision to normal have been supplied to all subjects.

(2) *Accommodation.* Factor B has significant loadings on all near tests (14–26) and on the near depth tests (35–36) as well. Aside from a few almost significant loadings on Sightscreener (10–11) and Telebinocular (12–13) far tests, the loadings for far tests, and the far depth tests are negligible or zero. Since the factor is highest or present on only near tests and since it is present for monocular (L and R) as well as for binocular (B) tests, the factor is easily identified as a *lens accommodation* factor.

In 23 of 28 opportunities, the accommodation loadings are higher for the right eye or binocular vision than for the left eye. This finding

indicates that the dominant eye (usually right eye) has the leading role in this function.

The few just nonsignificant loadings (0.15 and 0.20) for right eye and binocular sightscreener far tests indicate a possible technical defect in the instrument. It would appear that the lenses are set under 20 feet equivalent, thus requiring some extra accommodation by the normal subjects. It must of course be kept in mind that the error was found for only this one machine and may not apply to other instruments put out by the American Optical Co. In any case the fault determined 4 percent or less of the variance of the scores whereas retinal resolution accounted for from 56 to 81 percent of the test score variance.

The nearly significant negative loadings (−0.15, −0.15, −0.24, −0.25) on the Telebinocular for acuity tests (12 and 13) will probably be repeated, however, whenever this particular plate is used. The error here is also slightly more serious, accounting for from 2 to 6 percent of test score variance in contrast to the 38 to 52 percent accounted for by retinal resolution. The difficulty arises from the solid black field presented for the nontested eye. The test card presents no cues for distance fixation. Examinees good at accommodation are apt to try to see better by "looking closer" in case of difficulty, which only serves to increase the difficulty of the task and hence leads to negative loadings on the accommodation factor. This defect could be overcome by providing a fine boundary or other cues for bringing the views of the two eyes into focus.

(3) *Form (letter) perception.* Factor C has significant loadings on all wall charts which are also all letter charts (1–5 and 14–16). Significant or near significant loadings appear occasionally on machine letter charts (10–13, 21–22, and 25–26). Similarly, near significantly positive loadings (0.17 and 0.18) appear on the Telebinocular near, circles, test which requires form discrimination as between fields interrupted by lines or by dots; and nearly significant negative loadings (−0.13 and −0.17) are found for the Telebinocular far depth tests (29–30) which require the nearest of several objects to be given its form name of circle, square, heart, etc. Loadings are zero for all Orthorater tests and for the remaining depth tests. This factor is consequently called *Form Perception*, with difficulty centered at the letter identification level (probably relatively easy).

Loadings for Form (Letter) Perception are higher for the right eye or both eyes for far tests (1–5) where retinal resolution is the only other factor measure. This would indicate the leading role of the dominant eye in this function as was the case for accommodation. However, for the near tests (14–16), where accommodation as well as resolution is also required, the loadings are heavier for left than for right or both eyes. This indicates that, while the dominant eye plays a leading role in either control of accommodation or of form perception, there is a hierarchy within the two habit systems with the accommodation activity somewhat suppressing the form perception. This explains the fall in percent of test score variance accounted for by this factor (from about 16 to about 6) for right and both eyes, whereas for the left eye the percentage actually rises slightly.

A possible explanation of the negative loadings for the Telebinocular depth test is that subjects who were good at form perception and concentrated attention on identification of the varied forms were somehow diverted from the main problem of judging depth. At any rate this depth test has introduced an unnecessary complication.

(4) *Resistance to interference.* Factor D has its highest loadings on the Orthorater acuity tests (6–9 and 17–20), but occasionally has nearly significant loadings on Sightscreener and Telebinocular tests as well. (NOTE.—Presence of Form (Letter) Perception and Accommodation factors on these other tests makes this added Resistance factor of less *relative* importance for these other tests.) We might, therefore, call the factor simply a Machine Factor. But to so name it would not be very helpful. One or more elements known to interfere with visual perception can be found in connection with all test charts used in these machines. Some of these elements are:

(*a*) Inhibitory effect of large black areas in the visual field.
(*b*) Disturbance due to sharp or marked boundaries surrounding the stimuli.
(*c*) Reduced apparent brightness due to fusion of the dark area from the not-tested field of vision.
(*d*) Rivalry due to different fields for the two eyes.
(*e*) Distraction due to effect of bright areas near the stimuli.

A factor common to scores made under such varying distracting conditions is tentatively identified as *Resistance to Interference.*

The trend for the loadings in this factor to be higher on retest than on original test (true in 18 of 26 opportunities) indicates that this source of test score pollution tends to increase as the subject relaxes his vigilance when the material is more familiar. Thus for these tests we have decrease of efficiency with practice. This finding is borne out by the actually decreased means of Orthorater test scores on retest reported in OSRD Report No. 3969.*

The trend for such loadings to be higher for the left eye (12 of 16 times), indicates that subjects are more resistant when using the dominant eye.

*A Test-Retest Reliability Study of the Bausch and Lomb Orthorater with Naval Personnel, Applied Psychology Panel, NDRC, August 1944.

c. FACTORS IN DEPTH TESTS. (1) *Depth perception.* Factor E shows loadings on all depth tests and only on depth tests. It seems evident, therefore, that we are dealing here with *Depth Perception.* The loadings, however, vary from only moderate down to barely significant. Although these tests are labeled depth tests, the variance of their scores explained by that factor is in many cases no higher than that part of the variance explained by retinal resolving power. In the case of the Telebinocular, nonrelevant factors such as Form Perception and Resistance to Interference combined play a larger role than does Depth Reception *per se.*

(2) *Sightscreener depth specific.* Factor F has significant loadings only on the four Sightscreener depth tests, and is identified as something specific to that type of test for that instrument. That the Telebinocular depth test has a similarly high specific factor is shown by the residual of 0.38 remaining for the test-retest coefficient (29 with 30, table III). The Orthorater specific is probably erroneously dissipated among the depth factor loadings due to the similarity of task to that presented by the Sightscreener.

d. CAUTION INDICATED. The above discussion of factors indicates that all present visual tests, wall chart or machine, have large parts of their variance determined by nonpertinent variables. Several examples are given below to indicate just how bad the situation really is. No completely satisfactory test is available. The breakdown is according to percent of variance explainable by various factors. The pertinent factor and its numerical value are shown in italic.

Far Tests:

Snellen (RE)	Percent
Retinal Resolution	*57.8*
Form (Letter) Perception	25.0
Specific and Error	17.2
	100.0

Telebinocular (LE, Retest)	Percent
Retinal Resolution	*38.4*
Form (Letter) Perception	5.8
Resistance to Interference	5.3
Specific and Error	50.5
	100.0

Near Tests:

New London (LE, Retest)	Percent
Retinal Resolution	37.2
Accommodation	*30.2*
Form (Letter) Perception	15.2
Specific and Error	17.4
	100.0

Orthorater (RE, Retest)	Percent
Retinal Resolution	34.8
Accommodation	*38.4*
Resistance to Interference	12.2
Specific and Error	14.6
	100.0

Depth Tests:

Telebinocular, Far	Percent	*Sightscreener, Near*	Percent
Retinal Resolution	16.8	Retinal Resolution	13.0
Accommodation	----	Accommodation	4.0
Form (Letter) Perception	1.7	Form (Letter) Perception	----
Resistance to Interference	9.0	Resistance to Interference	0.8
Depth Perception	*10.2*	*Depth Perception*	*39.7*
Specific and Error	62.3	Specific and Error	42.5
	100.0		100.0

It should be noted that these tests are on the whole fairly reliable. Actual reliability coefficients (test-retest), for this sample, are:

Snellen, Far acuity	0.77
Telebinocular, Far acuity	.78
New London, Near acuity	.78
Orthorater, Near acuity	.85
Telebinocular, Far depth	.79
Sightscreener, Near depth	.76

One must not be led, however, into mistaking mere reliability for validity. What is measured is measured well, but apparently relatively little of what is measured is pertinent. It is concluded that much remains to be done if visual tests are to become really satisfactory. Furthermore, any validity studies based upon a single type of visual test can only, with great danger, be used to predict action based upon some other method of *supposedly* measuring the same function.

e. Factors in Phoria Tests. A few words as to the scales and methods used in measuring phorias is necessary to an understanding of the discussion of the factors.

Horizontal phorias are measured on a scale where low scores represent extreme esophoria and high scores represent extreme exophoria. Vertical phorias are measured on a scale where low scores represent extreme left hyperphoria and high scores represent extreme right hyperphoria. When the original measurements did not conform to this pattern scores were reversed to conform to those definitions.

A further difference, not resolvable by recoding, results from the treatment of the fusion problem. In the Sightscreener the subject is asked to report the farthest *excursion* observed at first glance before fusion brings the images to stable focus. On the other hand, Ortherater and Telebinocular measurements are usually taken *after* fusion has been maximized. One set of *excursion* readings was obtained on the Ortherater, however, and are included in this study. In the Maddox Rod, fusion is usually prevented by using an opaque card as an interrupter for the nonfocused eye. In this study the card was omitted, but fusion was probably not maximal.

Tables XV and XVI give the intercorrelations and residuals for two matrices of phoria tests. Table XV compares the Orthorater and Maddox Rod measurements while table XVI compares Orthorater, Sightscreener, and Telebinocular measurements. Table XVII shows the factor loadings used to obtain the residuals in tables XV or XVI. Three common factors of considerable importance, one other common factor and 4 specific factors removed by the analysis are described below.

(1) *Lateral phoria.* Factor A has loadings on all lateral tests, far (17–27) and near (28–38), and loadings are significant only for these tests. The factor appears then to be a general measure of *Lateral Phoria.* Loadings are higher on the far than on the near tests since the far tests measure only this general factor while the near tests all have loadings on a second factor (B) as well. The situation is analogous to the resolution-accommodation problem in far and near acuity tests. At far the Maddox Rod measures this factor common to all machine methods more poorly (24.0 percent) than does any machine (50.4 percent). Near tests do less well in general (6.2 percent to 49 percent of variance), while the Maddox Rod at this reduced distance does about as well (17.6 percent) as the other methods.

(2) *Near lateral phoria (convergence efficiency).* Factor B has loadings on only the *near lateral phoria* tests and so is given that name. Of these tests, p. 128, TM 8–300, says: "This is a test for imbalance at the ordinary reading distance, and may give the examiner some information as to the existence of refractive error, insufficiency of convergence, and may be indicative of a reduction of fusion control at usual reading distance." Reduction of fusion control can scarcely be the explanation since tests with control (32–33) have loadings as high as tests without such control (28–31). Since errors of refraction are considered important for all phorias this, too, can scarcely be the added factor. The best theory is that this is a factor dealing with *convergence efficiency* at the normal reading distance.

(3) *Hyperphoria (vertical phoria).* Factor C has loadings on all vertical phoria tests, far (1–9) or near (10–16). Loadings are about equal for the two sets of tests, although the Maddox Rod has a much higher loading when used as a far test. The loadings on this general vertical phoria factor are of about the same magnitude as those for near distance tests on the general lateral phoria factor but considerably lower than those for the far phoria tests on the general lateral factor.

(4) *Vertical-rest.* Factor D has loading only on the original tests for far vertical phoria. This may be due to better scores on original or worse scores on retest for those individuals whose defect was primarily due to refractive error and whose retest (afternoon) scores was consequently more affected by fatigue. The factor is not extremely important but does account for the higher loadings for retest on the Hyperphoria (Vertical) Factor above.

TABLE XV.—*Orthorater and Maddox Rod phoria tests, intercorrelations*

[Circled entries are remainders of reliability correlations]

Residuals

Test	1	3	4	9	10	12	13	16	17	19	20	21	22	27	28	30	31	32	33	38	39
1		48	41	35	36	35	29	31	22	16	08	12	09	15	34	28	25	21	14	18	-04
3	00		59	35	32	44	43	26	10	06	13	00	09	03	16	19	13	07	06	-04	-10
4	-07	(11)		41	50	50	58	37	06	00	11	03	03	13	13	12	12	05	08	10	-03
9	03	01	08		25	20	28	44	01	-02	06	-12	-07	-15	05	08	14	16	21	-15	-10
10	02	-04	02	-04		70	68	34	11	14	17	16	12	08	07	05	08	10	11	-05	-02
12	-02	04	-06	-11	-04		72	38	06	08	17	11	14	12	08	03	12	11	14	-03	-03
13	-07	03	05	-04	-02	(-05)		37	05	03	12	02	07	06	03	03	06	01	05	-06	-10
16	01	-11	03	13	02	06	-06		03	09	07	02	00	-17	-05	-02	04	01	00	-16	-23
17	-02	00	02	06	04	04	04	-04		86	77	80	71	59	59	48	35	22	16	44	15
19	-02	02	01	01	00	-01	-03	03	05		83	81	74	37	48	48	34	26	19	33	14
20	-12	07	09	03	-02	02	00	-02	-01	00		80	82	37	50	50	47	34	29	32	08
21	00	04	04	-04	-02	-01	-05	08	03	01	04		84	49	49	42	28	29	23	39	13
22	-07	07	00	-03	-04	-03	-03	01	01	03	13	(14)		46	45	45	36	34	30	37	11
27	-02	01	03	-05	00	-04	00	-01	07	-03	00	01	02		50	37	27	20	17	67	14
28	-01	00	01	-07	02	-01	-03	-05	01	-02	-02	-02	-06	-02		86	76	64	60	67	07
30	-01	05	04	-09	06	-06	-06	-06	00	02	-02	-04	00	00	08		76	68	62	57	07
31	-03	00	04	-07	01	-01	-04	-01	01	00	07	-04	02	-01	00	00		64	58	49	10
32	00	00	-01	05	04	01	-05	07	-07	00	02	-01	03	-12	-03	05	-04		86	39	-01
33	-04	01	02	09	03	-01	-03	06	-04	01	04	03	05	-11	00	04	-02	(32)		38	-02
38	02	-03	05	-07	-05	-11	02	07	-01	02	02	-04	-04	06	-03	03	-01	13	-10		14
39	01	-03	03	-03	00	00	-03	-14	03	02	-01	-01	00	00	02	05	11	-02	-02	04	

TABLE XVI.—*Phorias by three commercial devices, intercorrelations*

[Circled entries are remainders of reliability correlations]

Residuals

Test	1	2	5	6	7	8	10	11	14	15	17	18	23	24	25	26	28	29	34	35	36	37	40
1		55	43	26	34	21	38	27	09	25	20	18	17	21	14	19	34	33	33	25	19	16	00
2	(17)		31	39	24	41	47	47	30	29	14	16	08	22	10	09	15	15	19	20	04	01	-08
5	17	05		61	25	32	25	19	45	49	-01	-03	-04	08	02	10	11	14	00	01	04	06	-05
6	-02	01	(61)		23	40	31	27	76	80	-04	-05	-08	00	01	-03	01	02	03	-03	01	-03	07
7	-05	-04	-07	-10		63	24	19	20	30	06	-02	10	10	05	16	-03	01	05	02	-01	-13	-08
8	-03	07	01	01	(37)		29	32	33	42	-02	-10	-01	-01	-03	08	-08	-09	-07	-08	04	-11	-09
10	07	01	01	-10	-01	01		63	35	34	11	13	08	11	10	14	06	01	12	09	-02	-04	06
11	-02	03	-06	-11	-05	03	(26)		35	30	05	06	07	16	10	05	11	06	07	05	-04	-13	-11
14	-10	-05	-03	00	-07	-05	01	00		55	-12	-13	-08	-03	-14	-09	-12	-08	-04	-10	-11	-12	03
15	06	-01	00	06	06	10	07	00	(-05)		-02	-05	00	-05	04	-02	01	04	03	-05	07	03	03
17	02	-03	-04	-11	02	-01	04	-07	-06	04		88	62	65	64	64	59	51	55	51	41	36	02
18	-01	-02	-02	-08	-04	-05	-05	-06	-03	06	(-06)		62	65	63	63	62	58	57	55	45	40	11
23	-02	-09	-06	03	02	-04	-05	-05	04	06	04	02		74	56	57	53	48	64	53	45	40	03
24	02	04	04	06	04	-02	-02	04	05	-01	03	01	(21)		55	58	57	51	65	64	36	38	-05
25	01	01	-03	02	04	01	02	04	-07	05	00	-02	08	00		75	57	57	56	50	58	52	00
26	-01	-03	-01	-02	07	05	-06	-03	-03	-03	04	03	04	01	(19)		60	60	54	50	45	53	02
28	10	04	03	-01	-05	01	-02	05	-02	-03	-03	-04	-02	-02	-02	01		92	80	80	72	69	07
29	06	04	07	02	-01	00	-04	01	05	01	-05	-06	-03	-02	03	04	(13)		80	81	72	69	09
34	08	02	-04	08	01	-05	01	-03	-06	02	-03	-05	08	08	-01	-02	06	05		83	60	61	03
35	04	09	-05	06	00	-03	03	-01	-06	-03	-03	-03	00	09	-05	-06	08	07	(12)		63	66	03
36	-02	-03	10	00	03	13	-07	-10	-02	-01	-03	02	07	00	02	04	05	01	-03	-01		85	07
37	01	-05	01	01	-07	-02	-03	-15	-02	-04	-03	00	-04	-04	04	04	02	00	-03	02	(25)		02
40	-01	-08	-04	02	-02	-05	01	-01	00	-01	-04	04	-07	-03	-02	07	-01	-02	-03	-04	-04	-06	

791630—48——6

TABLE XVII.—*Summary of phoria factor loadings*

		A Lateral		B Near Lateral		C Vertical		D Vertical-Rest	E Orthorater (Near Verticle)	F Orthorater (Far Lateral)	G Sight Screener (Vertical)	H Maddox Rod
1	Far Vertical, Orthorater, 1	18	17	19	14	68	50	44	-01	04	-07	-02
2	Far Vertical, Orthorater, 1, Retest	—	12	—	1	—	73	-03	—	09	-05	—
3	Far Vertical, Orthorater, 2	01	—	08	—	67	—	—	05	—	—	-10
4	Far Vertical, Orthorater, 2, Retest	-01	—	07	—	71	—	—	22	—	—	05
5	Far Vertical, Sight Screener	—	07	—	00	—	40	20	—	-12	48	—
6	Far Vertical, Sight Screener, Retest	—	-08	—	04	—	57	11	—	-17	78	—
7	Far Vertical, Telebinocular (Reversed)	—	08	—	-16	—	41	46	—	-10	10	—
8	Far Vertical, Telebinocular (Reversed), Retest	—	-09	—	-18	—	49	07	—	-11	17	—
9	Far Vertical, Maddox Rod (Reversed)	-06	—	11	—	43	—	—	.13	—	—	-28
10	Near Vertical, Orthorater, 1	11	09	-04	-05	50	57	01	61	14	05	-03
11	Near Vertical, Orthorater, 1, Retest	—	08	—	-03	—	59	-04	—	06	06	—
12	Near Vertical, Orthorater, 2	07	—	02	—	54	—	—	76	—	—	07
13	Near Vertical, Orthorater, 2, Retest	02	—	00	—	73	—	—	71	—	—	-03
14	Near Vertical, Sight Screener	—	-12	—	-08	—	54	-02	—	01	57	—
15	Near Vertical, Sight Screener, Retest	—	-06	—	14	—	47	-01	—	-10	62	—
16	Near Vertical, Maddox Rod (Reversed)	00	—	-06	—	20	—	—	69	—	—	-35
17	Far Lateral, Orthorater, 1	91	81	-02	00	14	04	01	-16	50	03	11
18	Far Lateral, Orthorater, 1, Retest	—	82	—	05	—	04	01	—	56	-04	—
19	Far Lateral, Orthorater, 2	88	—	-01	—	-04	—	—	17	—	—	11
20	Far Lateral, Orthorater, 2, Retest	81	—	05	—	04	—	—	09	—	—	10
21	Far Lateral, Orthorater (Exc), 1	91	—	-08	—	04	—	—	02	—	—	-10
22	Far Lateral, Orthorater (Exc), 1, Retest	83	—	03	—	05	—	—	09	—	—	-17
23	Far Lateral, Sight Screener	—	71	—	10	—	10	-02	—	01	-04	—
24	Far Lateral, Sight Screener, Retest	—	75	—	05	—	11	-07	—	03	-08	—
25	Far Lateral, Telebinocular	—	71	—	15	—	-01	-02	—	11	-04	—
26	Far Lateral, Telebinocular, Retest	—	77	—	09	—	05	09	—	-06	03	—
27	Far Lateral, Maddox Rod	49	—	08	—	09	—	—	00	—	—	53
28	Near Lateral, Orthorater, 1	60	70	67	57	18	02	10	-11	10	02	16
29	Near Lateral, Orthorater, 1, Retest	—	64	—	63	—	03	16	—	08	-03	—
30	Near Lateral, Orthorater, 2	33	—	67	—	03	—	—	05	—	—	06
31	Near Lateral, Orthorater, 2, Retest	25	—	65	—	01	—	—	11	—	—	06
32	Near Lateral, Orthorater (Exc), 1	55	—	66	—	10	—	—	-10	—	—	-07
33	Near Lateral, Orthorater (Exc), 1, Retest	39	—	80	—	09	—	—	00	—	—	-12
34	Near Lateral, Sight Screener	—	69	—	47	—	09	02	—	03	01	—
35	Near Lateral, Sight Screener, Retest	—	66	—	49	—	03	05	—	01	00	—
36	Near Lateral, Telebinocular	—	42	—	70	—	08	04	—	06	-05	—
37	Near Lateral, Telebinocular, Retest	—	54	—	57	—	-01	-02	—	-11	02	—
38	Near Lateral, Maddox Rod	42	—	53	—	01	—	—	-02	—	—	58
39	Age	13	—	-05	—	-09	—	—	-08	—	—	12
40	Interpupillary distance	—	00	—	15	—	04	-10	—	13	02	—

(5) *Orthorater (near vertical) specific.* Factor E has loadings only on the three Orthorater near vertical tests. It is some function specific to those tests.

(6) *Orthorater (far lateral) specific.* Factor F has loadings only on the Orthorater far lateral tests. It is some function specific to those tests.

(7) *Sight Screener (vertical) specific.* The Sight Screener vertical phoria tests (both far and near) have relatively high loadings on Factor G. All other tests have insignificant loadings on this factor. Factor G is therefore some function specific to Sight Screener vertical phoria tests.

(8) *Maddox Rod specific.* Only the Maddox Rod tests have significant loadings on this factor (negative in sign for reversed scoring). Factor H is therefore some function specific to the Maddox Rod phoria tests.

That other devices possess similar specific factors is further indicated by the many large residuals remaining for the reliability (test-retest) coefficients found in tables XV and XVI. No attempt was made to remove such factors, since two tests do not adequately determine a factorial plane.

(8) *Interpupillary distance.* It was felt that this variable might be important for machine measurements, but was found to be unrelated to scores as measured by the present study. Since the fixed distances on the machines would be less or greater than actual individual differences, the results might have been expected to compensate for either esophoria or exophoria depending upon the direction of difference. If so, the near zero loading would be explained without demonstrating the unimportance of this variable. Further analysis might be indicated.

f. Warning on Phorias. As in the case of acuity and depth measurement, so for phorias, it appears necessary to warn against assumption of interchangeability of either interpretation of scores or validities from one machine or method to another. A few typical variance analyses, with the pertinent measurement in italic, are presented to show the dissimilarities.

Far Vertical:

Orthorater (2)	Percent	*Maddox Rod*	Percent
Far Vertical	*44.9*	*Far Vertical*	*18.5*
Specific and Error	55.1	Specific and Error	81.5
	100.0		100.0

Near Vertical:

Sightscreener	Percent	*Orthorater (2)*	Percent
Far Vertical	*29.2*	*Far Vertical*	*29.2*
Specific and Error	70.8	Specific and Error	70.8
	100.0		100.0

Far Horizontal:

Orthorater (2)	Percent	*Maddox Rod*	Percent
Far Lateral	*77.4*	*Far Lateral*	*24.0*
Specific and Error	22.6	Specific and Error	76.0
	100.0		100.0

Near Horizontal:

Telebinocular	Percent	*Orthorater (with Excursion) (1)*	Percent
Far Lateral	17.6	Far Lateral	30.2
Insuf. Convergence	*49.0*	*Insuf. Convergence*	*43.6*
Specific and Error	33.4	Specific and Error	26.2
	100.0		100.0

It would appear that here, too, considerable further research seems indicated.

g. Possible Relationship of Factors. Since the phoria factors were obtained from different matrices than those which yielded the acuity, depth, form, and resistance factors, it seemed highly desirable to ascertain the degree to which such factors were or were not interrelated.

Consequently a matrix involving 3 far acuity, 3 near acuity, 3 depth, 4 near lateral phoria, 3 near vertical phoria, 2 far lateral phoria, and 2 far vertical phoria tests was set up. Care was taken to include 3 form (letter) perception, 3 resistance, and several excursion (phoria) tests. The intercorrelations and residuals are reported in table XVIII.

In table XIX are reported the factor loadings which resulted in the residuals reported in table XVIII, along with the loadings (in parentheses) obtained for the same tests in their original matrices. With the exception of the special Sightscreener depth factor and other specific factors, all important factors, isolated previously in the separate matrices, reappear with loadings at least similar enough to leave all previous interpretations unchanged.

(1) *General lack of relationship.* The Retinal Resolution, Accommodation, Form (Letter) Perception, Resistance to Interference, and Depth Perception factors all have zero or near-zero loadings on the phoria tests. Similarly the Near Lateral (Insufficient Convergence) and Excursion (Anti-Fusion) factors have zero loadings on all acuity and depth tests.

TABLE XVIII.—*Factor integration matrix.*

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		64	72	39	49	33	32	24	34	00	-07	-08	-13	14	04	05	12	04	-09	03
2	00		74	52	45	36	49	35	40	11	01	-02	-10	07	08	11	20	04	02	-08
3	-02	00		47	58	33	38	29	41	-01	-04	-11	-22	12	13	-02	14	05	-03	-11
4	-03	-01	-03		50	49	27	27	22	-02	-09	-15	-24	11	17	10	00	-04	06	00
5	00	-01	00	02		42	24	26	21	-08	-14	-21	-30	05	08	-12	-02	-06	-05	-01
6	00	01	-01	00	01		06	18	17	01	-06	-07	-13	03	10	05	04	-01	00	06
7	-01	00	00	00	00	-01		53	46	06	05	-01	-08	-11	-06	01	06	-06	-10	-11
8	01	-02	01	-04	-01	01	-01		59	04	05	03	-12	-06	-09	-08	-03	06	-11	-10
9	02	-03	02	-02	-05	04	-02	01		05	16	08	-08	-15	-02	03	05	14	-06	-12
10	-01	03	01	-04	03	-03	05	-03	-03		58	80	61	12	-03	-04	55	44	03	08
11	00	00	02	-01	03	-01	03	-04	03	-02		72	64	-04	-19	-12	49	64	-12	-08
12	01	-02	00	00	-02	02	00	-02	-01	03	-03		72	06	-12	-06	59	55	01	03
13	02	00	-04	-05	03	03	-02	07	-01	00	07	-03		-02	-11	-11	41	40	00	01
14	-01	-03	00	00	00	-05	-01	01	-09	04	-02	00	00		35	34	10	10	31	25
15	-01	-02	01	-01	-01	01	06	-02	03	02	-05	00	04	-01		34	-12	-13	76	22
16	-02	04	03	01	00	03	-04	05	04	-07	10	-01	-04	03	00		-04	-13	36	42
17	01	01	-02	01	02	08	-01	00	-04	09	-01	02	01	-03	-01	-02		65	-04	-02
18	03	-04	-01	-03	-02	01	-05	-10	03	-01	04	-03	02	04	-03	08	00		-06	-07
19	00	00	01	01	01	00	02	-03	01	-01	-02	01	03	-01	00	-02	-01	-02		36
20	03	-01	-02	-02	03	-05	02	-04	-04	04	04	-03	04	-02	-10	06	02	02	02	

TABLE XIX.—*Factors in varied battery of visual tests.*

Name	A Retinal Resolution	B Accommodation	C Form (letter)	D Resistance to Interference	E Depth	F Lateral Phoria	G Near Lateral Phoria	H Vertical Phoria	I Maddox Maddox Rod	J Sight Screener (Vertical)
Far Acuity, Snellen	78 (70)	00 (11)	25 (41)	06 (10)	-06 (02)	-04	00	04	00	-13
Far Acuity, Orthorater, 1	77 (84)	12 (04)	08 (06)	29 (42)	10 (10)	02	01	05	09	04
Far Acuity, New London, Average	91 (90)	01 (06)	23 (33)	03 (-01)	-07 (00)	00	-07	[illegible]	03	-07
Near Acuity, Orthorater, 1	48 (50)	52 (66)	-05 (02)	24 (31)	01 (10)	-17	-01	17	15	-10
Near Acuity, New London, Average	49 (57)	43 (54)	41 (23)	01 (05)	02 (-10)	-17	-08	-07	14	05
Near Acuity, Telebinocular (Circle), Retest	39 (52)	47 (43)	00 (03)	01 (26)	-14 (61)	-17	07	11	01	-09
Depth, Far, Telebinocular	42 (41)	-06 (00)	04 (-13)	40 (30)	53 ([illegible])	-02	00	-17	01	00
Depth, Near, Sight Screener, Retest	33 (34)	24 (23)	05 (01)	00 (19)	73 (63)	-07	08	-12	14	07
Depth, Far, Orthorater, 1	46 (47)	06 (00)	-01 (11)	-01 (-01)	51 (71)	05	[illegible]	-12	09	01
Near Lateral, Sight Screener	03	02	-01	04	00	58 (69)	54 (47)	11 (09)	02	-04 (01)
Near Lateral, Maddox Rod	-06	05	-07	-01	07	65 (42)	42 (53)	-02 (-01)	33 (58)	-06
Near Lateral, Orthorater, 1	-06	00	09	00	07	74 (70)	62 (52)	05 (02)	-01 (16)	-02 (02)
Near Lateral, Telebinocular	17	-15	01	12	-10	53 ([illegible])	58 (76)	-03 (08)	-02	-02 (-05)
Near Vertical, Orthorater, 1	13	00	08	-07	-10	13 (09)	-07 (-05)	51 (59)	-11 (-03)	-04 (05)
Near Vertical, Sight Screener	08	05	05	-05	00	-16 (-12)	-04 (-08)	76 (54)	06	45 (57)
Near Vertical, Maddox Rod (Reversed)	08	-06	-13	12	12	-04 (00)	-06 (-06)	55 (20)	-52 (-35)	-08
Far Lateral, Orthorater, 1	17	03	01	12	06	81 (81)	-03 (00)	01 (04)	02 (11)	-01 (03)
Far Lateral, Maddox Rod	05	16	03	-06	00	79 (49)	01 (08)	-06 (09)	25 (53)	08
Far Vertical, Sight Screener, Retest	-04	01	01	00	06	-03 (-08)	-01 (04)	68 (57)	-10	56 (78)
Far Vertical, Maddox Rod (Reversed)	-04	11	06	-15	00	[illegible] (-06)	[illegible] (11)	41 (43)	-35 (-28)	03

(2) *Lateral phoria and accommodation.* The lateral phoria factor does have consistent negative, although not quite significant loadings (−0.17, −0.17, −0.17) on the three near vision tests (4–6). These loadings indicate a *positive* relationship between extreme *exophoria* and *difficulty on near acuity tests.* This finding has bearing upon the *accommodative theory* of phorias, as set forth in TM 8–300, pp. 131–132, which holds that—

> Esophoria and exophoria are manifestations of muscular asthenopia due to hyperopia and myopeia, respectively. In the hyperope accommodation exceeds convergence. The excessive or unused stimulation to converge * * * may lead to convergence excess * * * When the action of the fusion centers is lowered, one eye will deviate inward * * * In myopia the reverse occurs. Accommodation is less than convergence * * * When this occurs the urge to converge is weak, as accommodation is the dominating impulse. The external recti soon take advantage of the weak convergence stimulation and divergence excess results. When the action of the fusion centers is lowered one eye will deviate outward * * *.

