THE ATTEMPTED CONSTRUCTION AND VALIDATION

OF A TEST FOR SCIENTIFIC APTITUDE

By

Nancy Sharrington Bogl

Submitted as an Honors Paper
in the
Department of Psychology

THE WOMAN'S COLLEGE OF THE UNIVERSITY OF NORTH CAROLINA

GREENSBORO, NORTH CAROLINA

1958

THE ATTEMPTED CONSTRUCTION AND VALIDATION OF A TEST FOR SCIENTIFIC APTITUDE

By

Nancy Tharrington Boyd

Submitted as an Honors Paper
in the
Department of Psychology

THE WOMAN'S COLLEGE OF THE UNIVERSITY OF NORTH CAROLINA

GREENSBORO, NORTH CAROLINA

1958

Approved by

I. A. Burch, Director

Examining Committee

Kendon Smith

Victor M. Cutter, Jr.

ACKNOWLEDGMENTS

To the students and faculty who took part in this project, and without whom it could not have been done; to Dr. Victor Cutter, Jr., who helped administer the tests; and to Drs. I. A. Burch and Kendon Smith, my faculty advisors, who gave me valuable suggestions and encouragement; I wish to give my most sincere thanks.

CONTENTS

I. Introduction

Several test batteries have been fairly successful in predicting whether or not a student will succeed in science work in college, but these batteries make no claims other than that of predicting scholastic success or failure. They do not attempt to say whether the student will get along well in the world of science once he finishes school. Scholastic "scientific aptitude" and true scientific aptitude cannot be assumed to be the same, and there is an evident need for tests measuring the latter. But is there actually such a thing as scientific aptitude which can be isolated and measured and measured as an entity?

Lewis Terman did a study with a group of gifted children who were selected in 1921 and were followed closely through their careers to see if the ones who became scientists were different in interests, abilities, and personality from those who went into other fields.[1] Four detailed field studies were done on these persons: the studies of 1921, 1927, 1939, and 1950. There were all types of test data on them from about eleven years of age to about forty. From these individuals Terman selected those from the following areas for comparison:

1. physical sciences (research)
2. engineering
3. medical biological sciences
4. physical biological sciences (non-research)
5. social sciences
6. law
7. humanities

---

[1] Lewis M. Terman, "Are Scientists Different?" Scientific American, CXCII (January, 1955), pp. 25-29.

The members of the group who majored in social sciences were mostly in the field of business. Those who went into research, such as some of the economics majors, were not included: they did not fit into the category of businessmen, could not be considered as scientists, and were too few in number for statistical treatment as a separate group. Those used in this study were measured on scientific interests and abilities in eleven dimensions—some of which were measured in 1922 and some in 1940 after they were grown men. (Only men were used for this study of scientists, as there were only a few women out of the group of gifted children who became scientists.) They were also measured on fifteen factors of sociability. These measurements were taken from tests and ratings given in 1922, 1940, and 1950. The results showed that the science group scored significantly higher on science interest and lower on social traits than the businessmen-lawyer group or the humanities group. The businessmen and lawyers were the exact opposite and made the highest scores on sociability traits and scored lowest on science. The humanities group was in between these two groups on both traits. Of course, there was some overlapping of scores made by the scientists and the businessmen-lawyers; but most of the comparisons were statistically significant.

Terman concluded that differences do exist between scientists and other people; and furthermore, these differences were about as great when the persons were children as when they were adults and settled in their professions.

Another study, made by Anne Roe, has similarities with that done by Terman; and the results are essentially the same.[2] In addition to showing that there are differences in interests and abilities of scientists and non-scientists, Roe charted the occupations of the fathers of the scientists. Some of the fathers were scientists, but most came from other professions. Like Terman's group, these men were also interested in science when they were children.

Since there seemed to be some justification for attempting to construct a test for scientific aptitude, it was necessary to find a model for the new test. The test constructer can save himself a lot of trial-and-error guesswork if he has some sort of a model to go by. The model can be an empirical one patterned after other tests which have already had some degree of success, and/or it can be a theoretical model based on what the constructer thinks the test should be. In this case, it was decided that both types of model should be used. First, the theoretical model would be drawn up stating which traits were to be tested and defining those traits. After that was decided upon, other tests already in existence which claimed to test these traits would be examined. The best of these would be used as empirical models. For the experienced constructer who already knows much about making up test items, the empirical model would be perhaps unnecessary; he could make up all the test items entirely by himself. But in this case, the test constructer was quite inexperienced; and it seemed better that the items be taken from other tests. The fact that the questions came from other tests does not mean that the test would not be original; for though

---

[2]Anne Roe, "A Psychologist Examines 64 Eminent Scientists," Scientific American, CLXXXVII (November, 1952), pp. 21-25.

the items themselves would not be original, the particular combination in which they would appear would differ from that of any other test thus far written.