Thus being high on the general Lateral Phoria factor, which means being highly exophoric, will frequently be associated with myopia. The myope will be less able to read the near vision tests at *normal* reading distance and thus we have the negative loadings found in this study.

(3) *Vertical phoria and depth perception.* The three depth perception tests (7–9) all have negative just nonsignificant loadings (−0.19, −0.12, and −0.12) on the general Vertical Phoria factor. Of the relationship between phorias and depth perception TM 8–300, p. 109, says:

> In order to fix both eyes upon an object the defective muscle must receive a greater amount of enervation than its fellows. The increased amount of enervation is interpreted into distance by the cerebrum just as it is in refractive errors, therefore the relative position of objects is inaccurately judged.

To explain why vertical rather than lateral phoria has the loadings for the plates used in the study is fairly simple. Displacement on the printed charts is most likely horizontal for the circle (or other form) supposed to be discriminated as nearer. Even though further lateral displacement takes place, the stimulus object should still appear highest (have most displacement) if the phoria is in the same direction, or the only flat object, if visual displacement was opposite and equal to printed displacement.

In either case the stimulus object would be unique. In the case of vertical phoria however, the vertical visual displacement would be equal for all objects and all would appear equally solid. If this visual

vertical displacement were large in proportion to the printed lateral displacement, the latter might no longer be discriminated. Thus the low negative loadings.

h. Final Warning. In any case the two relationships are of extremely small scope, accounting for only about 3 percent of the near acuity or depth scores. In general it is concluded that all eight factors are relatively independent and that any thorough study of the role of visual factors in military or other occupational success must be aware of the possible importance of any one or any combination of them.

By way of summary, these factors have been finally or tentatively identified as—

- Retinal Resolution.
- Accommodation.
- Form (Letter) Perception.
- Resistance to Interference.
- Depth Perception.
- Lateral Phoria.
- Near Lateral Phoria (Convergence Efficiency).
- Vertical Phoria.

4. Frequency Distributions and Item Analyses of Tests Given at Fort Dix, N. J.

In this paragraph, some initial attempts to study how well the tests worked as measuring devices are reported. Since no independent criteria are available, these studies are confined to examination of the total scores attained and the behavior of individual items in terms of the usual practices of evaluation. There is little or no information available as to what the distributions of test scores should look like nor are there validity criteria to serve as a basis for evaluating the items. Thus we are forced to look at the tests as unstandardized instruments and assume that from the test data alone deficiencies can be detected and ways found to improve the tests for any future use.

Two sets of data are presented:

Frequency distributions of the scores on the 14 chart-type tests. These are presented in histogram form in figures 17 through 30. Test scores, so distributed, are based on scoring methods C and A as selected for the first-test and retest factorial analyses discussed in paragraph 2 of this section.

Item counts for 2 groups of 300 cases each, selected by systematic sampling with odd and even man-number as the basis. Each group of 300 was further divided into thirds on the basis of total test score (Method G, number right to three consecutive misses). This resulted in high, middle, and low score groups of 100 cases each for the two samples. Graphic item counts were made of each subgroup and the results are graphically presented in figures 31 through 44.

a. Frequency Distributions. The initial impression from examining figures 17 through 30 is that all the tests in one respect or another are lacking in the requirements of good measuring instruments. Most of the tests (1, 2, 3, 4, 7, 8, 9, 11, and 14) are obviously negatively skewed meaning that scores tend to pile up at the high score values (abscissa). This may indicate that the item values presented in these tests are too easy for a large part of the group tested. Or it may be that the "real" distribution of visual abilities tested are piled up in the same way. A third possibility is that the method of testing, starting with easy items and progressing through more and more difficult ones, may "push" the examinees to higher scores. In other words, the better eyes get considerable practice before the items become difficult enough to cause errors thus allowing such examinees time to try various modes of response. Which of these causes or what combination of them is producing negative skewness is difficult to determine. Probably all contributed to different extents for different tests.

Whether or not negative skewness was present, certain of the tests were conducive to inordinately high scores. Triangle discrimination (6) is such a test, with 3.1 percent of the sample coming up with perfect scores. This test is not predominantly skewed in the negative direction. A test that is skewed, AAF Constant Decrement (11), likewise has a walloping group close to the top score (8.4% in the 30–32 step interval). These examples reveal a basic weakness of such tests; the items do not extend far enough at the difficult end of stimulus range. The "real" distributions of abilities in acuity are unknown and in any case will be a function of the scales employed to measure them. It is possible that individuals will tend to pile up at the positive ends of such scales. The distribution of previous left-eye Army Snellen scores cited in table I (see sec. II) indicates that this may have been true of this sample. Yet, the Snellen test by its very design can lead to the impression that the population is skewed. In terms of the conventional fraction there are only two size categories of letters below the 20/20 line while six exist above 20/20. Naturally, there is a limit imposed on the ablest examinees by this scarcity of items at the more difficult levels. At the same time the unmotivated can quit anywhere along the line and hence may give the distribution the long tail in the negative direction.

These effects seem to have been eliminated partially by the inclusion of more difficult items in New London, AAF Letter and the checkerboard tests (5, 7, 8, and 13). In the distributions of these tests we find scores spreading out somewhat more toward the positive end of the scale, especially in test 13, and less tailing off toward the negative end.

However, there is an artifactual explanation of the latter effect; namely, these tests (and most of the other tests as well) started at too difficult levels.

Which brings us to the second obvious deficiency of the tests—bimodality. Tests 3, 8, 11, and 13 are good examples of this effect. Notice the percentages of cases which pile up on the lowest score interval on these tests (10.3, 10.3, 8.0, and 8.0 percent respectively). Other tests demonstrate this overdose of zero or near zero performance as well. In any case, the data indicate that the tests are too difficult for the poorer eyes to even get started. Only in the Practice Test (1), Snellen (2), and Vernier Test (14) do the instruments come close to getting all examinees off with equal opportunity to respond correctly on the first item.

An inspection of frequency distributions of scores attained by a larger sample (791) cases throws some light on this question of the tests starting at stimulus values too difficult for part of the group tested. These scores are based on scoring method G, number right to three consecutive misses, but there is no reason to believe that they differ appreciably from those obtained by method C in the sample of 261 cases. On the assumption that examinees who normally wear glasses represent a large part of the group for whom certain tests are too difficult, it was felt that exclusion of those wearing corrective lenses would eliminate the minor mode at near zero score and otherwise improve the bottom half of the distribution. Table XX shows the frequencies and percentages obtained on test 13 by total group and by subgroups on the basis of continuous, for reading only, and no use of corrective lenses. Test 13 is chosen for this purpose since it has good spread at the difficult end of the scale. In table XX it can be seen that eliminating those who wear glasses from the sample results in an improved distribution of the remaining cases. Comparing the No Correction and Total Group data of the table, it is quite apparent that the minor mode at near zero scores drops out when the test is applied only to the better eyes (No Correction). The other columns of table XX show how those wearing glasses continuously, or for reading only, distribute in test score. Note that they concentrate most in the bottom half of the distribution and that it is on this part of the score range that they effect the total-group distribution.

This analysis indicates that certain test score distributions can be expected to improve as the range of item values is selected to fit the populations. But in seeking to improve the curves the use to which the test will be put should be constantly considered. If the aim is to screen out individuals with poor acuity, the negatively skewed or bimodal distribution is desirable since the cut-off score may be set between the modes with considerable efficiency. Such a step depends on a knowledge of validity. If a positive relationship is shown between acuity test score and performance on some job, it is immediately apparent that in a test with negative skewness and a minor mode at zero score, a cut-off score for screening out the poor individuals can be set without worrying about the shape of the distribution. A good example is the present distribution for test 11 (fig. 27), in which a screening cut-off score of 12–14 might be set, thus eliminating the 18 to 20 percent of the people with lowest scores. Where the cut-off point is set depends, of course, on the validity found for given purposes and the selection ratios adopted. However, assuming validity, it is doubtful that switching to a scale which produces more normal distributions will do any better job for screening purposes.

TABLE XX.—*Frequency table for test scores obtained on Checkboard Variable Grid Test (18) by groups who did or did not wear corrections and by total group*
[Based on first-test scores of 791 enlisted men at Fort Dix]

Test score	Wear continuous correction		Wear correction for reading		No correction		Total group	
	F	%	F	%	F	%	F	%
42–44	0	0	0	0	0	0	0	0
39–41	0	0	0	0	11	1.7	11	1.4
36–38	0	0	0	0	8	1.3	8	1.0
33–35	2	2.7	0	0	14	2.2	16	2.0
30–32	1	1.3	2	2.4	30	4.7	33	4.2
27–29	0	0	1	1.2	42	6.6	43	5.4
24–26	1	1.3	2	2.4	85	13.4	88	11.1
21–23	4	5.3	10	12.0	148	23.4	162	20.5
18–20	11	14.7	16	19.3	171	27.0	198	25.0
15–17	5	6.7	12	14.5	71	11.2	88	11.1
12–14	3	4.0	2	2.4	15	2.4	20	2.5
9–11	8	10.7	10	12.0	22	3.5	40	5.1
6–8	8	10.7	8	9.6	8	1.3	24	3.0
3–5	31	41.3	20	24.1	8	1.3	59	7.5
0–2	1	1.3	0	0	0	0	1	.1
Total	75	100.0	83	99.9	633	100.0	791	99.9
Mean	10.43	......	13.35	......	21.54	......	19.63	
σ	8.20	......	7.85	......	6.30		7.73	

On the other hand any test designed to discriminate over the whole range of visual abilities should be revised to measure in ways most conducive to the uses intended. It is quite conceivable that jobs will be found on which different ranges of resolution ability will contribute to success. Some jobs may demand working to very close visual tolerances while others demand less exacting ones. The degree of skill required by the two jobs may appear on the same continuum, or factor, but a different difficulty level. When validity studies are

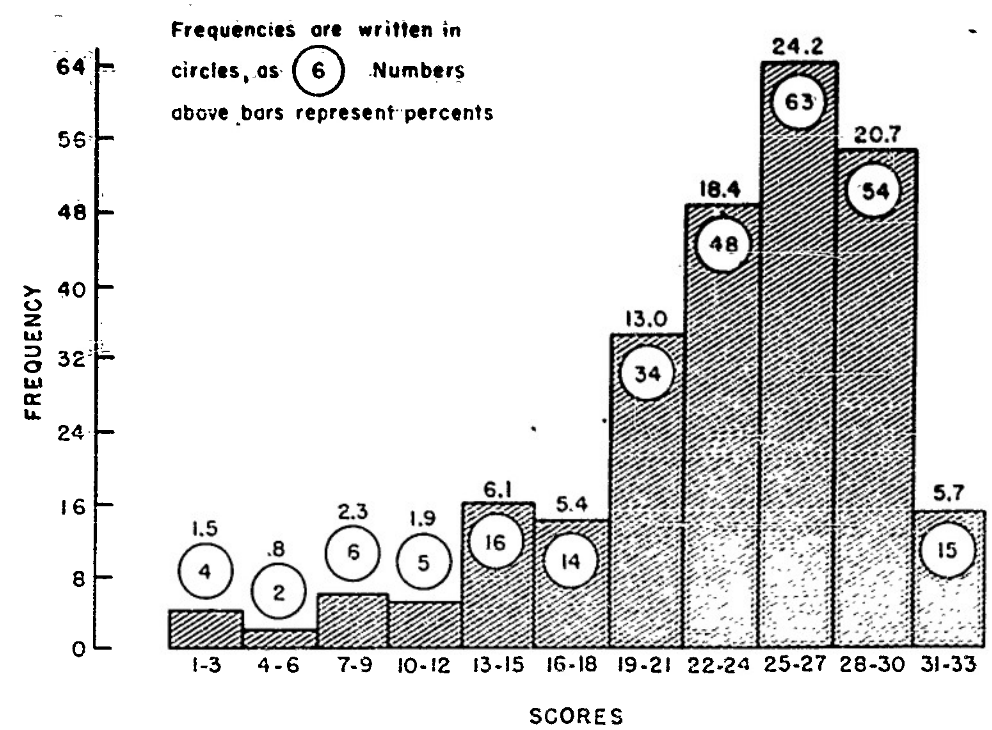
made, it may be found that differential resolution ability at the less difficulty level predicts success in the job requiring wide tolerances, while for jobs requiring close tolerances differential acuity at the difficult test level predicts success. If such is the case, tests will be required which measure sensitively over both difficulty ranges.

To summarize the salient features found in the distributions presented, each test is discussed briefly in turn.

(1) *Test 1 Practice* (fig. 17). This test starts with items permitting practically all examinees to get on the scale. However, the distribution of scores is negatively skewed with 5.7 percent of the cases achieving perfect or near perfect scores. As it stands this test may be a good screening test but requires more difficult items if selection and classification are to be accomplished at the positive end of the scale.

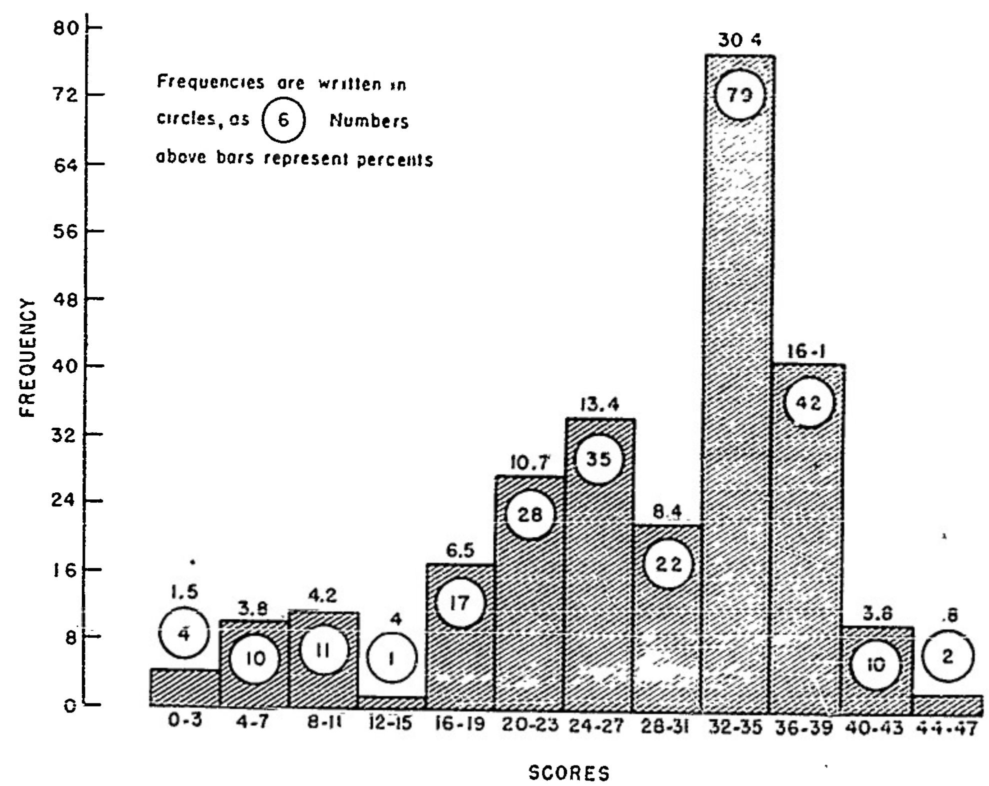
(2) *Test 2, Army Snellen* (fig. 18). The test at present starts with easy enough items but has an insufficient concentration at the difficult levels, hence, negative skewness. The final line of the test (nine items) seems difficult enough to spread men out at this end of the range; but below 20/20 there are only two scale values 20/15 and 20/10, an insufficient concentration of values to discriminate in this area. In addition, the test performs in spotty fashion throughout the range, with indication of multimodal distribution of the scores. This may be a result of inadequate standardization of the items leading to reversals in the letter difficulty continuum or it may be the result of heterogeneity in the population tested.

(3) *Test 3, Dot Variable Size* (fig. 19). The test starts at a level too difficult for many examinees (10.3 percent of the scores falling between 0 and 3). This can be corrected by including items at the larger size values. The test is not too easy. The highest score attained is in 36–39 category, while the test ran to 48 items. Nevertheless, negative skewness appears. This may be explained as either the result of coarseness of the scale of stimulus values at the points where scores pile up or the result of the impurity of the test. Certainly, the present test will be improved by the inclusion of more items at values where present scores are concentrated. But, beyond this corrective measure lies the possibility of the present distribution being a composite of two visual functions and correction may require teasing these out. Factorially, the Dot Variable Size Test has been shown to measure both resolution (factor I) and brightness discrimination (factor II). How these operate, together or singly, in the case of given individuals is obscure in the present data. Four possibilities are suggested: (1) Certain individuals always respond in terms of "seeing the dot" (resolution); (2) certain others always respond in terms of "seeing something" (brightness); (3) individuals at varying points in the test shift from one function (resolution) to the other one (brightness) as



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=23.2 σ=6.1).

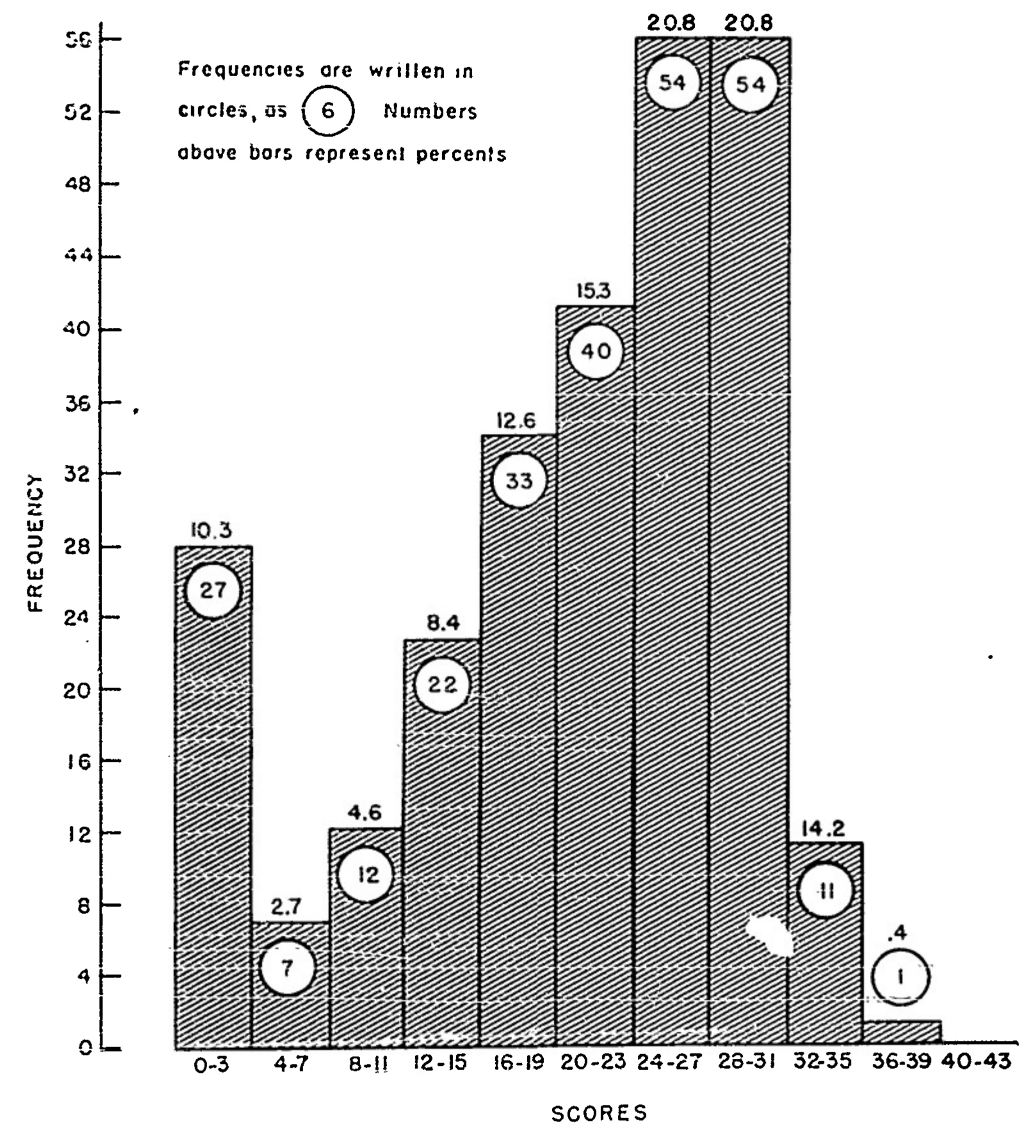
FIGURE 17.—*Frequency distribution for Practice Test—Test 1.*



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=28.2 σ=9.5).

FIGURE 18.—*Frequency distribution for Army Snellen Test—Test 2.*

the test becomes difficult; (4) individuals respond in terms of both functions simultaneously throughout the test. It appears that certain of these possibilities may account for part of the skewness in the present distribution. By combining functions or shifting to the brightness judgment, individuals may have acquired higher scores than would have been forthcoming on the basis of resolution judgment alone. Some evidence in support of this hypothesis is supplied by the performance of test 13 (fig. 29). As we have seen, this factorially rather pure test is less skewed although it employs a scale of stimulus values similar to that used for the Dot Variable Size Test. The Checkerboard Test values are slightly larger. This information suggests that the two-factor picture (I and II) in the case of the dot test



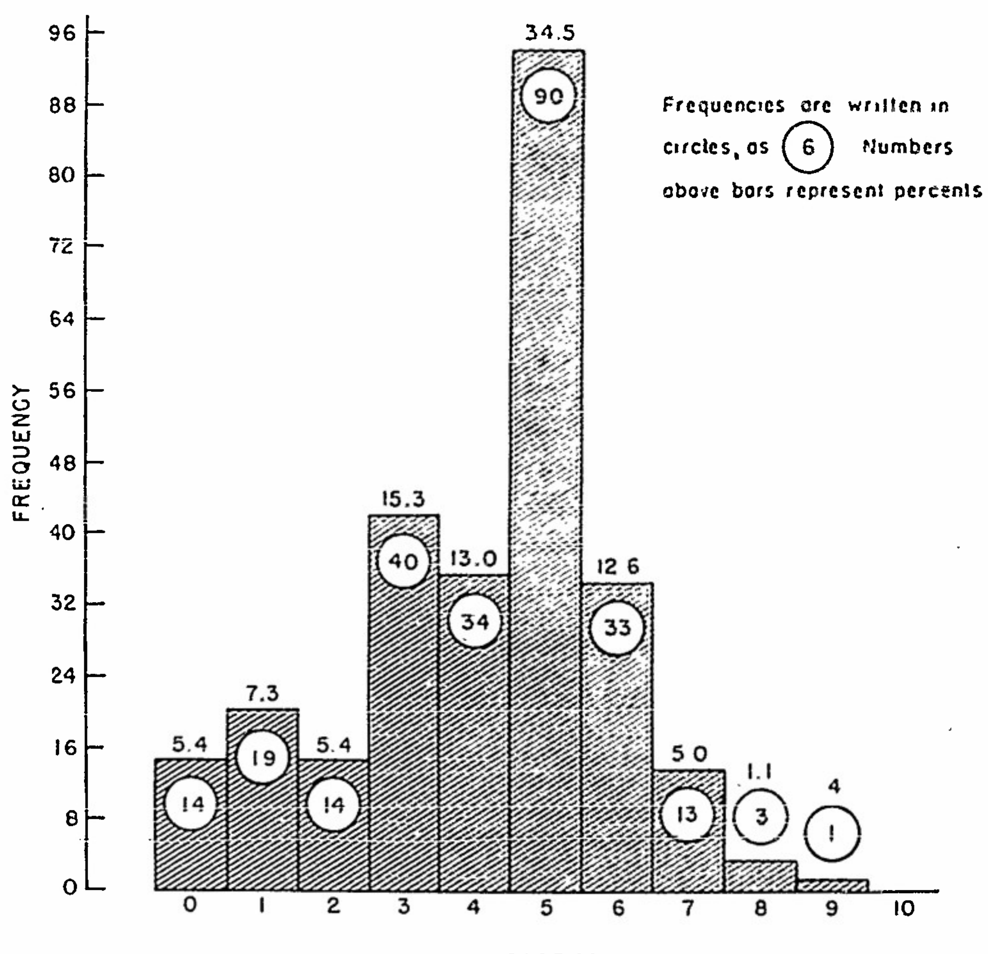
Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=20.1 σ=9.3).

FIGURE 19.—*Frequency distribution for Dot Variable Size Test—Test 3.*

is, in part, responsible for the skewness in that test. Unfortunately, we did not take the opportunity to get introspective reports of a sample of the population as to what method of responding was individually used throughout the test. If we had done so we might, by separate analyses, either have confirmed or placed in disrepute the hypothesis.

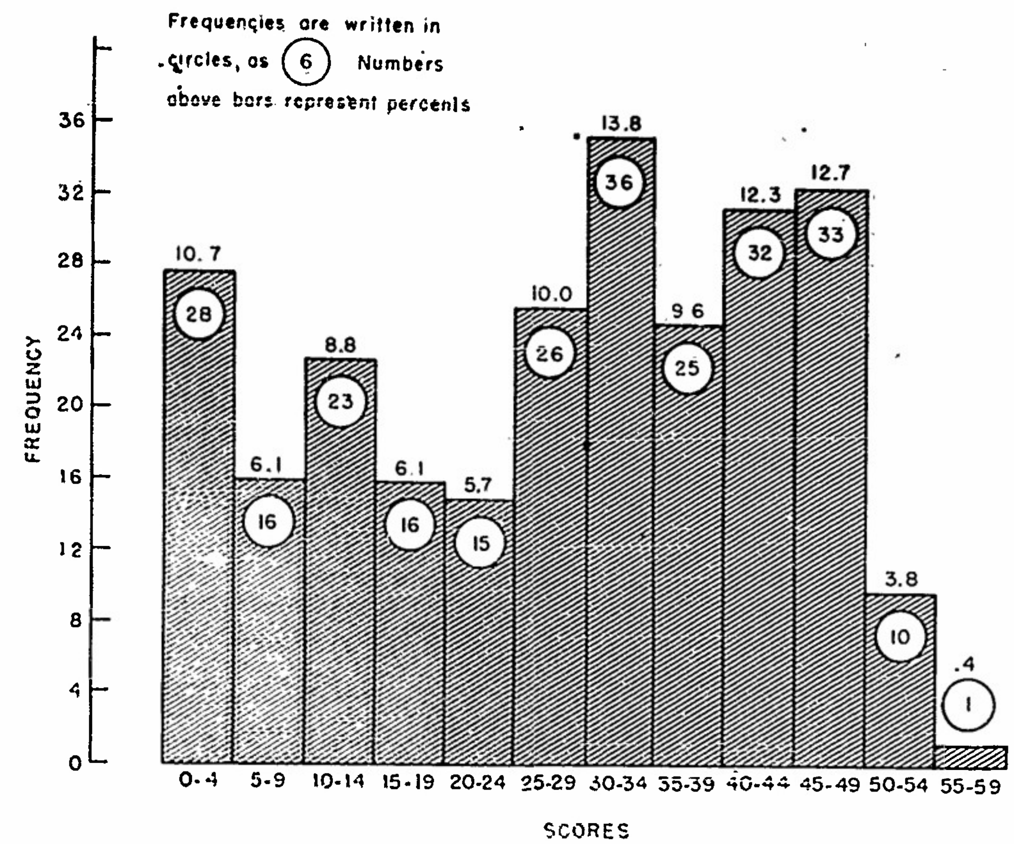
(4) *Test 4, Quadrant Variable Contrast* (fig. 20). This test, while too difficult for some examinees (5.4 percent at zero score), gives a rather good distribution. The tremendous group with scores of 5 (34.5 percent of the sample) indicates a break in the scale of items or something peculiar to items 5 and 6. This is confirmed by the item analysis which follows this discussion. For many individuals item 5 was no more difficult than item 4, but item 6 was extremely difficult for these people. The explanation of this result probably lies in improper scaling of the items or some other feature of the test structure.

(5) *Test 5, New London Letter* (fig. 21). The 10.7 percent of the



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=4.1 σ=1.9).

FIGURE 20.—*Frequency distribution for Quadrant Variable Contrast—Test 4.*



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=27.8 σ=15.6).

FIGURE 21. *Frequency distribution for New London Letter Test—Test 5.*

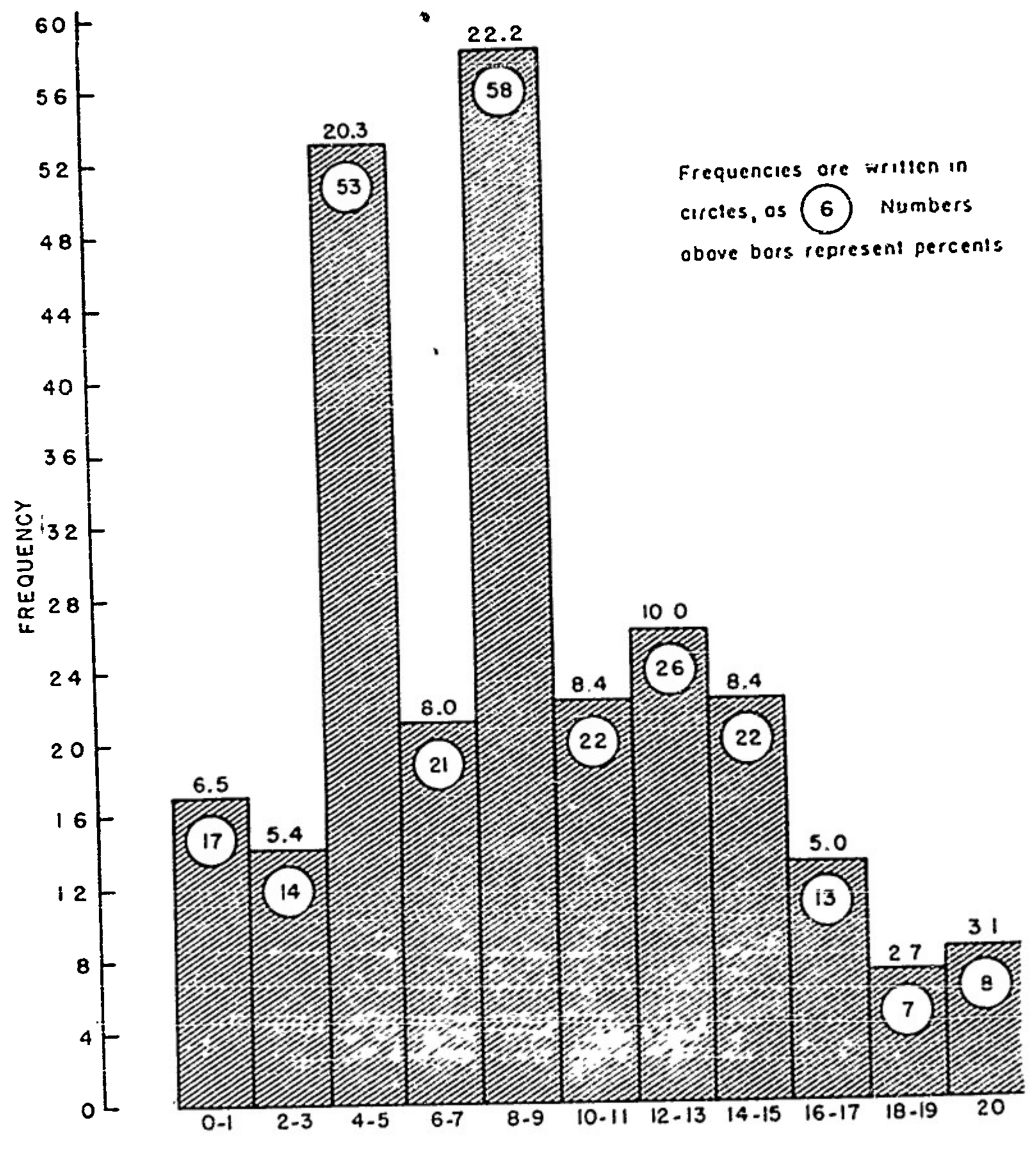
cases in the sample who have scores of from 0 to 4 indicates that this test began with items too difficult for the group tested. On the opposite end of the scale it is apparent that the test was not too easy. The "best" examinee scored 57 on a test 82 items in length. Between these extremes the test scores are distributed somewhat differently than in the case of other tests. Negative skewness is apparent to some extent; but, in general, the distribution is closer to a rectangular one than in the other instruments. The AAF Letter Test (fig. 24) which was composed of strictly comparable stimulus values (size) presents a different picture. Part of the explanation for the uneven distribution of scores in the New London Test is the high variability in difficulty exhibited by items of the same size. This will be discussed later in terms of the item analyses.

(6) *Test 6, Triangle Discrimination* (fig. 22). The test is too difficult for some and too easy for others as exhibited by the scores at 0 to 1 and 20 categories (6.5 and 3.1 percent of the sample, respectively). There is a slight tendency toward positive skewness meaning an excess of difficult items, but this is confused by the sharp modes at 4 to 5 and 8 to 9 score intervals (20.3 and 22.2 percent of the cases). Again,

701639—48——7

irregularity in the scaling of items is suspected as contributing to these jumps in the distribution.

(7) *Test 7, Bausch and Lomb Checkerboard* (fig. 23). The test is too difficult for certain individuals. It is not too easy for those with better vision. It also appears to be multimodal and leptokurtic (scores piling up within a narrow range of values, 23.8 percent at scores 14–15). The minormode at zero or near zero scores may be eliminated by the addition of easy items as indicated in the previous discussion of test 13 (see table XX and discussion). Otherwise the distribution seems good for purposes of screening and selection. A tendency to

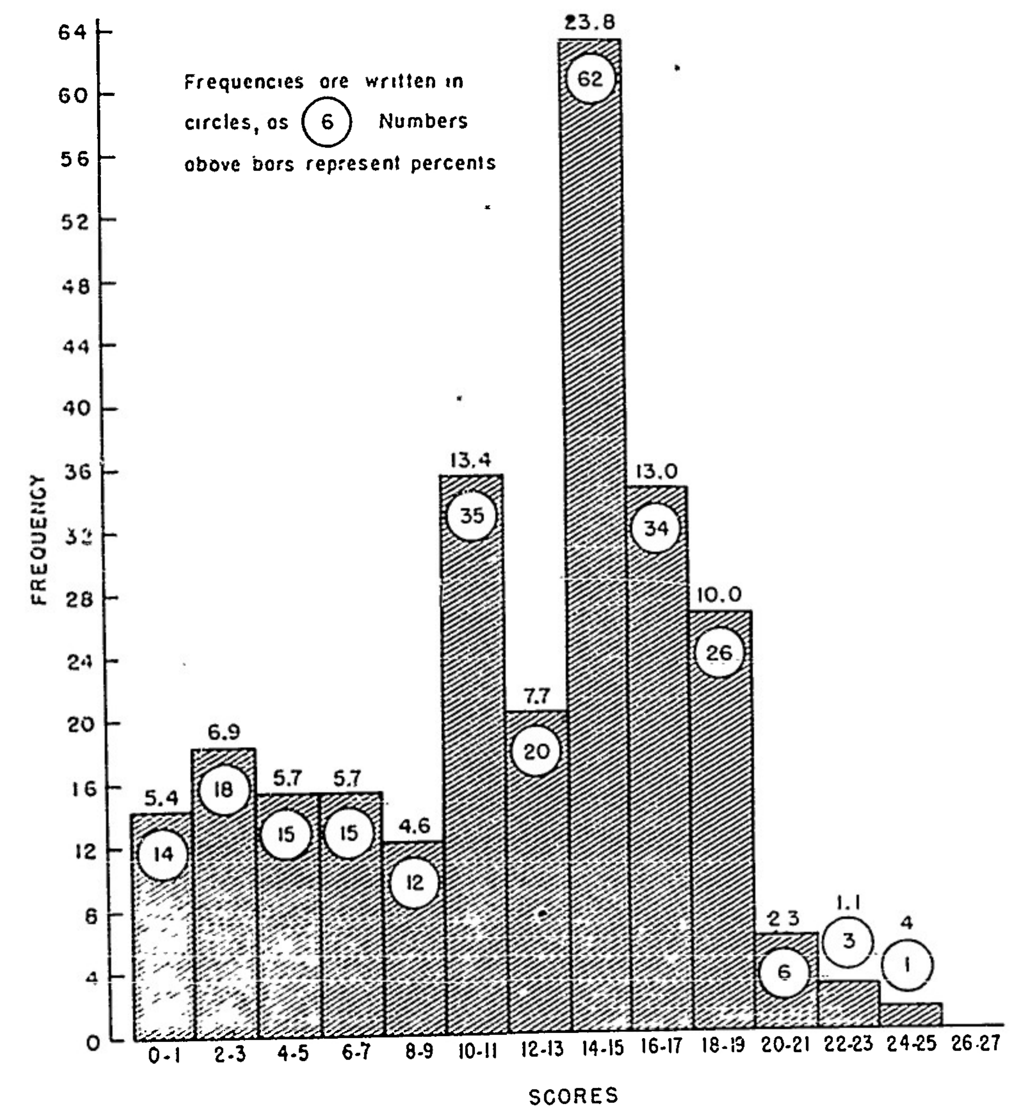


Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 ($M=8.8$ $\sigma=4.9$).

FIGURE 22.—*Frequency distribution for Triangle Discrimination Test 6.*

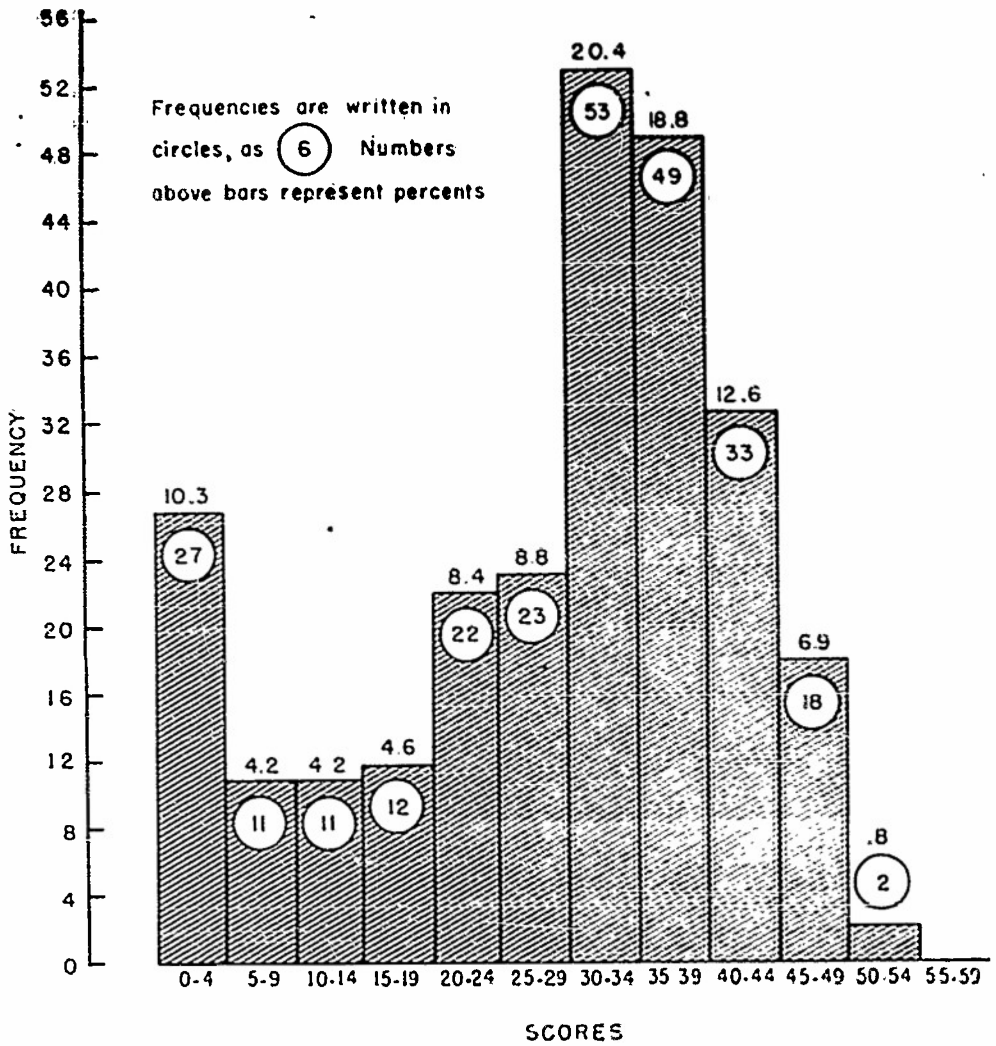
negative skewness may be adjusted by additional items at appropriate levels of difficulty.

(8) *Test 8, AAF Letter* (fig. 24). This is another test with a minor mode at zero and negative skewness or a tendency for scores to pile up at the positive end of the scale. No one, however, scores perfectly on the test. The smoother appearance of this distribution, comparing it with that for New London Letter (fig. 24) may be a function of better standardization of the items. The factorial analyses (see par. 2, this section) show that a form (letter) factor operates in all letter tests to a small but rather significant extent. It is inconceivable that this operates to the same hypothetical end as that suggested for the brightness discrimination factor in the case of the Dot Variable Size



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 ($M=11.8$ $\sigma=5.6$).

FIGURE 23.—*Frequency distribution for Bausch and Lomb Checkerboard—Test 7.*



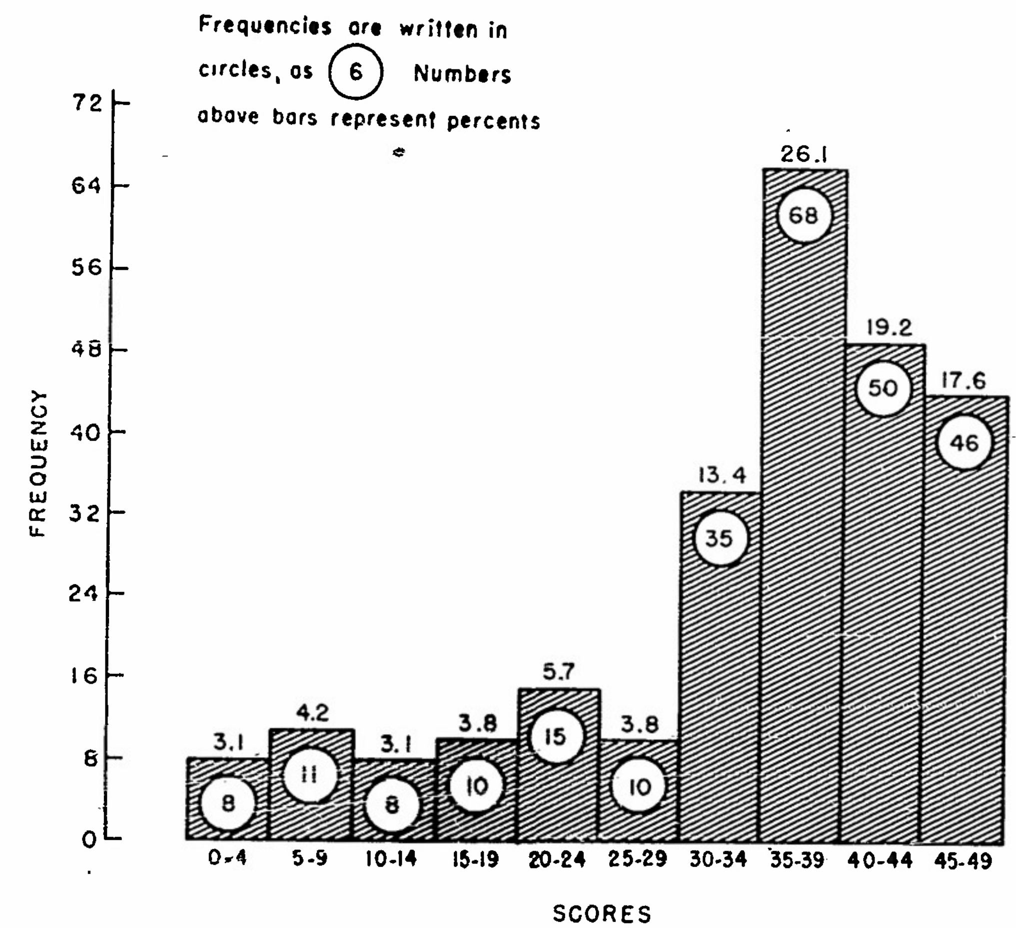
Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 ($M=28.2$ $\sigma=13.4$).

FIGURE 24.—*Frequency distribution for AAF Letter Test—Test 8.*

and Line Resolution Tests (tests 3 and 9). In those tests a switch from resolution to brightness judgments can apply to each of the succeeding items. However, in the letter tests it is possible that certain letter forms are more prone to being called correctly on the basis of their meaningful form than others. If such is the case, the uneven performance of certain items at different points in the scale may be explained. At the same time, the same phenomenon may explain the tendency for scores to pile up at the positive ends of letter scales. Sufficient correct responses to certain letters on the basis of ease of form identification may prolong the test beyond the point where certain other letters of the same size are too difficult and do not lend themselves to easy identification with any appreciable reduction in spatial cues. This is not particularly true of the AAF Letter Test alone. It is suggested here as a general possibility in all letter or form

identification tests and may well be a stumbling block in the path of scaling such items on a difficulty continuum.

(9) *Test 9, Line Resolution* (fig. 25). Negative skewness is immediately evident. The test for various reasons is too easy. We have seen the possible result of a shift to brightness judgments in the case of dot variable size. In this line test the effect may be even greater because the line remains constant in length. Reference to the specifications in appendix A reveals that the size (width of line) values



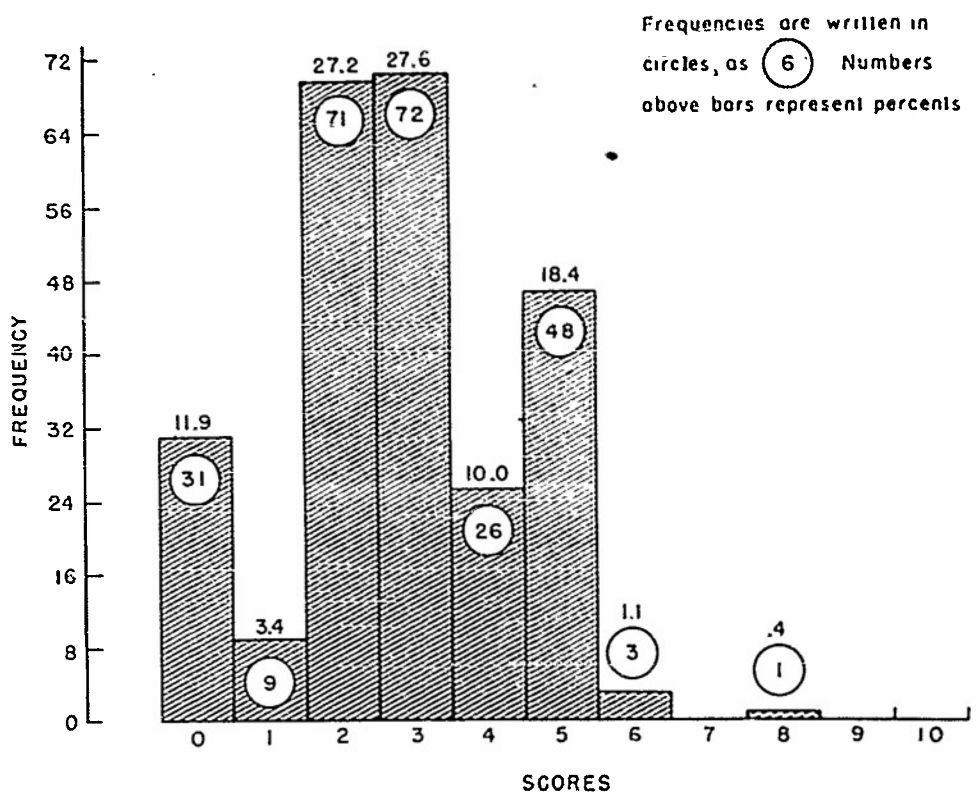
Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 ($M=34.0$ $\sigma=11.8$).

FIGURE 25.—*Frequency distribution for Line Resolution Test—Test 9.*

employed in this test are the smallest of all the test objects in the study. The smallest stimulus value is a line width of 0.0015 inch. This size is close to the limit of exact line production by the photographic methods employed in the present study. Any further decrements in stimulus size will require other methods such as the intaglio engraving process.

(10) *Test 10, Dot Variable Contrast* (fig. 26). This test is undoubtedly too difficult, 11.9 percent of the cases failing to score.

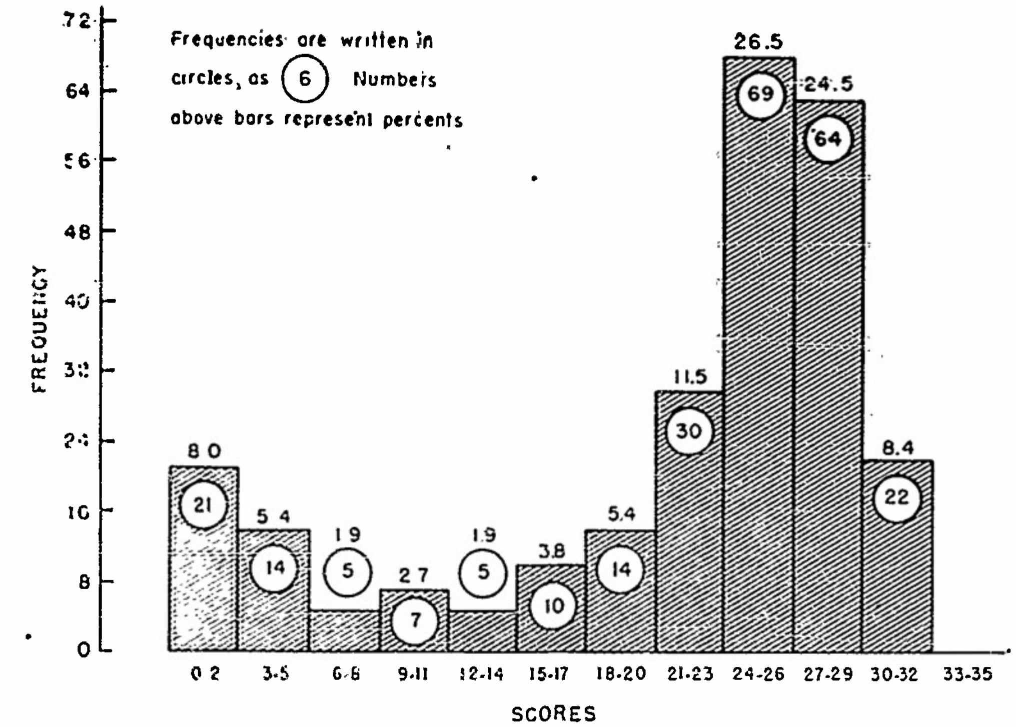
Beyond this obvious point and general positive skewness, little can be said about the distribution. Undoubtedly the size of the spot played a role in this test as well as the contrast variable. Future tests should incorporate larger spot areas which, as may be seen in the quadrant contrast test (fig. 20), will improve the distribution. At the same time, the impurity of these contrast tests should be remembered. They both measure resolution ability to a degree exceeding the amount of variance attributable to contrast discrimination.



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 ($M=2.8$ $\sigma=1.6$).

FIGURE 26.—*Frequency distribution for Dot Variable Contrast Test—Test 10.*

(11) *Test 11, AAF Constant Decrement* (fig. 27). Previous discussion of this distribution pointed out that it is clearly bimodal and negatively skewed. This is a natural consequence of the scale employed, one Snellen step decrement from item to item (20/40 to 20/5). Such scaling resulted in 20 items that subtended more than 1 minute of visual angle (20/20) and 15 that subtended less than 1 minute. This is counter to the logarithmically scaled letter tests in which the majority of items subtended less than 1 minute of angle. If closer to normal distributions are desired the logarithmic scale works best. The AAF constant decrement scale results in marked skewness, because of the coarseness of the scale, for items subtending less than one minute of visual angle.



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 ($M=21.2$ $\sigma=9.1$).

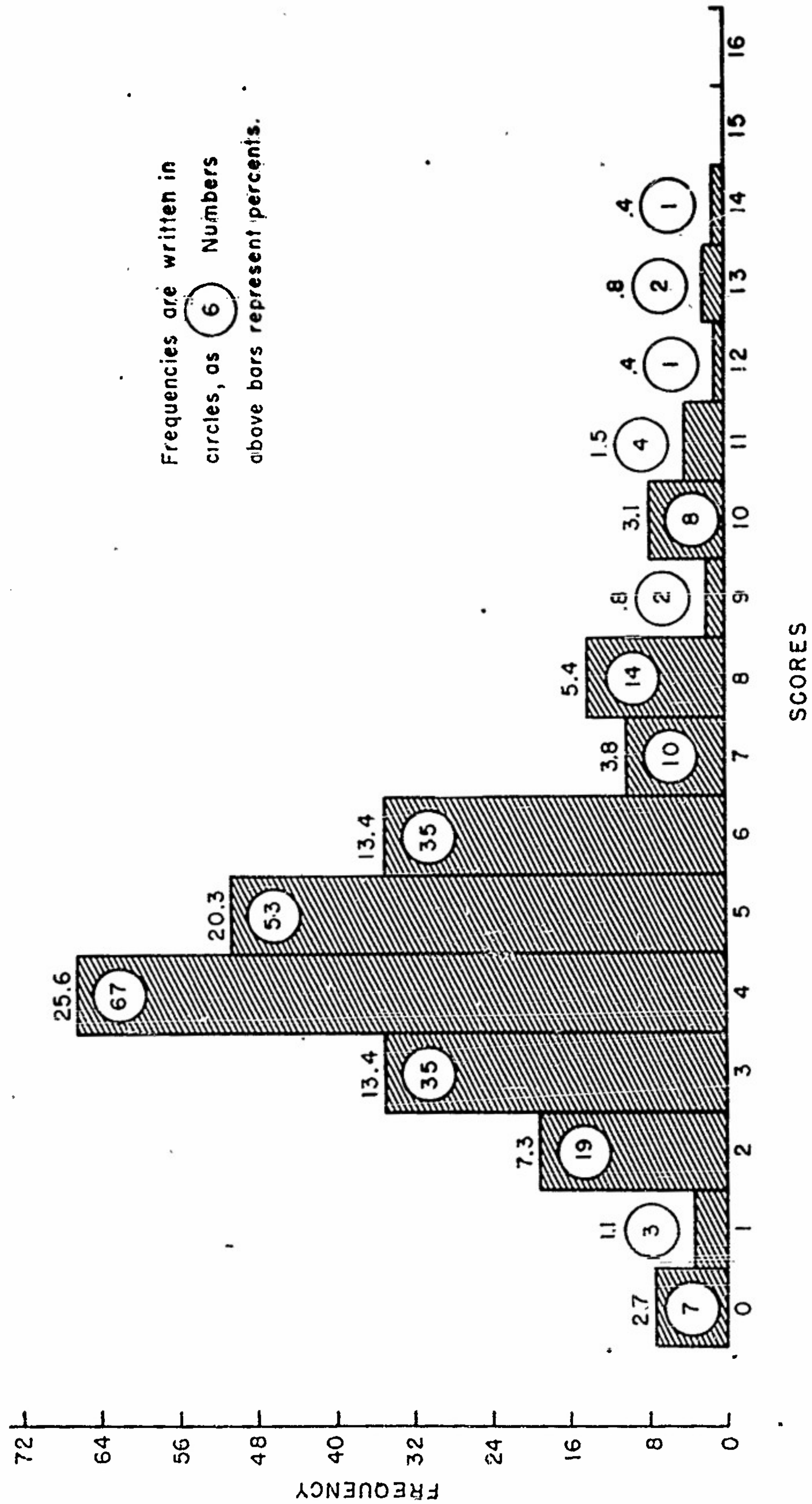
FIGURE 27.—*Frequency distribution for AAF Constant Decrement Test—Test 11.*

(12) *Test 12, Square Discrimination* (fig. 28). No test succeeded better than this one in spreading out the scores at the positive end of the scale. Whether this proves useful or not depends on validity determinations. At present the test is quite unreliable (about 0.44), and it may well be that in "cleaning up" the test the distribution picture may change. At present the test is not too difficult. It is, in fact, impossible to make the first item (perfect circle among the squares) any easier. So the 2.7 percent who failed to score are those who could not resolve at the basic sizes used for all stimuli.

(13) *Test 13, Checkerboard Variable Grid* (fig. 29). As previously discussed, this distribution has a minor mode at zero which will probably disappear when easier items are added to the test. Otherwise the distribution of scores is fairly normal.

(14) *Test 14, Vernier Acuity* (fig. 30). Negatively skewed with some indication that the most difficult items are not difficult enough, this distribution is similar to that for line resolution (fig. 25). It will be recalled that test-score variance was attributable to all the factors to some extent. Thus it is difficult to understand how individual scores were accomplished. In any case the distribution is at present adequate with the exceptions noted.

b. ITEM ANALYSIS. To further examine the action of the tests as measuring instruments per se, the performance of the individual items

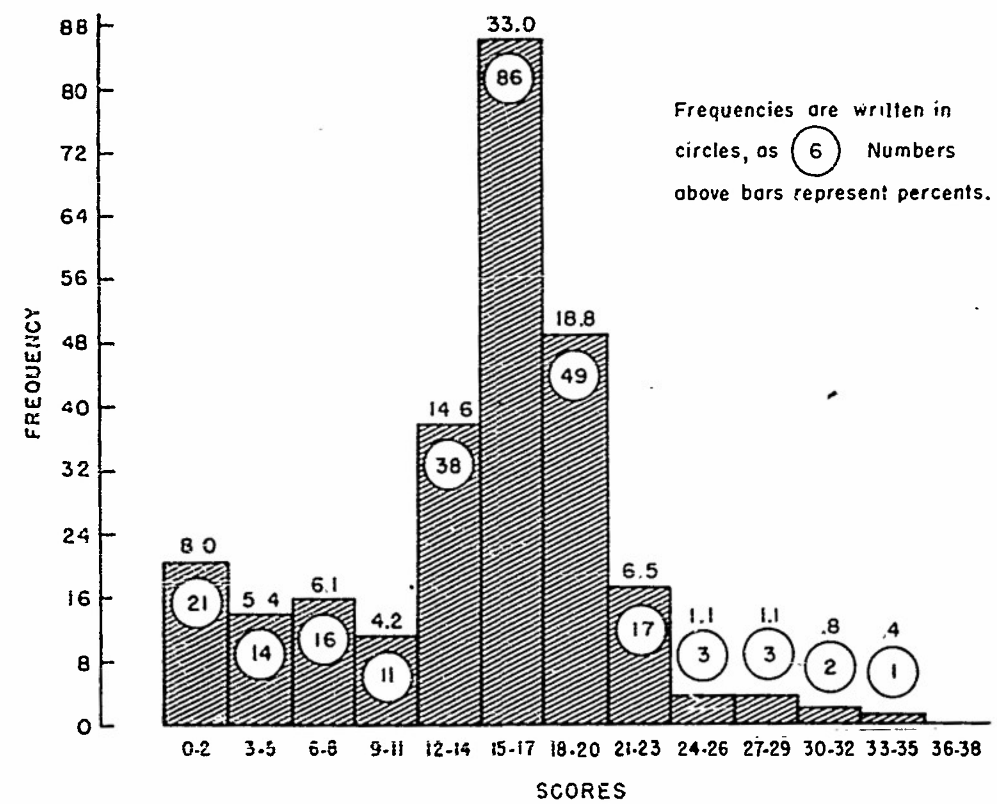


Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=4.8 σ=2.3).

FIGURE 28.—*Frequency distribution for Square Discrimination Test—Test 12.*

was investigated. As described at the beginning of this section of the results, this amounted to breaking the sample in thirds on the basis of total score and determining how well each item differentiated those in the high, middle, and low groups. Two methods of analyzing the action of the items were used, a computed index and graphic analysis. For purposes of this report the graphic method was considered most illustrative of the information.

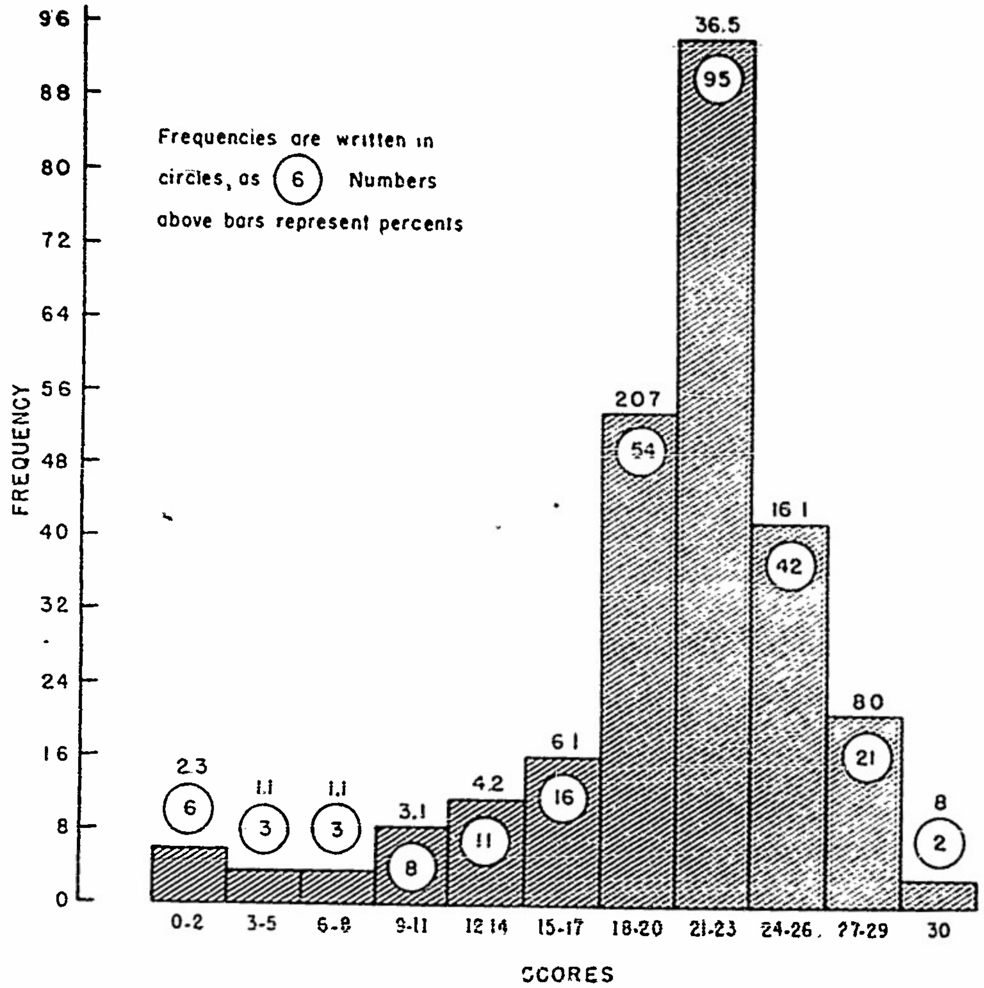
The graphs are presented in figures 31 to 44. To clarify the meaning of these graphic analyses the procedure is described as follows:



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=14.3 σ=6.4).

FIGURE 29.—*Frequency distribution for Checkerboard Variable Grid—Test 13.*

The sample is divided into 2 groups of 300 each, which in turn are split into high, middle, and low thirds of 100 cases each on the basis of total score on any given test. The results presented are for only 1 of the 300 men samples. Space limitations do not permit the inclusion of the second sample in this report. However, these data are available and they generally confirm the picture presented. For present purposes it is assumed that each of the thirds of the group of 300 represents a group of 100 individuals rather homogeneous as to the ability (high, middle, or low). Taking each of the groups in turn, the number of successes is plotted against the individual items arranged in order of presentation of the abscissa of each plot. Circles are used for plotting the high group data, crosses for the middle group, and triangles for the low group. It is apparent that the greater the proportion of individuals passing an item, the easier the item, and the less the proportion of individuals passing the more difficult the item. Since in all the present tests the items progressed presumably from



Based on first-test scores of 261 enlisted men at Fort Dix who were tested twice, September and October, 1946 (M=20.4 σ=5.5).

FIGURE 30.—*Frequency distribution for Vernier Acuity Test—Test 14.*

easy to difficult, the plots of numbers of successes against the items in serial order have the form of descending curves.

Theoretically, at least, an ogive form is expected in such a treatment. Taking a high ability group as an example, the 100 cases should all succeed on a long string of the initial, "easy" items of a test. Then, as the items which mark the limits of their ability are reached, a few individuals should err followed by an acceleration in individual errors and, finally, by a deceleration as the most difficult items are attempted. NOTE.—In the present testing procedure, individuals were stopped

after three consecutive misses. This means that an individual did not attempt items beyond the point where he stopped. It can be assumed that had they gone on, success would have been a matter of chance. We shall have more to say about this later. Meanwhile, the middle and low ability groups should behave in much the same way as the high group, but, naturally, the items over which the ogives occur should be found closer to the beginning of the test.

What can be learned from these ogives when they are plotted together (high, middle, and low groups) as they are in figures 31 to 44? It will be noted immediately that the points have not been connected in the same way for all tests. This is because the mode of item presentation varied. For tests in which stimulus value decreased with each succeeding item (item to item decrement) the points are continuously connected (tests 3, 5, 9, 10, and 11). For the remaining tests, in which stimulus value decreased by groups or lines of items (group to group decrement) only the points for a given stimulus value are connected. To make the analyses of the tests comparable, the average number of successes per line or group of items is indicated by a horizontal line across each group in those tests employing group to group decrement of the stimulus value. By tracing the horizontal lines for the high group (circles), middle (crosses), or low (triangles), the general form of any curve can be distinguished. Regardless of which method of connection is used it can be seen that the ogives are not always smooth. They differ considerably in slope and many of those for the high and low groups are incomplete. All these and other observations supply interesting information about the items. Each is discussed for the test in general and for each item specifically.

(1) *Discrimination of high, middle, and low ability.* How well do individual items contribute to the scaling of ability or separation of the three arbitrarily selected levels of ability? In the discussion of frequency distributions we have seen that in terms of total score the tests spread out individuals pretty well but with certain irregularities such as bimodality. Looking at the individual items it can be seen that, in general, they show orderly progressions in difficulty and effectiveness in scaling ability. However, in specific instances the item scales are quite inadequate. For example, examine test 1, the practice test (fig. 31). This test is a group-to-group decrement test, and the points for the items of like size are connected. Horizontal lines are drawn representing the mean number of successes for each group of items, except those too close to 100 or 0 successes to be clearly indicated. Taking each ability group in turn, the high group (circles) has nearly 100 percent success over each of the first six groups of items, then about 91, 81, and 33 of the group of 100 cases succeed on groups 7, 8, and 9 of the items. Within these last 12 items the present test differentiates the ability of the high group.

Notice that the last group of items is correctly called by a mean of 33 individuals. This means that addition of more difficult items is necessary to fully scale the high group ability. In the middle group (crosses) individuals begin to fail on easier items than in the high group and in about four stimulus values (groups 6, 7, 8, and 9) the number of successes goes from near 100 to near zero. The low groups, on the other hand, seem to have a much wider spread of ability as measured by this test. The slope of the ogive is much gentler, with three or four missing the first item group, and the number of successful individuals diminishes over some eight stimulus values of the test. Now reading up the columns above each stimulus value on the abscissa, notice how the item group means are separated in terms of the number of successes on the part of the high, middle, and low ability groups. The first five values fail to differentiate the high and middle groups. Hence, if only this part of the test were given to such populations, there would be no test at all. These items do discriminate a large portion of the low group, with only 54 out of 100 succeeding on the fifth group of items. Beginning with the sixth stimulus value the items separate the high and middle abilities, but it is over a very narrow range that this occurs (some 12 items). This indicates that the total scores of the high and middle groups will be concentrated at the high score end of the scale (negatively skewed) and that the scale is too coarse to make fine discriminations between individuals in these two groups. Meanwhile, the low group continues to be spread out over the 6, 7, and 8 stimulus values and the test may be said to measure this group most discriminatively.

To summarize, the high group is never fully tested by the items on the test, with some 33 out of 100 individuals still succeeding on the smallest stimulus values. Both the high and middle groups are measured over a small range of the most difficult items presented and with considerable overlap. Hence, the negative skewness found in the frequency distribution of total scores. The low group is discriminated more finely, which results in the long negative tail in the distribution.

(*a*) *Tests 1, 2, 9, and 14.* The results for test 2, Army Snellen (fig. 32), test 9, Line Resolution (fig. 39), and test 14, Vernier Acuity (fig. 44), are subject to similar interpretations as those given for test 1, Practice (fig. 1). Suggestions for improvement of tests 1, 2, 9, and 14 are—

1. Extend the scale at the difficult end by at least 3 additional stimulus values. This will eliminate the pile-up at perfect or near perfect scores. This applies to all 4 tests, but may prove difficult in Line Resolution (test 9) since the smallest line is close to reproduction limits.

2. Shift the Snellen (test 2) to a logarithmic scale of letter

sizes. Then add items to all four tests to get better discrimination of ability at the high score end. There is ample evidence that the addition of items by using smaller logarithmic decrements will improve the tests. The New London Letter (test 5) and AAF Letter (test 7), among others, use such scales and do a better job. Nevertheless, these two tests still show some evidence of negative skewness and, as we shall see, further reduction in the coarseness of the scales may be necessary. It should be remembered, however, that there are other possible causes of negative skewness, e. g., the multifactor nature of the tests. Another possibility is the order of presentation of the items, easy to difficult. The method used in all of the present tests is a modified method of serial exploration. This familiar technique as used in the laboratory, usually avoids two difficulties by suitable controls. First, ascending as well as descending series are employed to avoid the constant error of threshold estimation accompanying use of only one series or the other. By using only the descending series the present tests may overestimate ability; and, because of the nature of the present scales, this may lead to an overdose of high scores. Second, the presentation of a series "with knowledge" of what is coming leads to an influence of suggestion on the responses. In the present tests the examinees may anticipate difficulty on the ensuing items once they realize the decrement in scale values. The latter difficulty is not considered too important in the present cases, but the absence of ascending series in the present testing is an important consideration if any contemplated use of the test can conceivably be influenced by this error in measurement.

3. Use the same number of items per stimulus value whenever possible. The present Army Snellen should be revised on this basis wherever the size of the letters permit. In discussing the frequency distribution of this test (fig. 18) no mention was made of this point although it enters as an artifact explaining some part of the negative skewness in that test. Since the number of items per stimulus value increases by one for each successive value, those with high ability automatically accumulate more score. Even if a total score is not used and a line passage criterion is employed, the definition of chance performance or any percent correct in the single line will be variant as the number of items per line increases.

4. Add more stimulus values at the beginning of the test. In tests 1, 2, 9, and 14 the initial items are easy enough for almost all of the low ability groups to be successful. This is not true of all the other tests as we shall see. Although tests 1, 2, 9, and 14 accomplish this for the low ability group, there are reasons to believe that they are not good enough. Notice that this group has little opportunity to practice on a string of easy items, and this may account for the gentler slope of the ogives. In any case, the lack of practice makes the test a different proposition for this group, and one wonders what would happen if the middle and high groups began the tests with portions of their number failing the very first items. This also suggests that scaling of items on the same logarithmic continuum may not accomplish all that is desired. It may be necessary to choose the scale or scales for optimal use with divided ability groups. It is suggested that serious consideration be given to the use of a few "pilot" items, covering a wide range, to select the range of the scale in which any man's threshold occurs, to be followed by a more extensive group of items over the restricted range selected. This can conceivably be done by the use of only two or three of the selected ranges, e. g., one for low ability, and one for high ability.

(*b*) *Tests 3, 5, 7, 8, and 13.* The discriminative action of the items is similar for these five tests, namely, Dot Variable Size (fig. 33), New London Letter (fig. 35), Bausch and Lomb Checkerboard (fig. 37), AAF Letter (fig. 38), and Checkerboard Variable Grid (fig. 43). In each of these tests, sufficiently difficult values are used to fully measure the high ability groups (circles on the graphs). All three ability groups, high, middle, and low, are fairly well discriminated if the average success points per group of items of like stimulus value are considered. Individually the items do not do as well, especially in the New London Letter Test (fig. 35). This will be discussed in detail below. Notice that the ranges of items over which the high and middle groups of all four tests are discriminated cover more stimulus values than those found in tests 1, 2, 9, and 14 (figs. 31, 32, 39, and 44). The middle group (crosses) in figure 38 is the only exception with close to 100 successes on the fourth stimulus value, a mean of 80 on the fifth, 14 to 15 on the seventh, and almost no successes on the eighth. Over these three stimulus values the group passes from complete success to failure. Although the items on the five tests succeed in spreading out high ability, there is still room for improvement especially in the two letter tests (5 and 8). Negative skewness is apparent in all tests to a slight degree and may be a function of the coarseness of the scale or other causes mentioned in the discussion of tests 1, 2, 9, and 14.

Of much greater concern are the results for the low group on tests 3, 5, 7, 8, and 13. The ogives for this ability group are incomplete.

On the first stimulus value of Dot Variable Size (fig. 33) an average of 37 men out of the 100 man group commit errors. Similar results are obtained for tests 5, 7, 8, and 13: a mean of about 43 out of 100 fails the first item of test 5; 27 fail the first value of test 7; 49 fail in the case of test 8; and 32 out of 100 fail on test 13. Hence, ability of the low group is measured incompletely and this accounts for the minor mode of zero score observed in the frequency distributions for these tests. At the same time, the items do an excellent job of separating the low ability group from the middle and high on these tests. For screening purposes they are adequate; for purposes of scaling low ability, larger items are needed. Suggestions for improvement of tests 3, 5, 7, 8, and 13 are:

1. Add less difficult items than the easiest ones in the present tests. As mentioned in connection with tests 1, 2, 9, and 14, this may be accomplished in two ways. Extending the items at the easy end of the scale may effect a normal distribution. On the other hand, the same logarithmic scale of items may not apply equally well to different ability groups, in which case more than one scale of items may prove fruitful.
2. Increase the number of stimulus values, hence the fineness of the scale, to eliminate the negative skewness and spread out the higher ability groups, especially in the Letter and Dot Variable Size tests (figs. 33, 35, and 38).

(*c*) *Test 11.* The ogives for the AAF Constant Decrement Test (fig. 41) are like those for the Practice, Snellen, and Vernier tests in the sense that they reveal failure to discriminate at the high and middle ability levels. This is due to a coarseness in the scale below the 20/20 item. This test is unlike the Practice, Snellen, and Vernier tests since it starts at a much more difficult level (20/40) and 27 of the 100-man low ability group fail the first item. In this respect the test is similar to tests 3, 5, 7, 8, and 13. This is not surprising since all of these tests began with stimulus values physically equivalent to the 20/40 item of the present test. The net result is a test which discriminates poorly at the high ability end (negatively skewed) and has a minor mode at zero.

The irregularity of the curves for this test is a function of the decrement used, namely, item to item. Individual items do not retain the same order on the difficulty continuum as that expected in terms of physical size. This is also found on group-to-group decrement tests, but averaging the results of items at given stimulus values tends to smooth the curves for these tests. The variable action of individual items will be discussed more extensively below.

(*d*) *Test 4 and 10.* The Quadrant and Dot Variable Contrast tests (figs. 34 and 40) are constant decrement tests. Both have low reliability in their present forms, but are quite different with respect to item difficulty. In the quadrant test (4), the items discriminate ability over the entire test, but in the dot contrast test (10) most of the discrimination takes place over the first 6 items. The result is a fairly normal distribution in the case of the quadrant test and a positively skewed one for the dot test. In both tests there are insufficient items at either end of the scale to measure the highest and lowest abilities, and the individual items in both behave quite erratically. However, the quadrant test (4) stands out as most useful in its present form. The larger area of the "spot" in the quadrant contrast test is credited for its superior performance as a test. Item difficulty in the dot contrast test is probably a function of using an area too small for reliable measurement purposes. In addition, it is conceivable that orientation of the responses in the diamond frame of reference used in the dot test confuses the difficulty picture, and that discrimination of the diamonds becomes a serious limitation on performance.

(e) *Tests 6 and 12.* The items of Triangle and Square Discrimination (figs. 36 and 42) discriminate fairly well over the entire range, but the square test (12) is seen to be the more difficult of the two. In the square test the high and middle groups are poorly discriminated over the first few items while successes diminish rapidly in the case of both groups. In the triangle test (6) both high and middle groups are poorly discriminated over early items, but successes do not diminish at these points. In both tests initial and final items are, respectively, too difficult and too easy to measure the complete range of abilities. This is especially true of the final items of the triangle tests (6) in which 58 out of 100 and 8 out of 100 cases in the high and middle groups, respectively, succeed on the final item. Once again, the irregularities apparent in the action of individual items of these tests are important in considering revision of the tests and will be discussed below. Suggestions for improvement of tests 4, 6, 10, 11, and 12 are:

1. Add easier items to the AAF Constant Decrement Test (11).

2. If a normal distribution is desired on the AAF Constant Decrement Test (11) shift the size value to logarithmic ones as in the case of AAF Letter (8) and New London Letter (5).

3. Since the present initial items of the Triangle (6) and Square (12) Discrimination tests are as easy as they can be made in the present form, addition of easier items to these tests is a matter of attempting to increase their size and improve their spacing to a point where the initial item can be correctly judged. Spacing and item size should be corrected throughout these tests.

4. Increase the number of scale values at the difficult end of Triangle (6) and Square (12) Discrimination tests, more so in the case of the Triangle. Interpolate additional items at scale values where present discrimination is poor.

5. Increase the area of the "spots" in the Dot Variable Contrast Test (10).

6. Make both contrast responses as free as possible of spatial references if a purer test of brightness discrimination is desired.

7. Increase the number of scale values by addition at both ends of the contrast scales (4 and 10) and interpolation within the scales.

(2) *Difficulty of individual items.* In discussing the discrimination of ability by the items, in general, brief mention was made of the varying behavior of individual items. However, several interesting bits of information may be garnered from such an analysis and may prove quite important in future test designs.

In considering the action of individual items the first point becomes a matter of what one expects to occur in terms of the form of the items and the modes of presentation and response. Two forms of items are presented in the tests: Letters to which the responses are the names of said letters, and objects placed in two-dimensional space to which the responses are in terms of four positions—top, bottom, right, or left. As to presentation, the two forms of the items are used in tests with item-to-item decrements and others with group-to-group decrements. Both item forms are used in tests with one chart and in those broken up into two charts.

In the letter tests an absolute judgment is made about the patterns presented no matter what form of presentation is used. In this, the possibility of success by guessing alone is rather slim since the examinee has 26 letters from which to choose. This number is somewhat reduced if the examinee realizes that letters are repeated, but in the present letter tests this is likely to occur only in the AAF Letter Test in which the six letters are used from line to line. In general, it is fair to say that the chance letter response is indeterminate. Nevertheless, as the threshold is reached on the letter charts, one expects to find certain letters called correctly more often than others on the basis of consistently greater ease of certain forms.

In the nonletter tests an absolute judgment is also used, but in these cases the role of chance is determinate since one of four possible choices can be called for each item—top, bottom, right, or left. Each examinee by sheer guessing has one chance in four of success, on any item. In other words, one out of every four examinees is expected to succeed on the basis of chance; and at any line in the test, depending on the difficulty at the items for the group attempting them, about the same number of successes are to be expected on all four items of like size. Of course, this assumes that variables other than the specified sizes of the items have little or no influence on the response. Such variables are: the relative position of an item in lines of any chart;

variables introduced by the four different responses—top, bottom, right, and left; the shift from line to line of any chart; and the shift from chart to chart of any test. For nonletter items with the same specifications, some variability in the number of successes about the mean is to be expected on the basis of group fluctuations in ability. Any systematic departures from chance expectancy related to variables such as those cited above, may indicate serious flaws in the test or tests in which they are found. To prepare for an examination of the data with respect to such departures, chance expectancy should be carefully identified.

It has been seen that one out of four individuals will ordinarily succeed on a four-choice item by guessing. In other words, 25 percent of the individuals who are guessing on any line of items are expected to succeed on any item by chance alone. This means that, in addition to those who can actually see any given group of items of like size, 25 percent of those who are reduced to guessing will also be successful. But this assumes that the same number of individuals tries all the items of a given size. This may not be true in the present tests because an artifact arises in the method used to stop the tests. Tests are stopped after three consecutive misses, no matter where in the line, these occur. Theoretically, four possible modes exist in which three consecutive errors on any two lines of four items can occur.

As an example of this artifact, consider a group of N people for whom a given line is supraliminal and for whom the next line is subliminal. In such a case, where both lines "straddle" the limen, the pattern of responses for those people in the group who end the test on the second line will usually be—

	Item position 1	2	3	4
line *x*	S	S	S	S
line *y*	E	E	E	O

Where S is a correct response, E an error, and O indicates that no response was allowed.

In a four-choice item with one correct choice the probability of an error on any item, purely on a chance basis, is ¾. If line *y* is subliminal for this particular group, and they are guessing in the sense of responding on a chance basis, the probability of the above pattern of responses is $(\frac{3}{4})^3$ or .42. From this we may conclude that .42N will not respond to item 4, line *y*. Of this number ¼ would have been successful if allowed to continue. Therefore ¼ of .42N or about 10 percent of N is the amount by which the number of successes on item 4, line *y*, will be reduced. Since we ordinarily expect 25 percent of this guessing group to be successful on each item of line *y* it follows that the expected number of successes on items 1, 2, and 3 are 25 percent

of N, but on item 4, because the test was stopped for 42 percent of N, it will be only about 15 percent of N.

The special features of the group of N people discussed above should be emphasized:

Their limens were in the *gap* between line *x* and *y*.

The probability of making an error on line *x* is assumed to be negligible.

For this group, then, only the last or fourth item in the line was affected by the test stopping method and it would thus *appear* to be more difficult than the other three, where number of successes is the difficulty criterion.

When the test is stopped for any individual at any point we know that the pattern of the last four responses must be S E E E. Now of the 52 percent of N individuals in the above group who continue, the proportion that will drop out on item 4, line *y*, and who will not respond to the first item of the *next* line (say line *z*) is given by the probability of this pattern which is ¼ $(\frac{3}{4})^3$ or about 10 percent. Of these, 25 percent would have made an "S" response on item 1, line *z*. Therefore, the number of successes on this item will be (0.25) (0.10) (0.52) N or about 1.3 percent of N *less* than for item 4, line *y*. We can move the pattern S E E E forward in this manner one item at a time. In doing so it can be shown that the number of successes on each item is successively reduced.

The same type of argument would apply to a situation where two adjacent lines were supraliminal and liminal respectively if we defined the probability for an error at a limen in the usual way as ½.

In general we would expect, in a line where the probability for error was appreciable and constant and which immediately followed a supraliminal line, that the fourth item would act as indicated at the beginning. Thereafter, for lines following this line the decrement of successes would continue for each *successive item*. Since the combination of all these factors are additive the effects may be at times quite noticeable and this should be remembered in examining the results.

In summary, the effects of all the above will be to (1) decrease the number of successes of the fourth item in the neighborhood of a limen; (2) decrease the number of successes for each successive item after the point defined in (1); (3) increase the ordinate distances in terms of number of successes between lines after the point defined in (1).

In examining the results, letter and nonletter tests are treated separately.

(*a*) *Letter tests.*

1. Test 2. Army Snellen (fig. 32). Considerable variability in the number of successes on letters of like size is apparent in this test. The letters in any line are not always of equal difficulty. For the low ability groups, shown in the figure by triangles, the fluctuations noted are not too severe and may be considered a result of chance variation in the sample tested. For example, on line four of the items, 65 of those in the low ability group, who are still attempting the items, succeed on the first item; 81 on the second; 62 on the third; and 72 succeed on the fourth item in this line. The letters of the line are P, L, O, and E. However, when results on these same letters are examined for other lines of the test, a somewhat consistent picture of unequal difficulty is obtained. Taking the letter P in lines 4, 6, and 7 of the test, the number of successes by the low group in this letter is always lower than the mean for the line. The L's, on the other hand, seem to be consistently easier. In lines 3, 4, 5, 6, and 7 successes on the L's are always greater than the mean low ability group successes on the lines in which the L's appear.

Turning to the high and middle ability plots, the letters are found to be even more variable, and consistency of item difficulty from the high to middle groups is also observed. The last letter in line 7, P, is severely lower than the mean success on the line for both the high and middle groups. In fact, the letter was more difficult for the high group than most of the letters in the same line were for the middle ability group. Compared with a mean success of over 83 individuals in the middle ability group on line 7, only 74 of the high group successfully called the P, at the end of the line. In line 8 it is evident that the difficulty of the letters varies considerably. Of course, the extent of this appears greater in the plots for this line partly because this is a point in the test where greater fluctuation is expected, being at threshold or near threshold value for the individuals of the high and middle ability groups. However, in both ability groups the two L's and O are consistently easier while the two C's and D are consistently more difficult in comparison to mean success. In the high group plot for line 9 the second letter E is typically high in number of successes, along with the L, while Z, C, and T of the line are most difficult.

To summarize, there is a rather consistent finding that certain letters of the Army Snellen Test are easier while others are more difficult than the mean difficulty for the lines in which they are *placed*. As such they do not destroy discrimination between groups of different ability since the letters act rather consistently for these groups. However, from the point of view of measurement this

variability should be reduced by further standardizing the items.

2. *Test 5. New London Letter* (fig. 35). About the same amount of variability is found in number of success by the three ability groups on the lines of this test as that found on the Army Snellen. The first letter of the test, G, is quite atypical of the action of the other four letters in line one of the chart. In the high ability group only 78 out of 100 succeed on the letter compared with perfect group performance on the remaining 4 letters. In the middle ability group, 69 out of 100 succeed on the G while the other 4 letters are always correctly called. And in the low group the 31 successes on this letter are far below mean success of over 56 on the line. Nor does this mark the end of trouble with this letter. It is consistently more difficult than the other letters of like size, for all three ability groups, in lines 3 and 4. The results on later lines in which the G appears (lines 6 and 9) do not reveal that it is so atypical although in line 6 it is next to the most difficult letter of the line for the high ability group. Other letters of this test which rather consistently vary in difficulty from means for letters of the same size are the letters O, E, and V (easier); and the letters H and Y (more difficult). These results, naturally, are seen best at points in the test where the thresholds are being reached (lines 5, 6, 7, and 8 for the high and middle groups). As in the Snellen, considerable work is necessary to better equate the difficulty of items of like size if this test is to be an accurate instrument.
3. *Test 8. AAF Letter Test* (fig. 38). In this test the fact that the same six letters (K, M, G, S, Y, P) are used, in varying order, in each line makes analysis of the individual items more complete. Here, the same letter can be traced through every size value employed in the test. The first impression is that variability of difficulty among items of the same size value is less in this test than in the other letter tests. Letters which act in ways systematically different from others are quite rare. The letter P seems to fall above the mean success points on lines 6, 7, 8, and 9 for the high ability group; lines 5 and 6 for the middle ability group; and is quite inconsistent for the low ability group. It may be less difficult, but the evidence is inconclusive. The letter S is below mean success on lines 6, 7, 8, and 9 of the high group; but it does not act as consistently for the middle and low ability groups. In general, it can be said that improvement in the letter difficulty picture

is going to be arduous in the case of the AAF Letter Test, except for improvement in the letter P.

4. *Test 11. AAF Constant Decrement* (fig. 41). Some irregularity is noted in the steady progression of difficulty which should be obtained on the items of this test. Each successive letter is smaller than the last. In the plot for the low group (triangles) a cyclic pattern is seen with the letter V appearing to be easier than certain other letters. If any letters are consistently more difficult, they are the C's and Z's, but this tendency is not marked in the low group. For the high group V again appears easier at the twenty-eighth letter position, and it does in the same position for the middle ability group. Letters C and Z are rather consistently low in the high and middle ability groups.

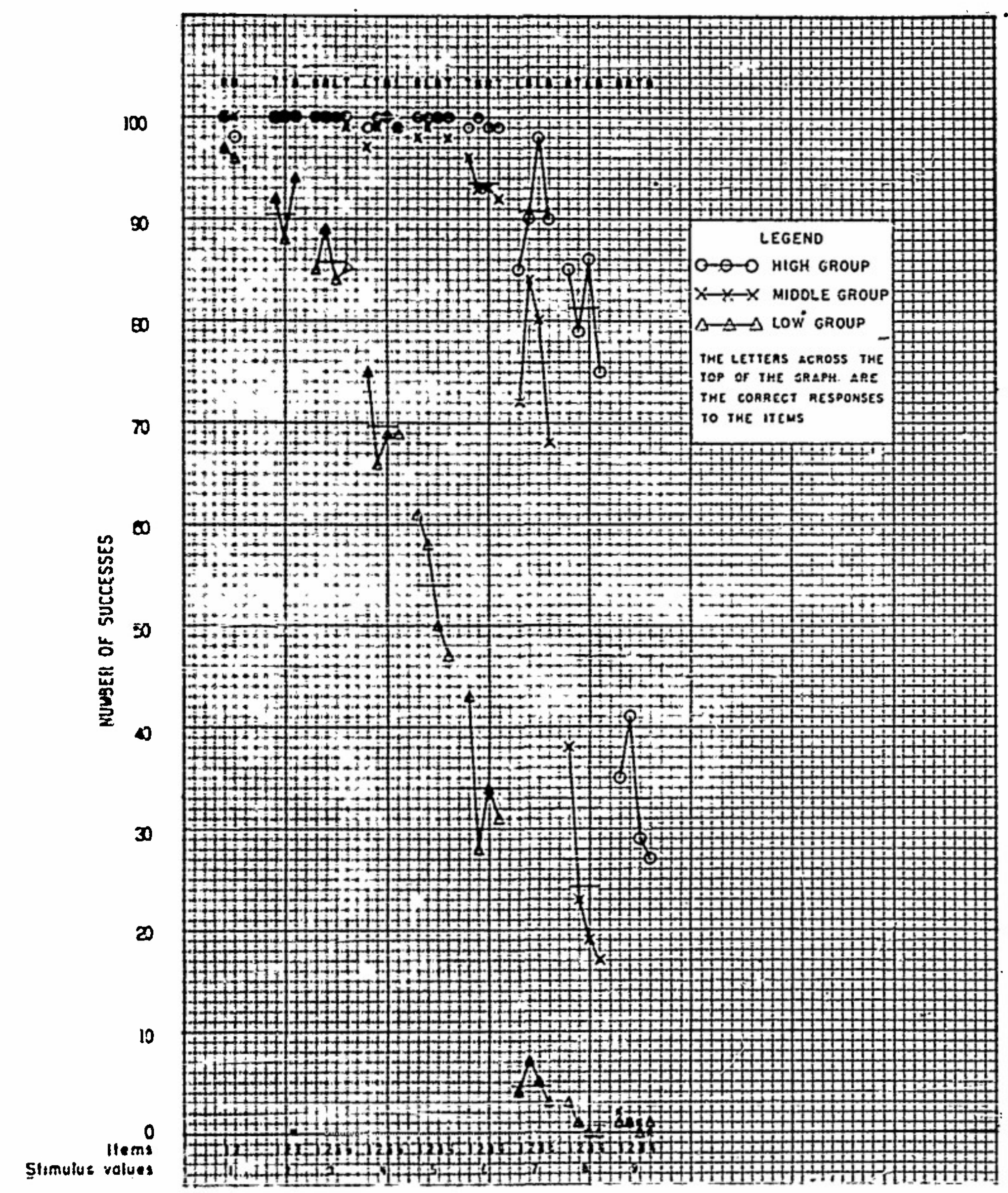
In this constant decrement test another point of interest is the break occurring between lines. From the present data there does not seem to be any systematic fluctuation in item difficulty as a result of shifting from line to line. The arrows on the abscissa indicate the points of shift, and the first item of the line following the break seems to be of lesser or higher difficulty on about a chance basis for all ability groups.

As corrective measures in this test, only modification of the difficulty of the letters V, C, and Z seems to be indicated by the present data.

(*f*) *Nonletter tests.* In the nonletter tests, interest in the individual items lies in the relative difficulty of items of the same size (four in a line or group); in any differential difficulty of top, bottom, left, or right modes of response; and in the break between charts when two charts are used to display all of the items of certain tests. The tests will be discussed in numerical order.

1. *Test 1. Practice* (fig. 31). This test, which employs a square modification of the Landolt ring, is contained in one chart and consists of groups of one to four items of like size. Four items are presented at each size category from size 3 through the remaining ones, but, for the first two sizes, two and three items are presented. Examination of figure 31 reveals that the items per groups of the same size are never as variable about the mean success points (horizontal lines) as was observed in the letter tests (2, 5, and 8). Certain groups of items, namely, group 5 for the low ability group and group 8 for the middle one, indicate that difficulty within the same size of items decreases progressively over the positions of the items within any given group. This is not conclusively demonstrated by casual perusal of

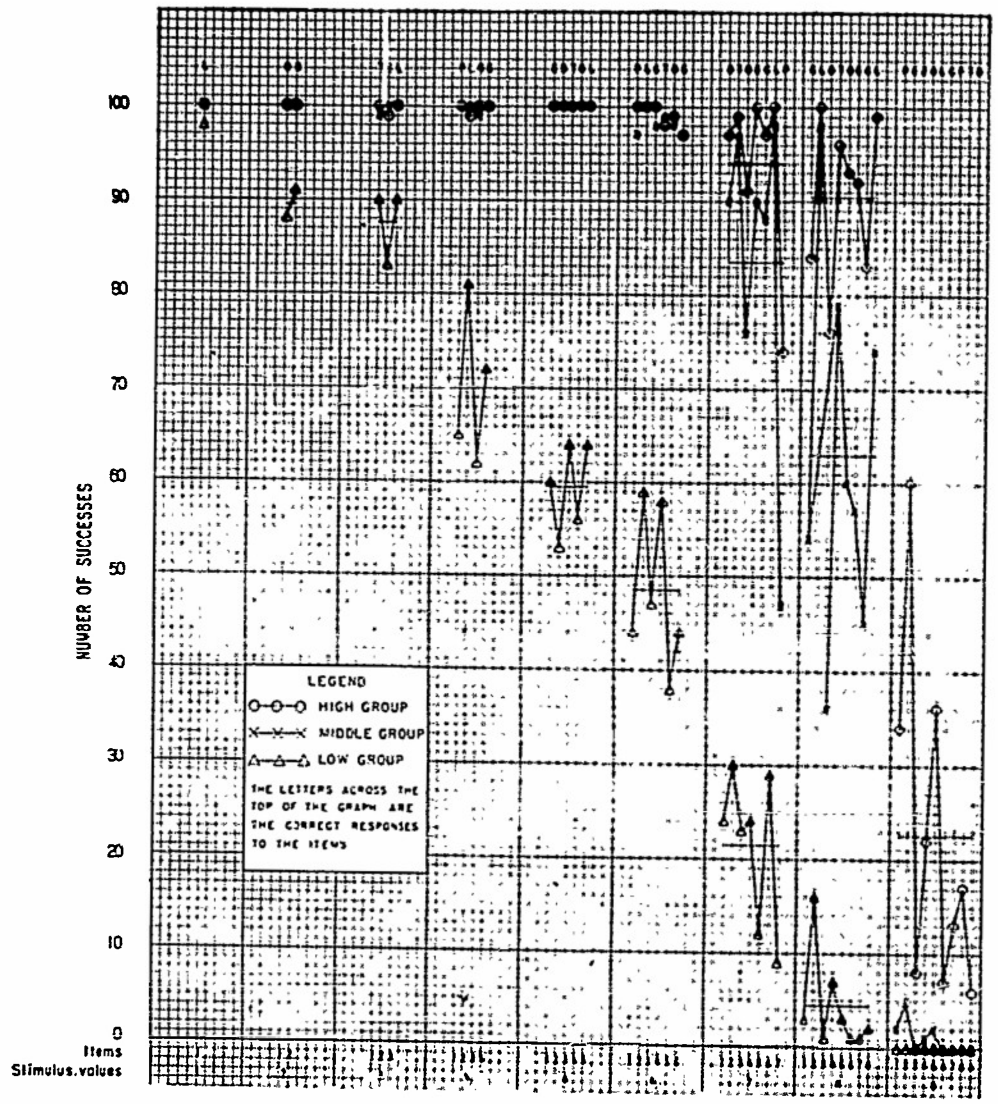
the plots although, in more cases than not, the first item in the groups seems to be less difficult than the last. Remembering the artifact introduced by the method of stopping the test, it is to be expected that somewhat fewer successes will be found on the last item than on the first. However,



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 31.—*Practice Test—Test 1.*

in the present test, the hypothesis that no variation in successes occurs as between positions of items in groups of like size except that due to chance random sampling has not been tested. Present status of the analysis makes this unavailable. Nevertheless, as will be seen, this impression is augmented as more test results are examined and may prove to be interesting from the point of view of test methodology.



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

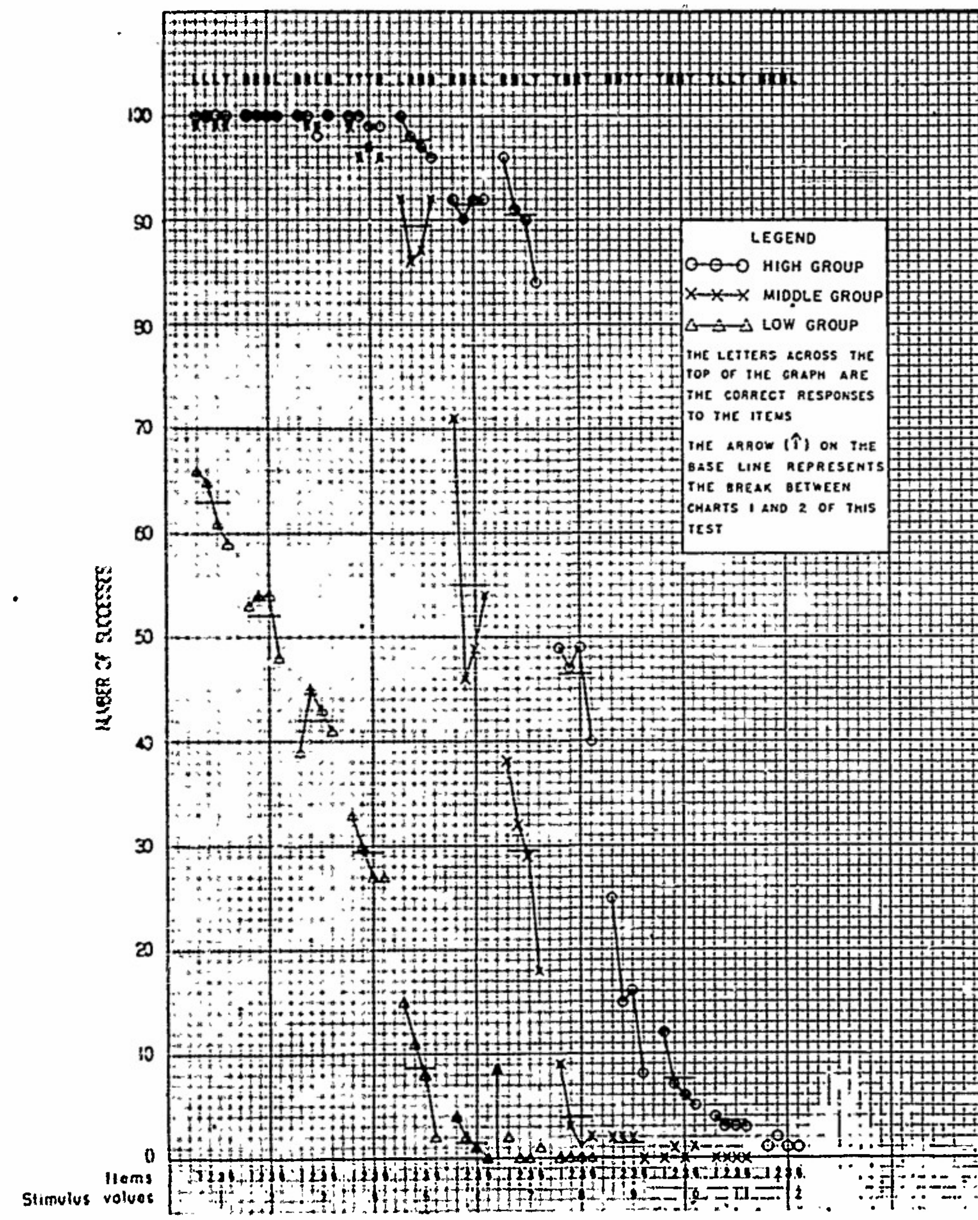
FIGURE 32.—*Army Snellen Test—Test 2.*

As to the top, bottom, right, or left categories of response, nothing conclusive can be said about any systematic difficulty attached to one or the other on the basis of current results on this test.

2. *Test 3. Dot Variable Size* (fig. 33). Once again, on this dot test less variability is noted among the items of like size than that observed in the letter tests. Beyond this, the same tendency is noted in this test as in the Practice test (1)

for last items in the line to be more difficult than the first. Again, the deviation of this result from chance expectancy cannot be shown readily from the present results.

In passing, note that lines 1 and 5 for the low groups,



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 33.—*Dot Variable Size Test—Test 3.*

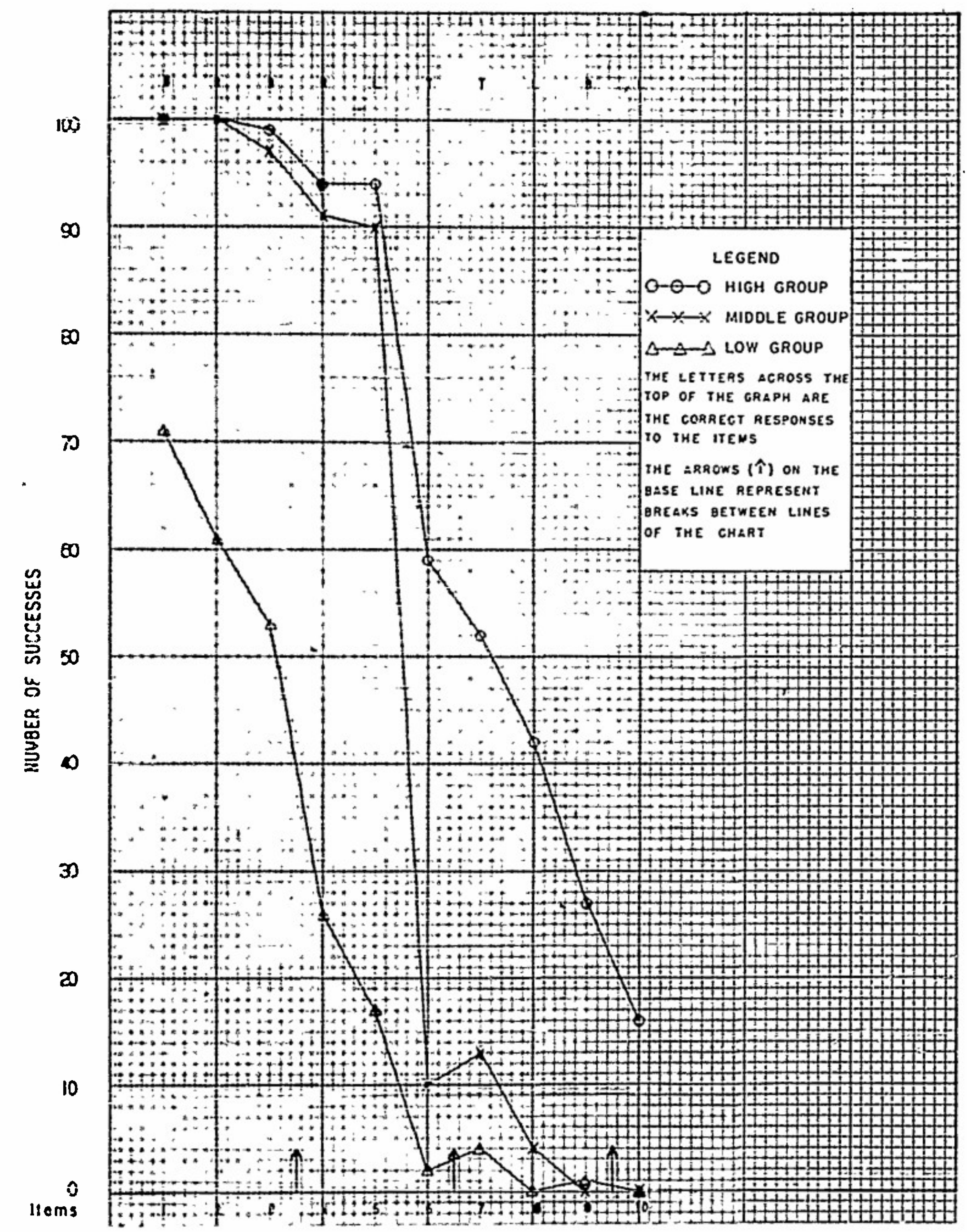
line 7 for the middle group, and lines 7, 9, and 10 for the high groups show consistently fewer successes from the first to the last items in lines of equal stimulus size. The result of using two charts is of considerable interest in this test. The arrow on the base line represents the point at which the second chart is introduced. Notice that the low group has reached a point of complete failure before the second chart is introduced. In the case of the high group, however, the introduction of the second chart is contiguous with the onset of many thresholds. It is interesting to find that the mean success on line 7 (the first line of the second chart) is practically the same as that for line 6 (last of the lines on chart 1). Whether this irregularity is due to the new constellation of stimuli presented in chart 2 or something specific to the first line of a chart is not clear.

In any case the occurrence of this atypical result may indicate that the use of two test charts to present the same test item continuum leads to error in testing. It may be better practice to seek a new method of presentation which avoids this deficiency, such as presentation of item after item or line after line by machine projection.

3. *Test 4. Quadrant Variable Contrast* (fig. 34). In this constant decrement test a continuous decrement of success is expected after each ability group reaches their threshold. Indications from the results (fig. 34) are that two serious irregularities occur in this test. First, item 5 appears to be as easy as item 4 (approximately the same number of successes). This indicates poor construction of the test since the observed deviation from expected decrement in successes does not appear to be related to any break between lines (indicated by the arrows on the base line). The second irregularity is a suggestion that the line break between items 6 and 7 results in a rise in successes on item 7, however small, as evidenced in the plots for the middle and low ability groups.

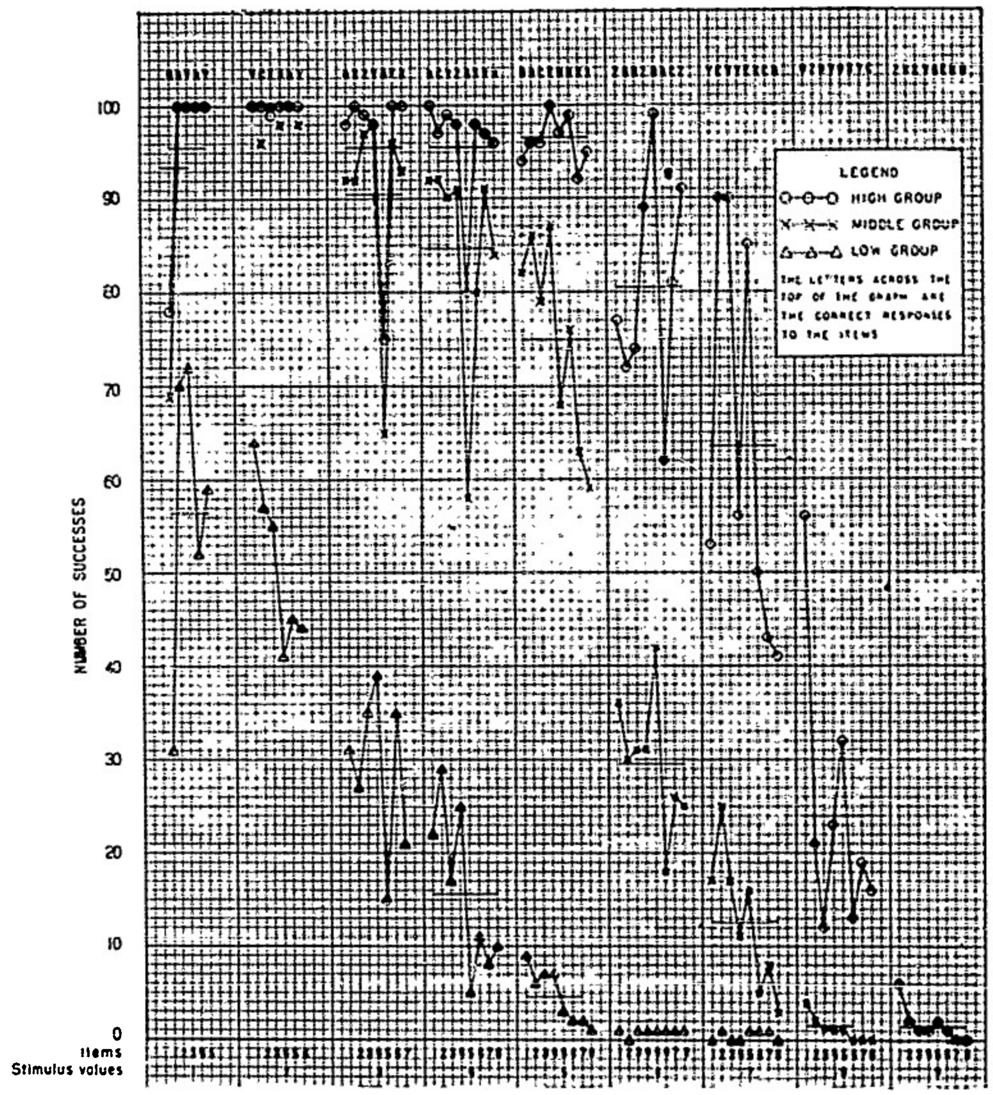
4. *Test 6. Triangle Discrimination* (fig. 36). This triangle test is another constant decrement test with four possible choices on each item. The outstanding feature of the results on this test is the consistent rise in number of successes achieved by all three ability groups as they passed from the first to the second chart. On item 10, the last item of chart 1, 79, 65, and 5 cases of the high, middle, and low groups respectively, succeeded. Successes on item 11, the first one on chart 2, amounted to 94, 90, and 23 men of the high, middle, and low groups. Clearly, this indicates that either the first item or line of a test chart is easier than one would expect from the item sizes or that the presentation of a new chart introduced a new approach on the part of examinees.

5. *Test 7. Bausch and Lomb Checkerboard* (fig. 37). The items of this checkerboard test behave in much the same



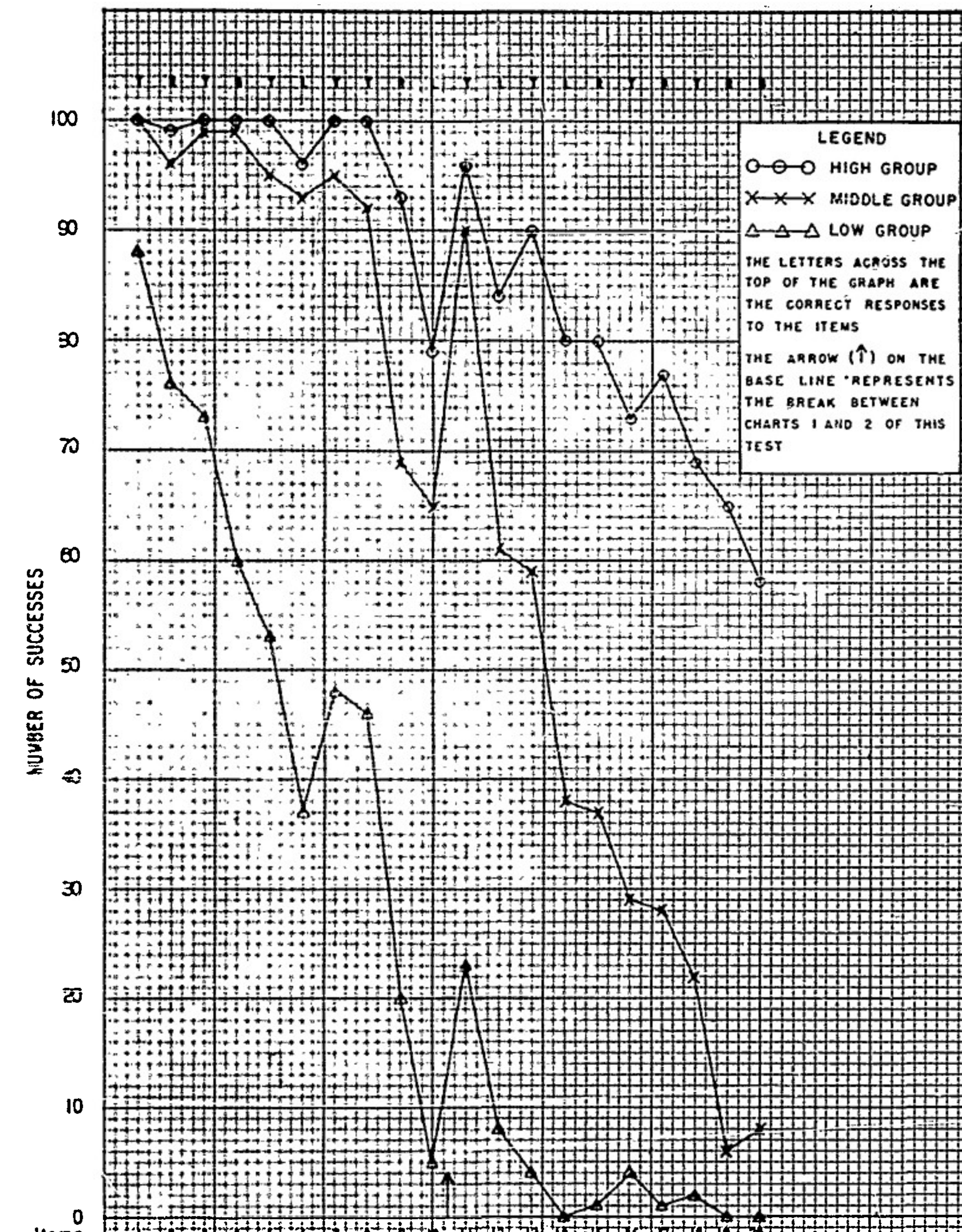
The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 34.—*Quadrant Variable Contrast Test—Test 4.*



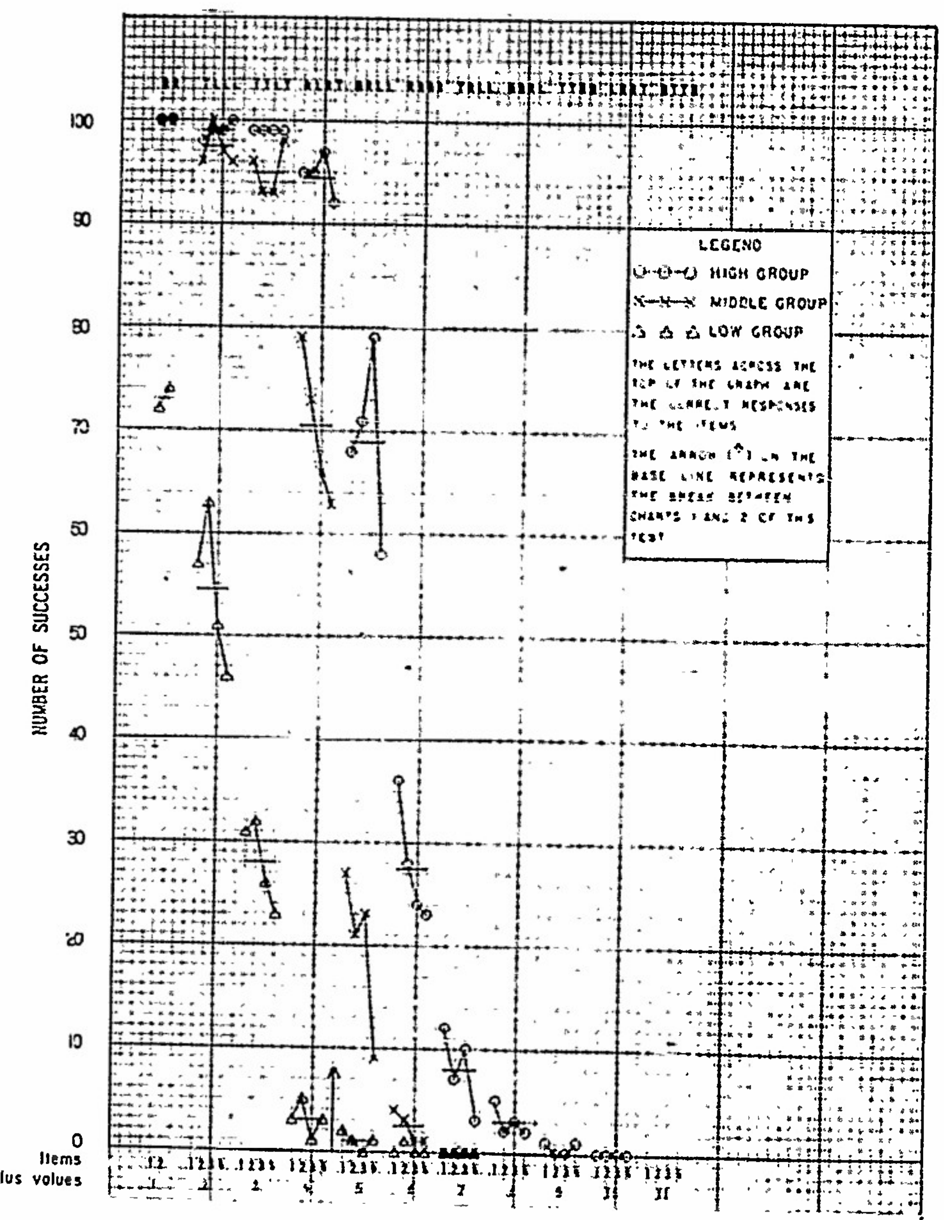
The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 35.—*New London Letter Test—Test 5.*



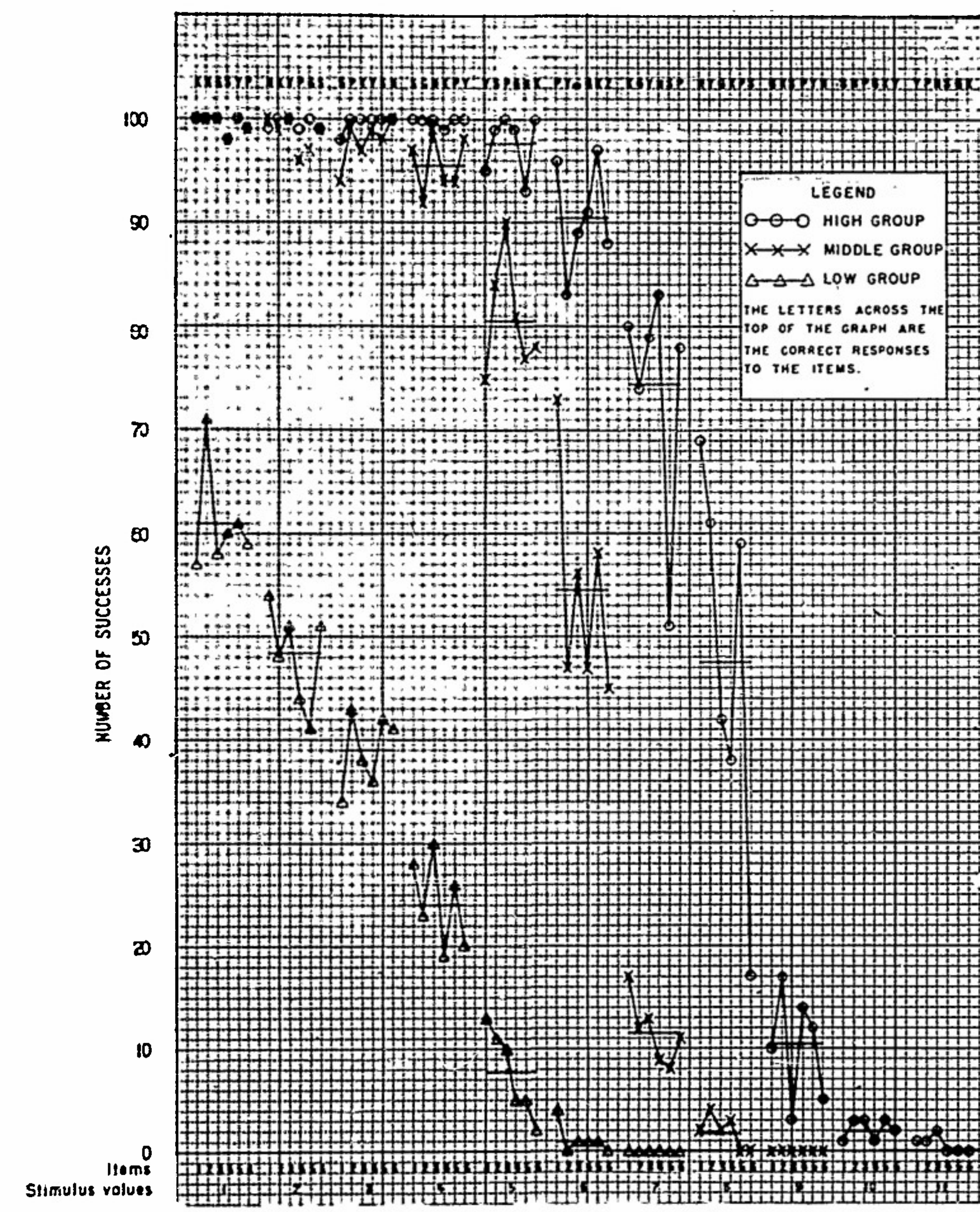
The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 36.—*Triangle Discrimination Test—Test 6.*



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 37.—*Bausch and Lomb Checkerboard Test—Test 7.*



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 38.—*AAF Letter Test—Test 8.*

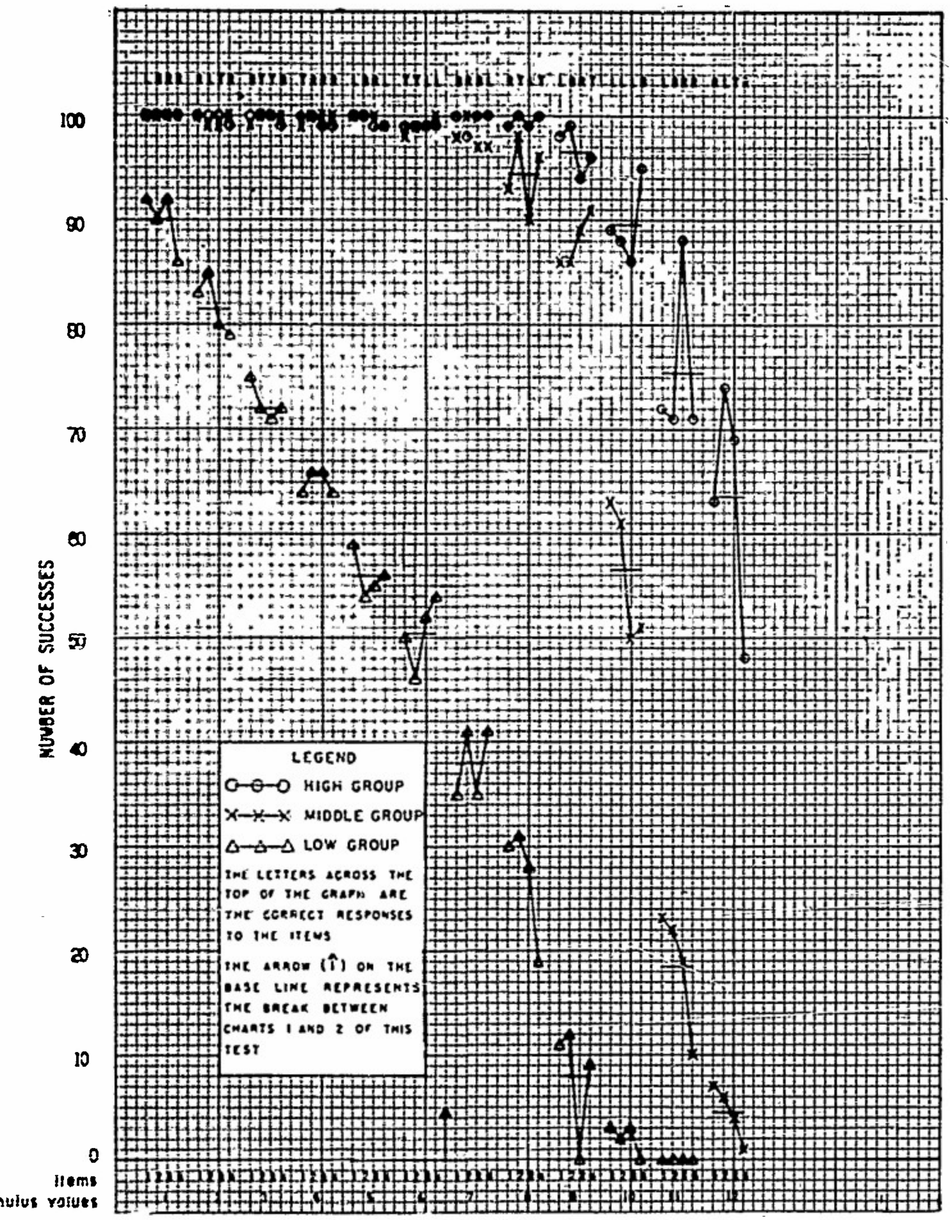
way as do those on the Dot Variable size test (fig. 33). Again there is some indication of a progression of difficulty from the first to last positions of the items in the lines. The break between charts has no observable effect in this test.

6. *Test 9. Line Resolution* (fig. 39). The results of this test show nothing that has not been discussed in connection with the dot and checkerboard tests. Again, items of the same size are seen to be less variable in difficulty than the letters.

7. *Test 10. Dot Variable Contrast* (fig. 40). The outstanding result in this test is the jump in successes on the part of the high ability group on item 7. This may be an item with a contrast value different from the specification given for it, or the rise in successes may be a function of the shift from line to line between items 6 and 7. Finally, the rise in successes may be a function of chance random sampling.

8. *Test 12. Square Discrimination* (fig. 42). The items of this test consist of six "squares" which take one line of the test per item. The second, third, fourth, and fifth positions in each line are the places where the curved "squares" may occur. To avoid any variation in difficulty that might arise because the correct response (square with curved sides) falls at the beginning or end position of a line, a true square is placed at both ends of each item. This means that each of the possible choices is completely surrounded by objects except those in the first and last items of the two charts. These items are not bordered by squares above the first item or below the last item of either chart. This may contribute to the irregularity in difficulty which appears for item 8 in this test, item 8 being the last item on the first chart. In both the high and middle ability groups a rather sharp increase in number of successes is found on item 8. Other than this, little can be said about the present items aside from the contention that in their present form they produce all kinds of configurational effects which, if possible, should be eliminated.

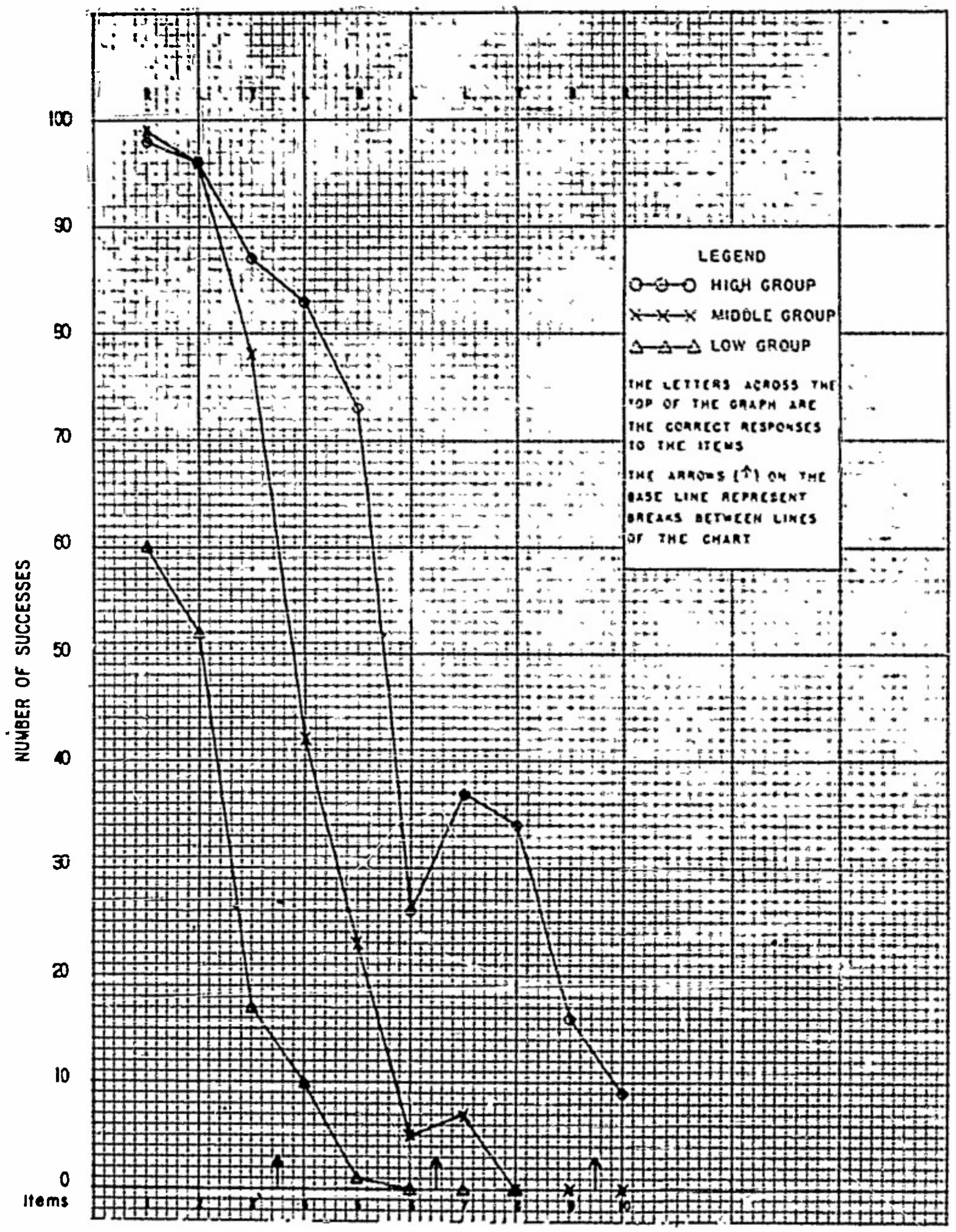
9. *Test 13. Checkerboard Variable Grid* (fig. 43). The items of this test show the same relative stability observed in other nonletter tests, but again, there is some indication of systematic variation in difficulty dependent on position of the items in any line. As mentioned, part of this can be explained in terms of the artifact which makes the chance expectancy of success on the last item in any line lower than that on the first three. To establish the num-



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

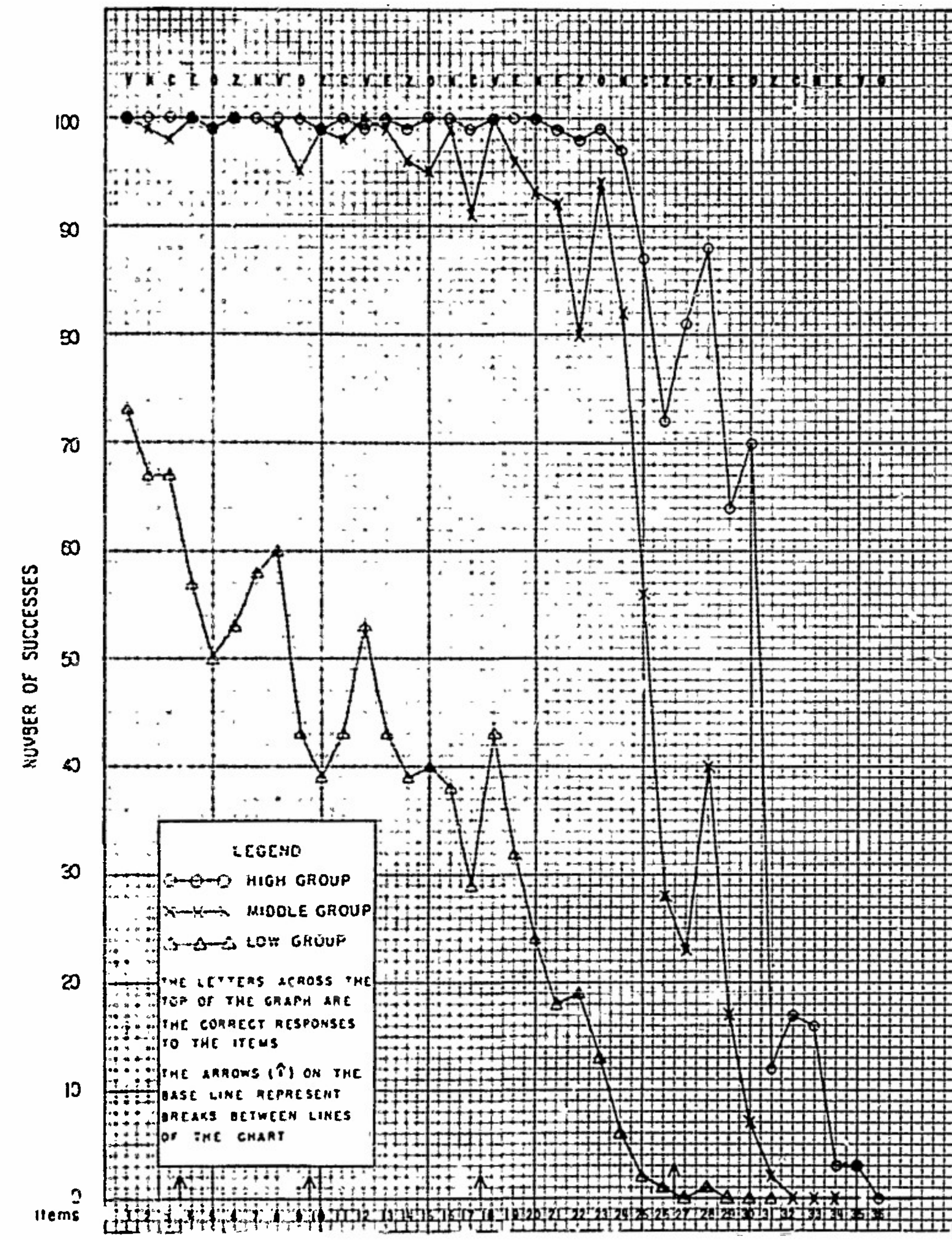
FIGURE 39.—*Line Resolution Test—Test 9.*

791630—48——9



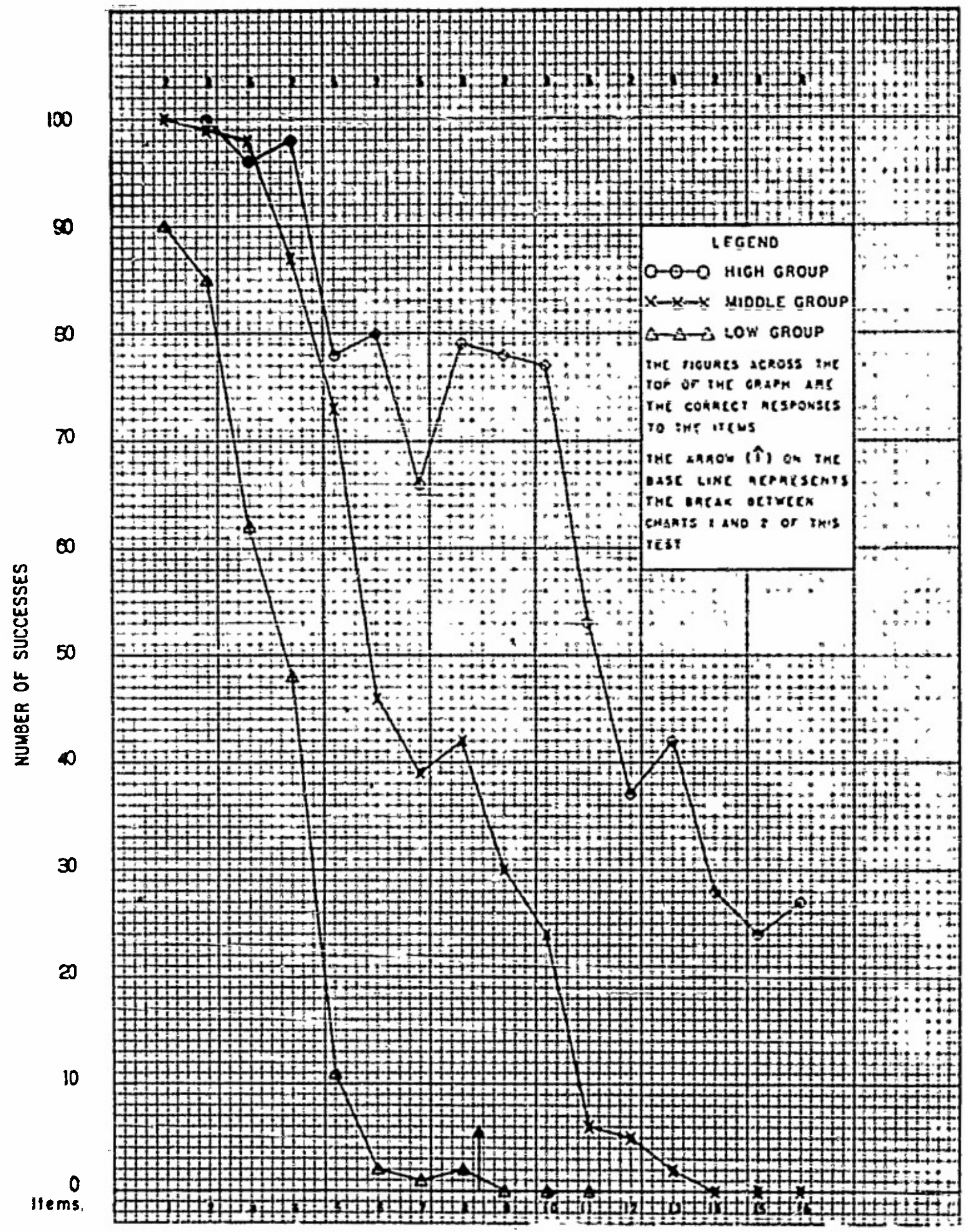
The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 40.—*Dot Variable Contrast Test—Test 10.*



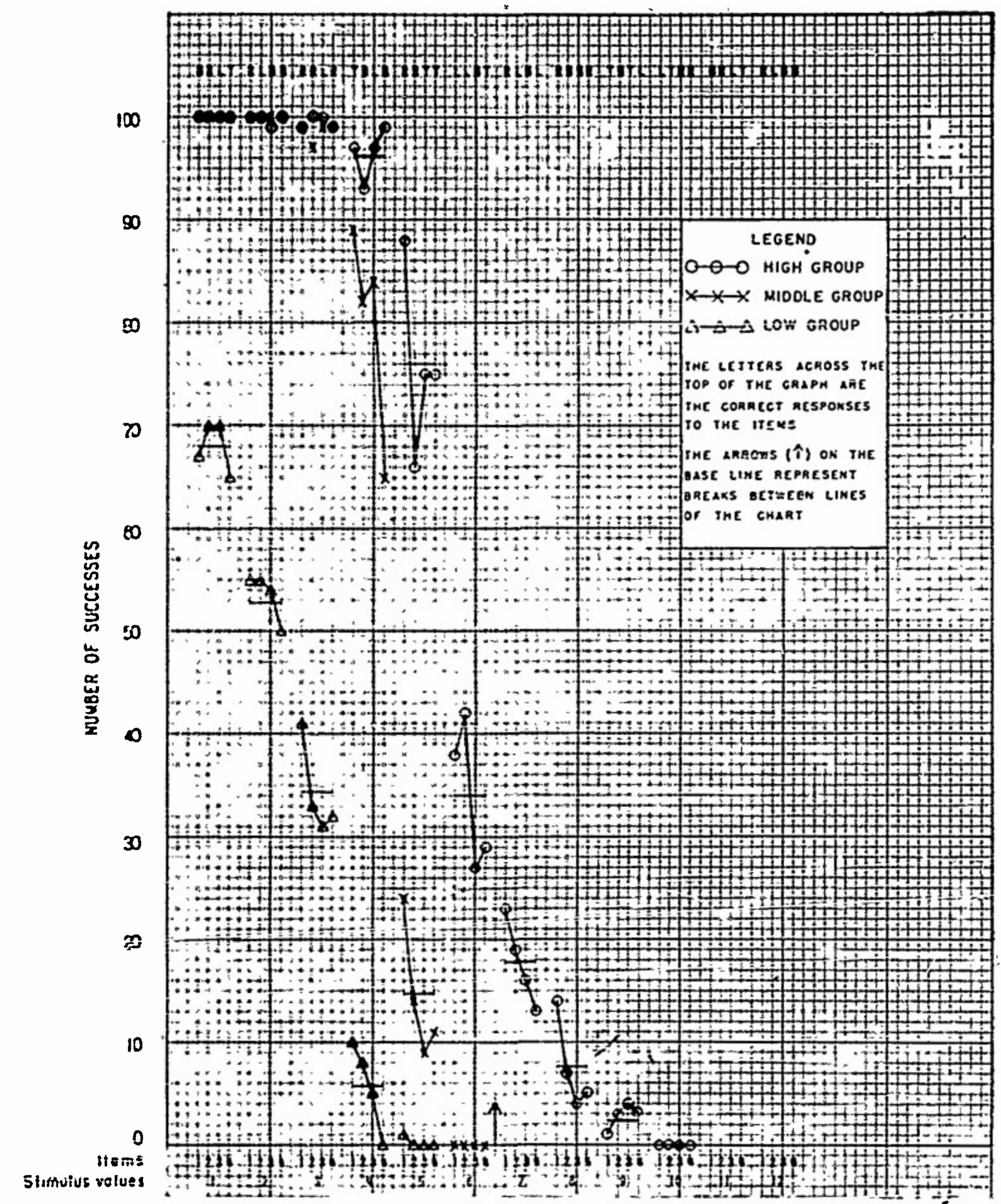
The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 41.—*AAF Constant Decrement Test—Test 11.*



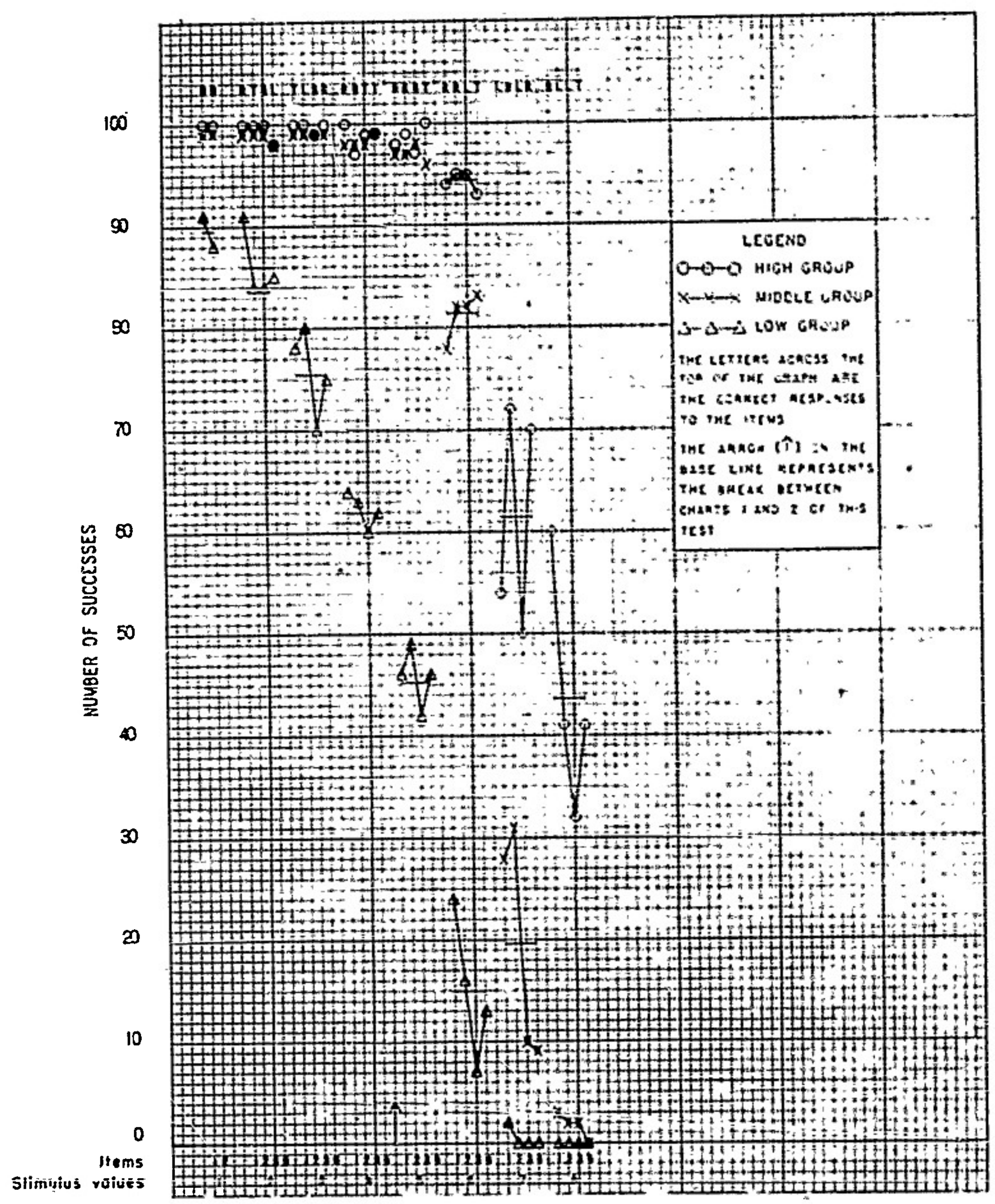
The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 42.—*Square Discrimination Test—Test 12.*



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 43.—*Checkerboard Variable Grid Test—Test 13.*



The 3 plots represent successes attained by high, middle, and low ability groups of 100 each, selected on the basis of number right to 3 consecutive misses.

FIGURE 44.—*Vernier Acuity Test—Test 14.*

ber of people guessing at any particular point, hence the theoretical expectancy values from which the results may deviate, is virtually impossible from the present data. However, a contingency table based on right (R) and wrong (W) responses can be used effectively since such a method automatically takes care of the fewer attempts accumulated on any item position in a given line. Table XXI presents the results of reworking the data of this test into a contingency table. The table is based on the data from lines 5, 6, and 7 of the high ability group.

TABLE XXI.—*Distribution of right (R) and wrong (W) responses for the attempts (A) made by the high group on the four item positions of lines 5, 6, and 7 of test 13, checkerboard variable grid*

		Item positions				Total
		1	2	3	4	
Responses	R	149(142)	127(134)	118(126)	117(109)	511
	W	109(116)	116(109)	112(104)	81 (89)	418
	A	258	243	230	198	929

From the data presented in the table above, the hypothesis that there is no difference in the difficulty of items attributable to their position in a line of the test chart, except that expected by chance, random sampling may be tested in terms of the distribution of χ^2. The theoretical frequencies are given in parentheses in the table. The value of χ^2 is 4.001. A χ^2 as large as this for 3° of freedom would be exceeded by between 30 and 20 percent of all random samples. There is no reason, therefore, for rejecting the hypothesis. However, this does not establish the hypothesis as true. Further search may demonstrate some relation between difficulty and line position. Nor does this result mean that one can confidentially assume that the hypothesis will not be rejected when applied to the data for other ability groups on test 13 and that for all ability groups on other tests of similar design. In addition, the test of the hypothesis as made in this small sample, does not take into consideration the mode of response—top, bottom, right, or left. Both the position in line and the mode of response may influence the result. Only when an equal sample of top, bottom, right, or left responses has been collected at all line positions will any thorough test of the line position hypothesis be obtainable. A similar condition applies to any study of differential difficulty in relation to the mode of responses. In this case line position must be controlled.

A rather startling result is found in terms of the use of two charts for the test. The break between charts occurs between lines 6 and 7 of the test as indicated by the arrow in figure 43. Reference to the specifications for these two lines of the test reveals that by accident the same stimulus value is employed in both. Hence, it is surprising to find that the numbers of successes on items of line 7 are so much lower than those on line 6 items. It indicates that serious errors in measurement may occur if tests are broken in half, and it may be best to present items on one chart or individually to avoid any possibility of breaks in procedure producing a shift in the level of difficulty of the task.

10. *Test 14. Vernier Acuity* (fig. 44). The action of individual items on this test is similar to that mentioned with respect to other nonletter tests.

5. Analyses of Test Conditions, Correlation of AGCT and Letter Test Scores, and Examinee Opinions About the Tests

This paragraph is devoted to certain minor considerations derived from the data collected at Fort Dix, N. J., which, while not of primary interest, are among the important items of information which may expand knowledge in the field of vision testing and point the way to future efforts.

a. Test Conditions. Variance in vision test score may not only be attributable to factors such as those mentioned in paragraphs 2 and 3 of these results but may also be a function of unreliability and other specific factors associated with the tests. Among the sources of error in individual examinations, such as those of the present study, are: The portions of variance attributable to divergent test booth conditions (brightness, condition of equipment); those introduced by different examiners; those introduced by the order of presentation of the tests; and those portions of variance attributable to the time of day at which the tests are administered. In analyzing the variance attributable to these test conditions the basic idea is to test certain null hypotheses, all couched in terms of expecting no influence of variables, aside from individual visual ability, other than that expected by chance random sampling. It is clear that in such a process the hypotheses are never established as completely true. What is learned is with what degree of confidence we may or may not reject the hypotheses. If the null hypotheses cannot be rejected with confidence, it is apparent that test conditions are not contributing enough systematic error to worry about them, although the question will always remain as to how much they might contribute if standards of administration

are relaxed. If, on the other hand, the hypotheses can quite confidently be rejected (assuming that an adequate random sample has been analyzed), then it behooves test administrators to look to the conditions of administration and to seek to eliminate the sources of significant error. In the present study it is interesting to examine this material, because it is hoped that such extra-ability sources of variance did not enter significantly.

(1) *Variance attributable to test booths.* In utilizing six test booths to collect examination data, there is a possibility that serious systematic variations may arise between the booths. Illumination, hence chart and wall brightness, may vary; and other sources of variance may accrue from the supervisors and examiners who worked in the different booths. In any case, the hypothesis to be tested is that no difference in test performance is attributable to test booth variations other than that expected by chance random sampling. To test this hypothesis the within group variance (that attributable to individual differences and other unanalyzed sources of error) is compared with the between groups variance (that attributable to differences in test booths). The ratio of these variances (between-group/within-group) is known as F, and the size of this statistic may be compared with known values of F which will occur by chance. The analysis is summarized in table XXII.

From the last column of table XXII it can be seen that for most tests the computed value of F is lower than 1.00. Entering a table of F it is seen that for 5 and 786 degrees of freedom, for the respective variances, a value as large as 2.22 is expected to occur by chance in only 5.0 percent of all random samples. Accepting this as the level of confidence, it is evident that in all but one test the computed value of F is lower than 2.22. This means that for these tests the null hypothesis cannot be rejected with sufficient confidence to declare that booth variance is significant. In the Quadrant Variable Contrast Test (4), however, an F of 2.72 is found. Such a high F (higher than 1.80) is not expected in more than 5.0 percent of all random samples. In this case, the hypothesis can be rejected with some confidence, and it may be said that variance attributable to differences in test booths is serious enough to cause some concern. That the Quadrant Contrast Test is the one effected (Note.—The Dot Contrast (10) has the next highest F value) is interesting since it was felt at the time of booth construction that poor lighting conditions would effect the contrast tests more seriously. The mean scores on the Quadrant Contrast Test (4) are 4.99, 4.69, 4.89, 5.09, 4.46, and 4.27 for booths 1 to 6, respectively. Notice that the last two booths, 5 and 6, have the lowest means. In seeking improvements these two booths should be examined, especially from the point of view of illumination conditions.

Before leaving the analysis of booth variance, it must be emphasized that the experimental design for this study did not prepare primarily for this approach. Hence, the present test-booth variable is confounded with any variation in examiners who happened to be continuously associated with this or that booth.

In the next analysis the examiner variable is studied separately and, here too, any significant variation in examiners may be due in part to systematic variations in the booths to which they are assigned.

(2) *Variance attributable to different examiners.* Certain test results may be expected to differ according to who gives the tests. Some examiners may motivate the examinees, while others may not. No matter what variables are introduced by variation in examiners, it is interesting to note the effect of this possible source of error on the various tests. The hypothesis to be tested is that no variation in test performance occurs due to the different individuals employed as examiners except that expected from chance random sampling. The same method is used to test the hypothesis in this situation as was used to examine the booth situation. The results are presented in table XXIII.

TABLE XXII.—*Between-groups (booths) and within-group variance, the degrees of freedom for each variance, and computed value of F for each of the 14 wall chart visual acuity tests*

Test	Variance between	d of f	Variance within	d of f	F
1. Practice	36. 8	5	46. 0	786	0. 80
2. Army Snellen	106. 4	5	95. 4	786	1. 11
3. Dot Var Size	35. 0	5	95. 3	786	. 37
4. Quad Var Contrast	11. 4	5	4. 2	786	2. 72
5. New London Letter	140. 0	5	261. 0	786	. 54
6. Triangle Discrim	21. 4	5	29. 0	786	. 74
7. B and L Checkerboard	25. 4	5	34. 2	786	. 74
8. AAF Letter	170. 6	5	198. 0	786	. 86
9. Line Resolution	107. 2	5	146. 2	786	. 73
10. Dot Var Contrast	6. 2	5	3. 9	786	1. 60
11. AAF Constant Decr	36. 4	5	77. 7	786	. 47
12. Square Discrim	7. 4	5	11. 0	786	. 68
13. Checkerboard Var Grid	43. 8	5	47. 7	786	. 92
14. Vernier	32. 6	5	36. 4	786	. 90

With the degrees of freedom given in table XXIII, it is discovered that an F value of 1.80 is expected by chance in only 5.0 percent of all random samples. The computed values of F for three of the tests exceed this value, 2.43 for test 4, 1.87 for test 10, and 3.08 for test 12. For these tests the null hypothesis can be rejected with the indication that examiner differences are significantly greater than one would expect by chance. For all other tests the hypothesis cannot be rejected with confidence.

Examiner veriability seems to play a large part in the test score variance of the two contrast tests (4 and 10) and the Square Discrimination Test (12). It is interesting to note that these are the three least reliable tests of the present battery. In the case of the Square Test (12) a considerable degree of error seems to be introduced by examiner variation while, as we have seen, the null hypothesis as to booth influence cannot be rejected. Before concluding that examiners made a mess out of the square or any other test, it is well to realize that the interaction of test and examiners may account for this. The test may be poor and difficult to administer, thus producing a sharp distinction between examiners. With simplification of instructions and other improvements the present finding regarding variance due to examiners may disappear. This does not mean that poor examiners should be kept on the job, but it seems economical of personnel to clean up an instrument as well as to take corrective measures with examiners.

TABLE XXIII.—*Between-groups (examiners) and within-group variance, the degrees of freedom for each variance, and computed values of F for each of the 14 wall chart visual acuity tests*

Test	Variance between	d of f	Variance within	d of f	F
1	60. 6	11	45. 8	776	1. 32
2	115. 0	11	95. 4	776	1. 21
3	43. 2	11	99. 8	776	. 43
4	10. 1	11	4. 2	776	2. 43
5	234. 4	11	261. 0	776	. 90
6	29. 6	11	28. 9	776	1. 02
7	26. 4	11	34. 2	776	. 77
8	144. 6	11	198. 6	776	. 73
9	168. 1	11	146. 2	776	1. 15
10	7. 2	11	3. 8	776	1. 87
11	51. 6	11	78. 0	776	. 66
12	32. 8	11	10. 7	776	3. 08
13	25. 6	11	48. 0	776	. 53
14	34. 6	11	36. 5	776	. 95

(3) *Variance attributable to the order of presentation of the test.* It is hoped that through systematic rotation of the order of presentation of the tests, certain variables, such as fatigue, have been equated. In this sense, it is interesting to find what effect is obtained on each test in terms of its relative position in the order of tests administered. For this purpose the Practice Test (1) is dropped since it always occurred first in any session. The hypothesis to be tested is that no variation in performance on any tests is noted as a function of the position of the test in the order of presentation except that variation which might be expected from chance sampling. Table XXIV gives the results of this analysis.

As can be seen from table XXIV no computed value of F is as great as the 1.76 value which the table of F lists as expected by chance in 5.0 percent of all random samples. The null hypothesis cannot, therefore, be rejected with enough confidence for any test. Apparently, no extra-chance variations arise due to the serial position of any test in the battery. This result tends to confirm the suspicion that test rotation would eliminate any variance in score due to position of the tests in serial order. However, the hypothesis is not demonstrated as absolutely "true." Such is not possible, as pointed out previously, and the hypothesis should be tested periodically in any multiple testing program until one can be certain that the order of presentation is not a source of error.

TABLE XXIV.—*Between-groups (order of test presentation) and within-group variance, the degrees of freedom for each variance, and the computed values of F for each of the wall chart visual acuity tests*

Test	Between variance	d of f	Within variance	d of f	F
1	(1)				
2	71.4	12	95.0	771	0.75
3	64.3	12	99.0	771	.65
4	5.7	12	4.2	771	1.36
5	166.3	12	260.9	771	.64
6	26.0	12	41.0	771	.64
7	36.5	12	34.0	771	1.07
8	139.2	12	198.2	771	.70
9	136.2	12	145.3	771	.94
10	2.7	12	3.9	771	.69
11	72.5	12	77.0	771	.94
12	6.3	12	11.0	771	.59
13	29.4	12	47.7	771	.63
14	23.0	12	36.0	771	.64

[1] Omitted.

(4) *Variance attributable to the hour of day in which the test session was held.* Diurnal (within the day) variations have long been suspected to have a lot to do with a variety of performances. If one eats a heavy meal at noon, it is felt that his subsequent afternoon activities will be slowed down. The same point is considered worthy of analysis in the present study. Perhaps certain vision tests invoke systematically better performance at certain hours of the day than at others. In analyzing the data the hypothesis is that no variation in performance can be attributed to the various hours of the day in which the testing occurred except that expected as a result of random sampling. The results of the analysis are given in table XXV.

In Table XXV, for only one test, Practice (1), can the hypothesis be rejected. Entering a table of F at the degrees of freedom involved,

it is discovered that an F as high as 1.89 will be found by chance in only 5.0 percent of all random samples. The computed F of 1.95 for the practice test exceeds this value, therefore the null hypothesis can be rejected.

Why should the practice test appear to be influenced by the time of day in which it is administered? One suggestion is that since this test was not rotated as to position in the test battery, the effect of the time of day operated in unequal fashion on tests given first at any particular time. This is, admittedly, *ex post facto* reasoning, but it makes some sense. If after lunch, for example, it takes a few minutes to warm-up to testing whereas at other times the diurnal variation has little effect, then any test which is not rotated should receive the

TABLE XXV.—*Between-groups (hour of testing) and within-group variance, the degrees of freedom for each variance, and the computed values of F for each of the 14 wall chart visual acuity tests*

Test	Between variance	d of f	Within variance	d of f	F
1	87.8	9	45.1	780	1.95
2	83.6	9	95.2	780	.88
3	108.8	9	98.1	780	1.11
4	2.6	9	4.2	780	.60
5	401.8	9	257.9	780	1.56
6	33.3	9	28.9	780	1.15
7	45.2	9	33.8	780	1.34
8	279.2	9	196.9	780	1.42
9	151.8	9	145.7	780	1.04
10	3.3	9	3.9	780	.86
11	79.9	9	77.4	780	1.03
12	14.7	9	10.9	780	1.35
13	74.1	9	47.3	780	1.56
14	51.4	9	36.0	780	1.43

full effect of this temporal variable. On the other hand, tests that are rotated as to temporal appearance in any session should be unaffected by the time of day. Examination of the means for different hours of testing for this practice test does not confirm the luncheon hypothesis. The highest mean score (24.9) on the practice test is achieved in the first afternoon session. The lowest means are found for the next to last sessions of both morning and afternoon testing. No rationale is offered for this rather unimportant practice test activity. Instead, it should be emphasized that, in general, the null hypothesis regarding diurnal variations cannot be rejected in all the other tests. Although time of day appears to be insufficiently important in the present study, this is no indication that it may not become significant as a function of certain laxities in testing methods.

b. CORRELATION OF AGCT AND LETTER TEST SCORES. In part 2 of the results, a discussion of factor III (form (letter) perception) contemplates the possibility that a verbal factor measured by AGCT may be related to something measured by the letter tests in the visual acuity test battery. Naturally, the correlations of these tests are of interest. They are presented for both a test and retest sample, of 205 cases each, in table XXVI.

Table XXVI reveals that practically zero correlation exists between either test or retest scores on the letter tests and AGCT score. As mentioned in connection with the discussion of the form (letter) factor, AGCT may not measure ability related to that required by the present letter acuity tests. In the letter tests, letter facility may be an important factor, where as in AGCT whole word facility is more likely to be measured. Another possible explanation of the lack of relationship demonstrated in table XXVI is the small amount of variance that can be attributed to the form (letter) factor in analysis of any of the letter tests. Since score on these vision charts varies so little on the basis of the factor, only slight correlation with AGCT is expected.

TABLE XXVI.—*Intercorrelations of AGCT standard scores with test and retest total score (number right to 3 consecutive misses) on the 4 letter tests of the wall chart tests studied at Fort Dix, N. J.*

Tests	Test with AGCT	Retest with AGCT
Test 2. Army Snellen	−0. 001	0. 000
Test 5. New London Letter	. 022	. 026
Test 8. AAF Letter	−. 013	−. 006
Test 11. AAF Const Decrem	. 009	−. 005

c. EXAMINEE OPINIONS ABOUT THE TESTS. Of somewhat less interest than other results, but, nevertheless, of some importance in planning for the future, are the opinions of examinees about the tests. To serve as a cue as to which of two similar tests is preferred by examinees the responses to the pertinent questions are presented in table XXVII.

From table XXVII the contrast tests (4 and 10) seem to be liked least and disliked most. On the other hand, tests 2 and 14 are the most liked tests and, at the same time, only slightly disliked. From the size of the "vote" there can be little doubt that these preferences exist. On other tests the group tested does not seem to hold the same overwhelming opinions. However, it can be seen that tests 1, 5, 8, 9, and 11 are more liked than disliked, while tests 3, 6, 7, 12, and 13 are mort disliked.

In general, it can be said that the letter tests are liked, while nonletter tests of about the same difficulty are disliked. The most disliked tests are those which by examiner remarks and actions are most difficult. These are the two contrast tests (4 and 10). Conversely, the fairly easy Vernier test (14) is well liked by the examinees along with the Snellen test (2). There is some question of how well the tests will be liked after suitable revisions make them better testing instruments. If difficulty is the key to preference, increasing test difficulty may reduce examinee preference for such tests. Yet, there is no reason to doubt that, aside from difficulty, some test objects are more appealing than others. In any choice of instruments for use, the preference of examinees may become the deciding factor after all other objective criteria have been applied in the selection process.

TABLE XXVII.—*Opinions of examinees as expressed in answer to questions: Which test did you like most? Which test did you like least?*

Tests	Like most	Percent	Like least	Percent
1. Practice	50	6. 5	1	0. 1
2. Army Snellen	175	22. 6	7	. 9
3. Dot Var Size	15	1. 9	33	4. 2
4. Quadrant Var Contrast	4	. 5	155	19. 6
5. New London Letter	39	5. 0	18	2. 3
6. Triangle Discrim	9	1. 2	31	3. 9
7. B and L Checkerboard	68	8. 8	123	15. 6
8. AAF Letter	42	5. 4	4	. 5
9. Line Resolution	94	12. 1	4	. 5
10. Dot Var Contrast	3	. 4	288	36. 5
11. AAF Contant Discrim	49	6. 3	5	. 6
12. Square Discrim	11	1. 4	67	8. 5
13. Checkerboard Var Grid	22	2. 8	35	4. 4
14. Vernier Acuity	193	24. 9	19	2. 4
	774	100. 0	790	100. 0

Section VI. SUMMARY AND CONCLUSIONS

1. Data

Two sets of data are analyzed to give the results presented in the report. The two sets are:

a. Data collected from the administration of 14 experimental wall chart tests of visual acuity to 792 enlisted men at Fort Dix, N. J. This data was collected by technicians of the Personnel Research Section, Personnel Research and Procedures Branch, The Adjutant General's Office. The tests, test booths, and procedures employed are described in sections II, III, and IV of this report.

b. Data collected from the administration of vision tests in three commercial devices (Bausch and Lomb Orthorater, American Optical Sightscreener, and Keystone Telebinocular), wall chart acuity tests, and Maddox Rod phoria tests to 128 subjects. This data was collected by United States Navy personnel at Medical Research Department, United States Submarine Base, New London, Conn. The devices, tests, and procedures employed are described in previous Navy publications.

2. Treatments of the Data

The main object of the study is to determine the aspects or factors present in the vision tests employed. From the study of the 14 wall chart tests, results are presented showing the factors in visual acuity (far) tests. From the study of three commercial devices and other tests, the factors in far and near acuity, depth, and phoria tests are presented. In addition, for the 14 wall chart tests, results are presented showing the test-retest reliability by 7 methods of scoring, the frequency distributions of test scores, analysis of item difficulty, analysis of testing conditions, correlation of letter tests with AGCT, and examinee opinions. All these studies are for the purpose of improving the wall chart tests as measuring instruments based upon determinations of how they operate in their present form.

3. Findings and Conclusions From Factor Studies

a. Analysis of the Fort Dix Data (Factors in Visual Acuity Test [Far]). Four factors are identified which account for test score variance in the 14 wall chart (far) acuity tests (see sec. V, par. 2, table X). They are:

Retinal Resolution.
Brightness Discrimination.
Form (Letter) Perception.
Simple Form Perception.

Factor I, *Retinal Resolution*, accounts for most of the test score variance on all of the tests. It is measured quite purely by the Bausch and Lomb Checkerboard and the Checkerboard Variable Grid Test (7 and 13). Letter Tests (5, 8, 2, and 11), Dot Variable Size Test (3), and Line Resolution Test (9) measure this factor to about the same extent as do the checkerboard tests. About 80 percent of test score variances on all of these tests is accounted for by factor I.

Factor II, *Brightness Discrimination*, is measured best by the Quadrant (4), and Dot (10) Variable Contrast Tests, but other tests, such as Line Resolution (9) and Vernier Acuity (14), have sigificant loadings for this factor. The amount of test score variance accounted for by this factor is never as high as that accounted for by factor I. On the two contrast tests factor II accounts for about 16 percent of score variance while factor I accounts for about 35 to 45 percent of the variance.

Factor III, *Form (Letter) Perception*, accounts for as much as 10 percent of test score variance on Letter Tests (5, 8, 2, 11). It also appears, though less consistently on the Square Discrimination Test (12), but does not have significant loadings on any other test.

Factor IV, *Line Form Perception*, accounts significantly, but to a small extent (about 10 percent) for test score variance on the Triangle (6) and Square (12) Discrimination Tests. It does not have significant factor loadings on any other test.

From these and other findings it is concluded that—

(1) Retinal Resolution (I), the general factor, is the basic concept underlying visual acuity measurement at 20 feet, with the present set of tests.

(2) Only the Bausch and Lomb Checkerboard and Checkerboard Variable Grid Tests purely measure Retinal Resolution. For all of the other tests, lesser loadings, are found on either Brightness Discrimination (II), Form (Letter) Perception (III), or Simple Form Perception (IV) factors.

(3) Letter tests, Dot Variable Size and Line Resolution tests, measure Retinal Resolution as well as do the checkerboard tests if the relatively small impurities introduced by other factors create no objections to the use of these tests for given purposes.

(4) Tests which measure Brightness Discrimination (II), Form (Letter) Perception (III), and Simple Form Perception (IV), do so inadequately in their present form since Rational Resolution (I) accounts for so much more of test score variance.

(5) Of the factors other than Retinal Resolution only Brightness Discrimination seems capable of measurement in pure form. To accomplish factorial purity of such tests, spatial (resolution) references in the form of the response to items should be eliminated.

(6) Chances of discovery of other factors in the wall chart type of

acuity test, other than those specific to different perceptual tests, are slight since the present factors leave little test score variance unaccounted for except that due to error in measurement.

b. ANALYSIS OF SUBMARINE BASE DATA. Nine factors have been finally or tentatively identified by the analysis of tests employed in the three commercial devices, wall charts, and Maddox Rod phoria tests (see sec. V, par. 3). They are:

Retinal Resolution.
Form (Letter) Perception.
Resistance to Interference.
Depth Perception.
Lateral Phoria.
Near Lateral Phoria (Convergence Efficiency).
Vertical Phoria.
Antifusion Factor.

(1) *Factors in acuity tests.* Factor A, *Retinal Resolution*, has significant loadings on all acuity tests (1–26) and on all depth tests (27–36) as well. The loadings are highest for the far tests (1–13) and slightly lower for the depth tests (27–36) than for the near tests (14–26). This factor is undoubtedly the basic measure of visual acuity, and is accordingly named *Retinal Resolution.*

Factor B, *Accommodation*, has significant loadings on all near tests (14–26 of table XIV) and on near depth tests (35–36 of the table) as well. With few exceptions, loadings of the factor on other tests are insignificant.

Factor C, *Form (Letter) Perception*, has significant loadings on all wall charts which are also letter charts (1–5 and 14–16, table XIV). Significant or near significant loadings appear on machine letter charts (10–13, 21–22, and 25–26), and Telebinocular near (circles) test (23–24). Nearly significant negative loadings appear on Telebinocular far depth tests (29–30). All require form naming. On other tests loadings are insignificant.

Factor D, *Resistance to Interference*, has its highest loadings on the Orthorater Acuity Tests (6–9 and 17–20, table XIV) with occasional nearly significant loadings on Sightscreener and Telebinocular tests. One or more elements known to interfere with visual perception can be found in connection with the test charts employed, hence, the tentative name given to this factor.

(2) *Factors in depth tests.* Factor E, *Depth Perception*, has loadings on all depth tests and only on depth tests. The loadings vary from moderate to barely significant and are in many cases no higher than the loadings for Factor A, Retinal Resolution, on the same tests.

(NOTE.—A Sightscreener Depth Specific factor is identified as Factor F in the analysis.)

(3) *Factors in phoria tests.* Factor F, *Lateral Phoria* (identified

as factor A in the analyses, see table XVII), has loadings on all lateral tests, far (17–27 in the table) and near (28–38), and loadings are significant only for these tests. Generally, much higher loadings are obtained for the far tests. The Maddox Rod measures this factor more poorly (36.0 percent) than does the poorest machine test (50.4 percent).

Factor G, *Near Lateral Phoria (Convergence Efficiency)* (identified as B in the analysis), has loadings on only the near lateral phoria tests. Tests with control of fusion (32–33 and 38) show loadings almost identical to those for tests having no such control (28–31 and 36–37).

Factor H, *Hyperphoria (Vertical Phoria)* (identified as C in the analysis), has loadings on all the vertical phoria tests, far (1–9) or near (10–16).

Factor I, *Antifusion* (identified as D in the analysis), has loadings on all Maddox Rod measurements (9, 16, 27, 38), on the excursion measurements on the Orthorater (21–22 and 32–33) and on the Sightscreener vertical phoria tests near (14–15) and far (5–6). An attempt to prevent the action of the fusion center is common to all these tests.

NOTE.—A Vertical-Rest factor is identified as factor E in the analysis with loading only on the original far vertical phoria tests.

An Orthorater (Near Vertical) Specific factor is identified as Factor F in the analysis.

An Orthorater (Far Lateral) Specific factor is identified as Factor G in the analysis.

Other devices are seen to possess similar specific factors by the many large residuals remaining in tables XV and XVI.

(4) *Relationship of factors* (see table XIX). Retinal Resolution, Accommodation, Form (Letter) Perception, Resistance to Interference, and Depth Perception factors all have zero or near zero loadings on the phoria tests. Similarly, the Near Lateral (Insufficient Convergence) and Excursion (Antifusion) factors have zero loadings on all acuity and depth tests.

The lateral phoria factor has consistent negative, although not quite significant loadings on the three near vision tests (4–6). These loadings indicate a positive relationship between extreme exophoria and difficulty on near acuity tests.

The three depth perception tests (7–9 in table XIX) all have negative although not quite significant loadings on the general Vertical Phoria Factor (C in the analysis).

(5) From these and other findings it is concluded that—

(*a*) Retinal Resolution (A) is the basic concept underlying measurement of visual acuity.

(*b*) Near acuity and depth tests are greatly influenced by the resolving power of the retina. Near acuity and depth test scores should either be corrected on the basis of far test scores or should be determined only after proper lenses to bring far vision to normal have been supplied to all subjects.

(*c*) The Accommodation Factor (B) plays a greater role in near acuity tests of the dominant eye (usually right eye).

(*d*) Loadings of the Accommodation Factor on certain far acuity tests appear as errors introduced by correctable defects in the specific instruments or tests.

(*e*) The Form (Letter) Perception Factor (C), plays a lesser role in the tests and may enter test scores as an unnecessary complication.

(*f*) Resistance to Interference, Factor (D), is a source of test score pollution related to defects in the design of certain test charts, which interfere with visual perception.

(*g*) Since all presently available acuity and depth vision tests, wall chart or machine, have large parts of their variance determined by nonpertinent factors much remains to be done to make them really satisfactory.

(*h*) Any validity studies based upon a single type of visual test can only with great danger be used to predict action based upon some other method *supposedly* measuring the same function.

(*i*) Lateral Phoria, Factor (F), appears to be a general measure of Lateral Phoria, far and near, with machine tests measuring it better than the Maddox Rod test.

(*j*) Near Lateral Phoria, Factor (G), appears to deal with convergence efficiency at the normal reading distance rather than with errors of refraction or control of fusion.

(*k*) Interpupillary distance is unrelated to phoria scores as measured by the present study.

(*l*) As in the case of acuity and depth tests, so for phorias, it is dangerous to assume interchangeability of either interpretation of scores or validities from one machine or method to another.

(*m*) There is a general lack of relationship between acuity and depth factors and phoria factors.

(*n*) In general it is concluded that all nine factors are relatively independent and that any thorough study of the role of visual factors in military or other occupational success must be aware of the possible importance of any one or any combination of them.

4. Reliability of the 14 Wall Chart Tests (see Sec. V, par. 1)

Of the seven methods of scoring attempted in this study there is little to choose between method C, number right to two consecutive misses and method G, number right to three consecutive misses. Both

are highly intercorrelated. If anything, method G gives the highest test-retest intercorrelations. They are:

Test	R's	Test	R's
1. Practice	0.80	8. AAF Letter	.89
2. Army Snellen	.88	9. Line Resolution	.85
3. Dot Variable Size	.85	10. Dot Variable Contrast	.43
4. Quadrant Variable Contrast	.57	11. AAF Constant Decrement	.87
5. New London Letter	.88	12. Square Discrimination	.44
6. Triangle Discrimination	.69	13. Checkerboard Variable Grid	.81
7. Bausch and Lomb Checkerboard	.79	14. Vernier Acuity	.80

When the Spearman-Brown prophecy formula is applied to tests 4, 6, 10, and 12 of this list, doubling test length raises the respective correlations to 0.74, 0.84, 0.66, and 0.61.

From these and other findings it is concluded that—

a. Most of the present tests have satisfactory reliability, exceptions being Quadrant Variable Contrast, Triangle Discrimination, Dot Variable Contrast, and Square Discrimination.

b. Doubling the present length of the Triangle Discrimination Test will most likely raise its reliability to a satisfactory level. Other presently unreliable tests must be improved in other additional ways to become more reliable.

5. Frequency Distributions of Test Scores on the 14 Wall Chart Tests (see Sec. V, par. 4, and fig. 17–30)

a. Skewness. Most of the tests are severely negatively skewed. These are tests 1, 2, 3, 4, 7, 8, 9, 11, and 14, showing that they are either too easy in their present forms or are prone to methodological influences leading to negative skewness. Tests 6, 10, and 12 are positively skewed. Test 5 is irregular, but shows some negative skewness.

b. Minor Modes. Several of the tests are too difficult for the poorer eyes tested. They all show minor modes at zero or near zero score and are tests 3, 4, 5, 6, 7, 8, 10, 11, 12 and 13. Tests with pile up of near perfect scores include tests 6 and 9.

c. From these and other findings it is concluded that none of the present tests show all the requirements of adequate measuring instruments and all should be revised to improve their range and discrimination of ability in accordance with the use to be made of them.

6. Item Difficulty in the 14 Wall Chart Tests (see Sec. V, par. 4, and fig. 31–44)

Item difficulty analyses confirm the inadequacies of the present tests as measuring instruments. The relation of number of successes made by high, middle, and low ability (total score) groups on each of the test items of the tests indicates specific deficiencies on all tests. From the point of view of discrimination of ability, separation of the high, middle, and low groups is fairly good. The main deficiencies are discrimination over too narrow ranges of items and incomplete measurement of the high or low groups. Specific suggestions are made for the improvement of all tests on the basis of these analyses. (See sec. V, par. 4.)

Inspection of the action of individual items shows that greater variability in the difficulty of items of like size occurs in letter charts than in nonletter charts. Letters of the Army Snellen and the New London Letter tests are particularly variable in this respect with the AAF Letter test items being considerably less variable.

There are some indications that the method of test administration as well as the mode of presentation of items may bear considerably on the action of individual items. From these and other findings it is concluded that—

a. All tests need revision on the basis of present suggestions from the item analyses and future studies.

b. Lines of items of equal difficulty should be completely attempted rather than stopping tests in the middle of any line.

c. Studies should be made of the mode of presentation of items with special respect to the effects of using two charts, line-to-line or item-to-item decrements, and serial order of presentation of the items.

7. Analysis of Variance Among Test Conditions for Administration of the 14 Wall Chart Tests (see Sec. V, par. 5, tables XXII–XXV)

An analysis of the variance due to differences in test booths shows that in only the Quadrant Variable Contrast are booth differences significantly higher than what is expected by random sampling (5 percent level).

An analysis of the variance due to differences in examiners employed in the study shows that in only the Quadrant Variable Contrast, Dot Variable Contrast, and Square Discrimination tests are examiner differences significantly higher than what is expected by chance (5 percent level).

An analysis of the variance accounted for by the order in which tests were presented indicates that on no tests did this variable account for differences significantly greater than that expected by random sampling (5 percent level).

An analysis of the variance due to diurnal variations in testing time shows that in only the Practice Test are differences found that are other than what would be expected by chance (5 percent level). From these findings it is concluded that—

a. On most of the tests variables in testing procedure resulted in

no significant variations other than those expected by chance, i. e., controls were adequate.

b. Scattered significant influences of variables in the testing procedure (on Quadrant Contrast, Dot Contrast and Square Discrimination Tests) indicate the need for continuous study of testing conditions.

8. Relationship Between Letter Test Scores and AGCT (see Sec. V, par. 5)

Letter tests scores, both test and retest, from the data collected at Fort Dix, N. J., show close to zero correlation with AGCT standard scores for the same individuals. From this finding it is concluded that the Form (Letter) Perception factor found in all letter tests is not related to factors known to account for test score variance in AGCT.

9. Preferences of Examinees Among the 14 Wall Chart Tests (see Sec. V, par. 5)

From tabulations of opinions as to which tests, of the 14 wall chart types employed at Fort Dix, examinees liked most and least, the Army Snellen and the Vernier Acuity tests stand out as most preferable; the Quadrant Variable Contrast and the Dot Variable Contrast stand out as least preferable. In the light of future revisions which are definitely called for on the basis of other findings of this study, preferences of examinees on the present tests lead to no conclusions.

Appendix A

DETAILED SPECIFICATIONS FOR THE CONSTRUCTION OF TEST CHARTS

Theory of construction. As briefly described in section III of this report, the test charts for the study were constructed according to two principles of scaling item characteristics: Snellen type units bearing relationship to a theoretical concept of normal acuity, resolution of two points separated by a space subtending one minute of visual angle; and units bearing a logarithmic relationship to the visual angle.

The preconceived opinions of normal visual acuity generally held with respect to the Snellen units in the measurement of visual acuity has made it appear expedient for the Personnel Research Section, The Adjutant General's Office, to abandon the use of these units in favor of an arbitrary system completely divorced from any preconceived concepts of normal acuity. It was decided that the physical basis of unit to be used would be the visual angle, and that the AGO visual acuity unit would bear a logarithmic relationship to its physical counterpart, the visual angle (the usual psychophysical relationship). Starting from the assumption that perceivable increments in visual discrimination bear a constant relationship to the point of change and arbitrarily setting this constant at 2, the formula for the derivation of AGO visual acuity units was set at:

$$U=(10-\text{Log}_2 A=\text{Log}_2\frac{2^{10}}{A}$$

in which U is equal to the AGO visual acuity unit and A is equal to the visual angle. This unit has the property of being positive throughout the normal range of vision and yet describes in small enough units to be easily handled. Visual angles may readily be converted to AGO visual units with the use of a 5-place table of logarithms to the base 10. The following formula may be used for this purpose:

$$U=10-\frac{\text{Log}_{10} A}{0.30103}$$

Similarly AGO visual units may be converted to visual angles by the formula:

$$\text{Log}_{10} A=(0.30103)\ (10-U)$$
$$A=\text{Antilog}_{10}\ (0.30103)\ (10-U)$$

The attached table is provided for converting AGO units to visual angles in minutes. It also gives computations of the physical equivalents of the visual angles in Snellen units, Grow (Navy) units, and subtenses in inches at the 20 feet observation distance.

Visual acuity conversion table

$\text{Log}_2 \frac{2^{10}}{\text{(angle)}}$

AGO score	Visual angle (minutes)	Snellen unit	Grow unit	Inches at 20 feet
15. 50	0. 022	20/0. 4	909. 1/20	0. 00155
	. 026	20/0. 5	769. 2/20	. 00183
15. 00	. 031	20/0. 6	645. 2/20	. 00218
	. 037	20/0. 7	540. 5/20	. 00260
14. 50	. 044	20/0. 9	454. 5/20	. 00310
	. 052	20/1. 0	384. 6/20	. 00366
14. 00	. 062	20/1. 2	322. 6/20	. 00436
	. 074	20/1. 5	270. 3/20	. 00520
13. 50	. 088	20/1. 8	227. 3/20	. 00619
	. 105	20/2. 1	190. 5/20	. 00733
13. 00	. 125	20/2. 5	160. 0/20	. 00873
	. 149	20/3. 0	134. 2/20	. 01039
12. 50	. 177	20/3. 5	113. 0/20	. 01235
	. 210	20/4. 2	95. 2/20	. 01466
12. 00	. 250	20/5. 0	80. 0/20	. 01746
	. 297	20/5. 9	67. 3/20	. 02078
11. 50	. 354	20/7. 1	56. 5/20	. 02470
	. 421	20/8. 4	47. 5/20	. 02933
11. 00	. 500	20/10. 0	40. 0/20	. 03492
10. 75	. 595	20/11. 9	33. 6/20	. 04156
10. 50	. 708	20/14. 2	28. 2/20	. 04940
10. 25	. 842	20/16. 8	23. 8/20	. 05866
10. 00	1. 000	20/20. 0	20. 0/20	. 06984
9. 75	1. 189	20/23. 8	16. 8/20	. 08311
9. 50	1. 416	20/28. 3	14. 1/20	. 09880
9. 25	1. 685	20/33. 7	11. 9/20	. 11733
9. 00	2. 000	20/40. 0	10. 0/20	. 13968
8. 75	2. 378	20/47. 6	8. 4/20	. 16622
8. 50	2. 828	20/56. 6	7. 1/20	. 19760
8. 25	3. 364	20/67. 3	5. 9/20	. 23466
8. 00	4. 000	20/80. 0	5. 0/20	. 27936
7. 75	4. 757	20/95. 1	4. 2/20	. 33244
7. 50	5. 657	20/113. 1	3. 5/20	. 39520
7. 25	6. 727	20/134. 5	3. 0/20	. 46932
7. 00	8. 000	20/160. 0	2. 5/20	. 55872
6. 75	9. 514	20/190. 3	2. 1/20	. 66488
6. 50	11. 131	20/222. 6	1. 8/20	. 79040

TEST 1.—*Practice Test.*

The figures are squares with sides inclined at 45° to the horizontal. The width of the side lines is one-fifth the length of a side. One corner of each square has a square break equal in width to the width of a side line. The location of the break varies from figure to figure in chance order. The following table shows the number of figures, the width of the figures, and the width of the stroke in each figure, for each row. Dimensions are given in inches.

Row	Group	AGO unit	Visual angle of stroke (minutes)	Subtense of stroke (inches)	Snellen unit equivalent	Number of items
1	1	6.67	10.000	0.700	20/200.0	2
2	2	7.24	6.762	.472	20/135.2	3
3	3	7.81	4.556	.318	20/91.1	4
4	4	8.38	3.081	.215	20/61.6	4
5	5	8.95	2.076	.145	20/41.5	4
	6	9.51	1.404	.098	20/28.0	4
6	7	10.08	.945	.066	20/18.9	4
	8	10.66	.631	.044	20/12.6	4
	9	11.22	.430	.030	20/8.6	4

TEST 2.—*Army Snellen*

The sizes of all letters decrease from the top row down, and the number of letters per row increases as the size decreases. The width of the stroke decreases with the size. In Snellen steps, the letters in succeeding rows from top to bottom are:

Row	AGO unit	Visual angle of stroke (minutes)	Subtense of stroke (inches)	Snellen unit equivalent	Number of items
1	6.68	10.000	0.700	20/200.0	1
2	7.68	5.000	.349	20/100.0	2
3	8.19	3.500	.244	20/70.0	3
4	8.68	2.500	.175	20/50.0	4
5	9.00	2.000	.140	20/40.0	5
6	9.42	1.500	.105	20/30.0	6
7	10.00	1.000	.070	20/20.0	7
8	10.29	.750	.052	20/15.0	8
9	11.00	.500	.035	20/10.0	9

TEST 3.—*Dot Variable size*

A dot of constant blackness is located in one of four corners of a constant-sized white diamond on a field of uniform gray. In each row, there are four diamonds containing dots of equal diameter. The diameter of the dots decreases from the top row down. There are two charts, chart B being a continuation of chart A. The dots are located in the corners of the diamonds in chance order.

The following table gives the diameters of the dots in inches for each row:

Row	AGO unit	Visual angle of Dot diameter (minutes)	Dot diameter (inches)	Snellen unit equivalent	Number of items
(Chart A)					
1	9.25	1.685	0.117	20/33.7	4
2	9.50	1.416	.099	20/28.3	4
3	9.75	1.198	.083	20/23.8	4
4	10.00	1.000	.070	20/20.0	4
5	10.25	.842	.059	20/16.8	4
6	10.50	.708	.049	20/14.2	4
(Chart B)					
1	10.75	.595	.042	20/11.9	4
2	11.00	.500	.035	20/10.0	4
3	11.25	.421	.029	20/8.4	4
4	11.50	.354	.025	20/7.1	4
5	11.75	.297	.021	20/5.9	4
6	12.00	.250	.017	20/5.0	4

TEST 4.—*Quadrant Variable Contrast*

Each figure consists of 3 quadrants of a constant-sized square similar in shape to an arrowhead. The shade of the figure varies from darker than the background to a shade which almost matches the background. Contrast between figure and background decreases from left to right within a row and from the top row down. The apex of each figure points in any one of four directions horizontally or vertically in a chance order.

Reflectance values for the figures and the background and the contrast values for each item are as follows:

Items: Chevrons (co's)	Reflectance: R (percent)	Reflectance* difference ($R_{bg}-R_{co}$)	Contrast $\frac{(R_{bg}-R_{co})}{R_{bg}}$	Row
1	63.4	9.8	0.134	1
2	65.3	7.9	.108	1
3	67.1	6.1	.083	1
4	68.7	4.5	.061	2
5	69.9	3.3	.045	2
6	70.8	2.4	.033	2
7	71.4	1.8	.025	3
8	72.0	1.2	.016	3
9	72.4	.8	.011	3
10	72.8	.4	.005	4

*Reflectance of the background (bg): 73.2 percent.

Test 5.—*New London Letter*

The letters in each row are equated for difficulty and are of equal size and equal stroke width. The size of letters and the width of stroke decreases for succeeding rows from top to bottom. The specifications for letters in each row are as follows:

Row	AGO unit	Visual angle of stroke (minutes)	Subtense of stroke (inches)	Snellen unit equivalent	Number of items
1	9. 00	2. 000	0. 140	20/40. 0	5
2	9. 25	1. 685	. 117	20/33. 7	6
3	9. 50	1. 416	. 099	20/28. 3	7
4	9. 75	1. 189	. 083	20/23. 8	8
5	10. 00	1. 000	. 070	20/20. 0	8
6	10. 25	. 842	. 059	20/16. 8	8
7	10. 50	. 708	. 049	20/14. 2	8
8	10. 75	. 595	. 042	20/11. 9	8
9	11. 00	. 500	. 035	20/10. 0	8
10	11. 25	. 421	. 029	20/8. 4	8
11	11. 50	. 354	. 025	20/7. 1	8

Test. 6—*Triangle Discrimination*

Each figure consists of 4 parts. These are placed about a circle in 90° intervals, their apexes touching the circle. Three of the parts are equilateral triangles, measuring 1.5 inches on a side. The fourth part is identical with the others in all respects except that the side opposite the apex is bowed convexly. The convexly bowed side is constructed by drawing an arc of radius (*r*) from a point at distance (*d*) from the original line of the true triangle so that this part is equal in area to each of the other three parts. The direction of the convexly bowed side is distributed in chance order among the figures.

The following table gives the radius (*r*) of the arc for the convexly bowed side and the distance (*d*) from the edge of the true traingle to center of arc in inches. There are two figures for each row.

Row	r	d	r	d
(Chart A)				
1	1. 060	0. 9764	1. 403	1. 3394
2	1. 856	1. 8076	2. 456	2. 4193
3	3. 250	3. 2220	3. 614	3. 5885
4	4. 020	3. 9971	4. 470	4. 4494
5	4. 971	4. 9521	5. 529	5. 5114
(Chart B)				
1	6. 149	6. 1339	6. 838	6. 8233
2	7. 604	7. 5916	8. 457	8. 4448
3	9. 405	9. 3943	10. 459	10. 4495
4	11. 632	11. 6217	12. 937	12. 9281
5	14. 387	14. 3809	16. 000	15. 9923

Test 7.—*Bausch and Lomb Checkerboard*

This is the same test used in the Bausch and Lomb Orthorater increased in size for use at 20 feet. It consists of a square grouping of 4 squares, one of which contains a grid coarser than the other 3. The coarse grid is a checkerboard type made of rows of alternate black and white squares, while the grids of the other 3 squares in the group consist of rows of black dots. The square with the coarser grid varies in location from figure to figure in chance order. Each group decreases in size from top to bottom. The individual squares comprising the checkerboard grid decrease in size as follows:

Row	AGO unit	Visual angle of grid square (minutes)	Subtense of grid square	Snellen unit equivalent	Number of items
(Chart A)					
1	9. 00	2. 000	0. 140	20/40. 0	2
2	9. 25	1. 685	. 117	20/33. 7	4
3	9. 50	1. 416	. 099	20/28. 3	4
4	9. 75	1. 189	. 083	20/23. 8	4
(Chart B)					
5	10. 00	1. 000	. 070	20/20. 0	4
6	10. 25	. 842	. 059	20/16. 8	4
7	10. 50	. 708	. 049	20/14. 2	4
8	10. 75	. 595	. 042	20/11. 9	4
9	11. 00	. 500	. 035	20/10. 0	4
10	11. 25	. 421	. 029	20/8. 4	4
11	11. 50	. 354	. 025	20/7. 1	4

Test 8.—*AAF Letter Test*

There are 6 letters in each row equated for difficulty and of equal size and stroke width. The size of letters and the width of stroke decreases for succeeding rows from top to bottom. The specifications of the letters in each row are as follows:

Row	AGO unit	Visual angle of stroke (minutes)	Subtense of stroke (inches)	Snellen unit equivalent	Number of Items
1	9. 00	2. 000	0. 140	20/40. 0	6
2	9. 25	1. 685	. 117	20/33. 7	6
3	9. 50	1. 416	. 099	20/28. 3	6
4	9. 75	1. 189	. 083	20/23. 8	6
5	10. 00	1. 000	. 070	20/20. 0	6
6	10. 25	. 842	. 059	20/16. 8	6
7	10. 50	. 708	. 049	20/14. 2	6
8	10. 75	. 595	. 042	20/11. 9	6
9	11. 00	. 500	. 035	20/10. 0	6
10	11. 25	. 421	. 029	20/ 8. 4	6
11	11. 50	. 354	. 025	20/ 7. 1	6

TEST 9.—*Line Resolution*

A line of constant blackness and length extends from the center and points to 1 of 4 corners of a constant-sized white diamond on a field of uniform gray. In each row there are 4 diamonds, each containing a line of uniform width which points to a different corner of each diamond in chance order. There are 2 charts, chart B being a continuation of chart A. The width of the lines decreases from the top row down as indicated below.

Row	AGO unit	Visual angle of line width (minutes)	Line width (inches)	Snellen unit equivalent	Number of items
(Chart A)					
1	10. 00	1. 000	0. 0698	20/20. 0	4
2	10. 50	. 708	. 0494	20/14. 2	4
3	11. 00	. 500	. 0349	20/10. 0	4
4	11. 50	. 354	. 0247	20/7. 1	4
5	12. 00	. 250	. 0175	20/5. 0	4
6	12. 50	. 177	. 0124	20/3. 5	4
(Chart B)					
1	13. 00	0. 125	. 0087	20/2. 5	4
2	13. 50	. 088	. 0062	20/1. 8	4
3	14. 00	. 062	. 0044	20/1. 2	4
4	14. 50	. 044	. 0031	20/0. 9	4
5	15. 00	. 031	. 0022	20/0. 6	4
6	15. 50	. 022	. 0016	20/0. 4	4

TEST 10.—*Dot Variable Contrast*

Rows of uniformly gray diamonds of constant size are on a background of uniform white. In one corner of each diamond a gray dot of constant size, always darker than the gray of the diamond, is located. Contrast between dot and diamond decreases in each row from left to right and from the top row down. The dots are located in the corner of the diamonds in chance order.

Reflectance values for the dots and diamonds are as follows:

Items: dots (*dt*)	Reflectance R (percent)	Reflectance* difference ($R_{bg}-R_{dt}$)	Contrast $(R_{bg}-R_{dt})/R_{bg}$	Row
1	72. 0	1. 2	0. 016	1
2	71. 4	1. 8	. 025	1
3	70. 8	2. 4	. 033	1
4	69. 9	3. 3	. 045	2
5	68. 7	4. 5	. 061	2
6	67. 1	6. 1	. 083	2
7	65. 3	7. 9	. 108	3
8	63. 4	9. 8	. 134	3
9	61. 2	12. 0	. 164	3
10	57. 8	15. 4	. 211	4

*Reflectance of the background (*bg*): 73.2 percent.

TEST 11.—*AAG Constant Decrement*

The size of each letter and width of stroke decrease from letter to letter across each row and from top to bottom of the chart. Each letter decreases one Snellen step in size from left to right and top to bottom. Specifications for the size of letters are:

Item	AGO unit	Visual angle of stroke (minutes)	Subtense of of stroke (inches)	Snellen unit
(Row I)				
1	9. 00	2. 000	0. 140	20/40
2	9. 04	1. 950	. 136	20/39
3	9. 07	1. 900	. 133	20/38
(Row II)				
4	9. 11	1. 850	. 129	20/37
5	9. 15	1. 800	. 126	20/36
6	9. 19	1. 750	. 122	20/35
7	9. 23	1. 700	. 119	20/34
8	9. 28	1. 650	. 115	20/33
9	9. 32	1. 600	. 112	20/32
(Row III)				
10	9. 37	1. 550	. 108	20/31
11	9. 41	1. 500	. 105	20/30
12	9. 46	1. 450	. 101	20/29
13	9. 51	1. 400	. 098	20/28
14	9. 57	1. 350	. 094	20/27
15	9. 62	1. 300	. 091	20/26
16	9. 68	1. 250	. 087	20/25
17	9. 74	1. 200	. 084	20/24
(Row IV)				
18	9. 80	1. 150	. 080	20/23
19	9. 86	1. 100	. 076	20/22
20	9. 93	1. 050	. 073	20/21
21	10. 00	1. 000	. 070	20/20
22	10. 07	. 950	. 066	20/19
23	10. 15	. 900	. 063	20/18
24	10. 23	. 850	. 059	20/17
25	10. 32	. 800	. 056	20/16
26	10. 41	. 750	. 052	20/15
(Row V)				
27	10. 51	. 700	. 049	20/14
28	10. 62	. 650	. 045	20/13
29	10. 74	. 600	. 042	20/12
30	10. 86	. 550	. 038	20/11
31	11. 00	. 500	. 035	20/10
32	11. 15	. 450	. 031	20/9
33	11. 32	. 400	. 028	20/8
34	11. 52	. 350	. 024	20/7
35	11. 74	. 300	. 021	20/6
36	12. 00	. 250	. 018	20/5

Test 12.—*Square Discrimination*

The areas of all figures are equal to 1 square inch. Of the 6 figures in each row 1 is not a square. The sides of this odd figure are arcs of radius (r) drawn from a point at distance (d) from the line of the true square so that an equal area is maintained for all figures. The odd figure is distributed in chance order among the center 4 figures in each row. The following table gives the radius (r) of the arc for each odd (nondiamond) figure and the distance (d) from edge of the original square to center of arc in inches.

Row	r	d	Row	r	d
(Chart A)			(Chart B)		
1	0. 564	0. 4998	1	8. 572	8. 5657
2	1. 019	. 9822	2	9. 371	9. 3655
3	1. 840	1. 8188	3	10. 245	10. 2400
4	3. 323	3. 3107	4	11. 200	11. 1960
5	6. 000	5. 9928	5	12. 245	12. 2400
6	6. 560	6. 5528	6	13. 387	13. 3835
7	7. 172	7. 1631	7	14. 635	14. 6307
8	7. 841	7. 8350	8	16. 000	15. 9966

Test 13.—*Checkerboard Variable Grid*

This test uses the same figure as the Bausch and Lomb Checkerboard Test, except that the figure size remains constant while the individual squares making up the checkerboard grid of the odd square decrease in size for each succeeding row. The checkerboard grid is of contsant size for each of the 4 figures in a row. The location of the square containing the checkerboard grid is distributed in chance order among all the figures. There are two charts, chart B being a continuation of chart A. The sizes of the individual squares making up the checkerboard decrease as follows:

Row	AGO unit	Visual angle of grid square (minutes)	Subtense of grid square	Snellen unit equivalent	Number of items
(Chart A)					
1	9. 00	2. 000	0. 140	20/40. 0	4
2	9. 25	1. 685	. 117	20/33. 7	4
3	9. 50	1. 416	. 099	20/28. 3	4
4	9. 75	1. 189	. 083	20/23. 8	4
5	10. 00	1. 000	. 070	20/20. 0	4
6	10. 25	. 842	. 059	20/16. 8	4
(Chart B)					
7	10. 25	. 842	. 059	20/16. 8	4
8	10. 50	. 708	. 049	20/14. 2	4
9	10. 75	. 595	. 042	20/11. 9	4
10	11. 00	. 500	. 035	20/10. 0	4
11	11. 25	. 421	. 029	20/8. 4	4
12	11. 50	. 354	. 025	20/7. 1	4

Test 14.—*Vernier Acuity*

The opposite arms of the cross have a combined length of 2.8 inches and a width of line of 0.22 inch. A segment of 1 arm half the distance from center of the cross to end of arm is displaced in a clockwise direction. The amount of displacement in inches and the number of figures with the same displacement are given below. The figures are in order of diminishing amount of displacement, the displaced segment being located in chance order. Specifications are as follows:

Group	Row of the charts	AGO unit	Visual angle of displace (minutes)	Displace-ment (inches)	Number of items, total per row in ( )
(Chart A)					
1	1	8. 16	3. 583	0. 250	2 (2)
2	1, 2. 3	8. 34	3. 150	. 220	4 (1, 2, 1)
3	3, 4	9. 08	1. 891	. 132	4 (2, 2)
4	5, 6	9. 82	1. 131	. 079	4 (3, 1)
5	6, 1 of Chart B	10. 57	. 673	. 047	4 (1, 3)
(Chart B)					
6	2, 3	11. 32	. 400	. 028	4 (2, 2)
7	3, 4, 5	12. 04	. 243	. 017	4 (1, 2, 1)
8	5, 6	12. 81	. 143	. 010	4 (2, 2)

791639 48 - 11

Appendix B

ANSWER SHEET FOR 14 VISUAL ACUITY TESTS USED AT FORT DIX, NEW JERSEY

TO BE FILLED BY ATTENDANT

1. Name ______ Last ______ First ______ Initial ______ 2. ASN ______ 3. Glasses: Yes ______ No ______ (Reading Only ______ Habitual ______)
3. Age Last Birthday ______ 4. AGCT ______ 5. School Grades Completed ______ 6. Rest Previous Evening ______ hours
6. Station Color ______ 7. Previous Snellen (LEFT Eye Only) ______ 10. General Health: Excellent ______ Good ______ Average ______ Poor ______

SIDE 1 TEST LEFT EYE ONLY SIDE 1

PRACTICE TEST

Test 2 DOT VARIABLE SIZE CHART A / CHART B

Test 4 NEW LONDON LETTER

Test 5 TRIANGLE DISCRIMINATION CHART A / CHART B

Test 1 ARMY SNELLEN

Test 3 QUADRANT VARIABLE CONTRAST

BAUSCH & LOMB CHECKERBOARD CHART A / CHART B

WD AGO PRT-668

DO NOT WRITE IN THIS BOX

TO BE FILLED IN BY EXAMINER

11. Date of Test ______ Hour ______
12. Check One: First Test ______ Retest ______
13. Drugs Taken Last 24 Hours: Yes ______ No ______ Kind ______
14. Chart Center Foot Candles ______
15. Chart Number of First Test ______
16. Examiner's Name ______

SIDE 2 TEST LEFT EYE ONLY SIDE 2

Test 7 AAF LETTER

Test 8 LINE RESOLUTION CHART A / CHART B

Test 10 AAF CONSTANT DECREMENT

Test 12 CHECKERBOARD VARIABLE GRID CHART A / CHART B

Test 9 DOT VARIABLE CONTRAST

Test 11 SQUARE DISCRIMINATION CHART A / CHART B

Test 13 VERNIER ACUITY CHART A / CHART B

Appendix C

CODING SHEET FOR PERSONAL DATA AND VISUAL ACUITY TEST RESULTS

VISUAL ACUITY
4075

NAME ______________________________ ASN ______________________________
last first middle

1	CARD NO	(1-4)	7	5	0	1	26	(3)	DOT VAR SIZE	(35-36)	
2	MAN NO	(5-8)					27	(4)	QUAD VAR CONT	(37)	
3	ORIENTER	(9)					28	(5)	NEW LON	(38-39)	
4	EXAMINER	(10)					29	(6)	TRI DISC	(40)	
5	ASTIGMASTISM	(11)					30	(7)	B & L CHECK	(41-42)	
6	TEST LIKED MOST	(12)					31	(8)	AAF LETTER	(43-44)	
7	TEST LIKED LEAST	(13)					32	(9)	LINE RES	(45-46)	
8	EYES TIRED?	(14)					33	(10)	DOT VAR	(47)	
9	IRREGULARITIES	(15)					34	(11)	AAF CON	(48-49)	
10	GLASSES	(16)					35	(12)	SQ DISC	(50)	
11	AGE	(17)					36	(13)	CHECK VAR GRID	(51-52)	
12	AGCT	(18)					37	(14)	VERN ACUITY	(53-54)	
13	SCHOOL	(19)					38	(1)	PRACTICE	(55-56)	
14	REST	(20)					39	(2)	SNELLEN	(57-58)	
15	PREV SNELLEN	(21)					40	(3)	DOT VAR SIZE	(59-60)	
16	HEALTH	(22)					41	(4)	QUAD VAR CONT	(61)	
17	DATE	(23-24)					42	(5)	NEW LONDON	(62-63)	
18	HOUR	(25)					43	(6)	TRI DISC	(64)	
19	TEST-RETEST	(26)					44	(7)	B & L CHECK	(65-66)	
20	DRUGS	(27)					45	(8)	AAF LETTER	(67-68)	
21	FIRST TEST	(28)					46	(9)	LINE RESOL	(69-70)	
22	RACE	(29)					47	(10)	DOT VAR CONT	(71)	
23	BOOTH NO	(30)					48	(11)	AAF CONST DEC	(72-73)	
24	(1) PRACTICE	(31-32)					49	(12)	SQ DISC	(74)	
25	(2) SNELLEN	(33-34)					50	(13)	CHECK VAR GRID	(75-76)	
							51	(14)	VERN ACUITY	(77-78)	

Appendix D

INSTRUCTIONS FOR INDIVIDUAL TESTS USED IN THE FORT DIX, NEW JERSEY, VISUAL ACUITY STUDY

INSTRUCTIONS TO EXAMINER — Test 1

PRACTICE TEST

Say to Examinee

First, there will be a practice test to help you understand the test procedure. One chart will be used in this test.

Show Hand Chart 1

Here is a sample of the figures you will see on the chart. Notice that one corner of each figure is missing. You are to tell me which corner of each figure is missing by saying: top, bottom, right, or left. The missing corner gets harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row of figures. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready. Keep both eyes open.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 2

ARMY SNELLEN

Show Hand Chart 2 and Say

Here is a sample of letters you will see in this test. Call off each letter as you come to it. The letters get smaller and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 3

DOT VARIABLE SIZE (CHARTS A AND B)

Show Hand Chart 3 and Say

Here is a sample of what you will see in this test. Notice the dot in one corner of each diamond. Tell me where you see the dot in each diamond by saying: top, bottom, right, or left. The dot gets smaller and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There are two charts for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 4

QUADRANT VARIABLE CONTRAST

Show Hand Chart 4 and Say

Here is a sample of what you will see in this test. Each figure is like the head of an arrow. Tell me in which direction each arrow is pointing by saying: top, bottom, right, or left. The arrow gets dimmer and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 5

NEW LONDON LETTER

Show Hand Chart 5 and Say

Here is a sample of letters you will see in this test. Call off each letter as you come to it. The letters get smaller and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 6

TRIANGLE DISCRIMINATION (CHARTS A AND B)

Show Hand Chart 6 and Say

Here is a sample of what you will see in this test. Notice the 4 triangles in each figure. One triangle has a curved outer side. Tell me which triangle has a curved side by saying: top, bottom, right, or left. The curved side gets harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right.

There are two charts for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 7

BAUSCH AND LOMB CHECKERBOARD (CHARTS A AND B)

Show Hand Chart 7 and Say

Here is a sample of what you will see in this test. Notice the 4 small diamonds in each figure. One diamond is like a checkerboard. Tell me which is the checkerboard by saying: top, bottom, right, or left. The checkerboard gets harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There are two charts for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 8

AAF LETTER TEST

Show Hand Chart 8 and Say

Here is a sample of letters you will see in this test. Call off each letter as you come to it. The letters get smaller and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER — Test 9

LINE RESOLUTION (CHARTS A AND B)

Show Hand Chart 9 and Say

Here s a sample of what you will see in this test. Notice the line in one corner of each diamond. Tell me where you see the line in each diamond by saying: top, bottom, right, or left. The line gets smaller and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There are two charts for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER Test 10

DOT VARIABLE CONTRAST

Show Hand Chart 10 and Say

Here is a sample of what you will see in this test. Notice the dot in one corner of each diamond. Tell me where you see the dot in each diamond by saying: top, bottom, right, or left. The dot gets dimmer and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER Test 11

AAF CONSTANT DECREMENT

Show Hand Chart 11 and Say

Here is a sample of the letters you will see in this test. Call off each letter as you come to it. The letters get smaller and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There is one chart for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER Test 12

SQUARE DISCRIMINATION (CHARTS A AND B)

Show Hand Chart 12 and Say

Here is a sample of what you will see in this test. Notice the six figures in each row. For the test they are called 1, 2, 3, 4, 5, and 6. In each row, Nos. 1 and 6 are always diamonds. One of the others—2, 3, 4, or 5—is rounded. Tell me which figure is rounded, by saying: 2, 3, 4, or 5. The rounded figure gets more like a diamond and harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Go from row to row. There are two charts for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER Test 13

CHECKERBOARD VARIABLE GRID (CHARTS A AND B)

Show Hand Chart 13 and Say

Here is a sample of what you will see in this test. Notice the 4 small diamonds in each figure. One diamond is like a checkerboard. Tell me which is the checkerboard, by saying: top, bottom, right, or left. The checkerboard gets harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There are two charts for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.

INSTRUCTIONS TO EXAMINER Test 14

VERNIER ACUITY (CHARTS A AND B)

Show Hand Chart 14 and Say

Here is a sample of what you will see in this test. Notice that one arm of each cross is broken. Tell me which arm is broken, by saying: top, bottom, right, or left. The broken arm gets harder to see. Keep trying. Guess if you are not sure. Remember, start with the top row. Read from left to right. There are two charts for this test. Keep going until I say, "That is enough."

Are there any questions? Get ready.

Go ahead.