The purpose of the test constructed in this project was to differentiate between the scientist and the non-scientist. Yet, it is very difficult to construct a model for the ideal scientist; because in reality the individual requirements for going into different sciences and different scientific jobs differ greatly. Therefore, any one model could not possibly include all the qualities which might be necessary in a particular scientific job; it could only include those which would be essential for all scientists. Nevertheless, these are the traits selected as being indicative of scientific aptitude and necessary for success in the field of science:

1. Critical Thinking Ability

   The ideal scientist should be able to examine a situation thoroughly and critically before forming final conclusions. Being exact and cautious does not exclude originality, daring ideas, or taking chances. But the scientist should be aware of the chances he is taking and know their possible results. He should also know the difference between the situation in which he is following an accepted course and that in which he is introducing something unorthodox or new into his conclusions.

2. Ability to See New Relationships

   This would involve applying old principles to new things and devising new principles and uses for old things.

3. Mathematical Ability

Since mathematics is important in all branches of science, the scientist should have some knowledge of the subject and be able to deal with formulas and mathematical symbolism. A test of arithmetic, however, would not be adequate, as arithmetic is more of a clerical skill than a mathematical one.

4. Interest in Scientific Matters

If a scientist is to be successful, he must be interested in science as well as possessed of scientific abilities. This means that he probably does a lot of reading in scientific areas, and it will be evident in a good scientific vocabulary.

5. Creative Ability

It is difficult to define creative ability and even more difficult to test it. Unfortunately, there is no really satisfactory test for creative ability on the market as yet; and this being the case, it was decided not to try to test this trait, except in the sense implied by "ability to see new relationships" discussed above.

With the foregoing considerations in mind, we now briefly review pertinent tests already published.

II. Tests of Scientific Aptitude Already Published

One of the first attempts to develop a test of scientific aptitude resulted in the Stanford Scientific Aptitude Test by D. L. Zyve. This test is composed of eleven sub-tests considered to be measures of components of scientific aptitude. The hypothetical components are experimental bent; clarity of definition; suspended versus snap judgment; reasoning; recognition of inconsistencies; fallacies; induction, deduction, and generalization; caution and thoroughness; discrimination of values in selecting and arranging experimental data; accuracy of interpretation; and accuracy of observation. This test is intended for high school juniors and seniors and for college students. It is a power test, i. e., one with no time limit. Most people finish the test in one to two hours.[3]

The scores on the test taken by fifty research students at Stanford University had a correlation of .74 with ratings made by faculty members of their scientific aptitude. However, later investigations have failed to confirm the suggested high validity of the test. Studies at Yale and at the College of the City of New York found correlations with science grades to be in the .30's. They reported a reliability of only .60 for the entire test.[4]

Even though the validity of the Stanford Scientific Aptitude Test has failed to be confirmed, the test is still an intriguing one. One might wonder why it has not been revised and properly standardized. A further look at this test will show that revising it would be more trouble than creating a new test "from scratch." The test as it stands now has no

---

[3]David L. Zyve, Explanatory Booklet for the Stanford Scientific Aptitude Test, pp. 1-16.

[4]Anne Anastasi, Psychological Testing, p. 515.

separate answer sheet and must be hand-scored. In order to convert the test to the machine-scoring type, it would be necessary to reconstruct the entire test using different items for most of the sub-tests. Many of the sub-tests do not appear always to be testing what their titles would lead one to think.

Milton Mandell devised a battery of tests which was designed to measure originality in the physical sciences.[5] This battery was made up of the following tests: figure analogies, Gottschaldt figures,[6] spatial relations tests, formulation, letter series, table reading, vocabulary, interpretation of data, hypotheses, scrambled sentences, and scientific subject matter. This study was done for the U. S. Civil Service Commission for purposes of personnel selection. There were three comparisons made: (1) correlation of test scores with ability in basic research, (2) relationship of test scores to ratings on originality, and (3) critical ratios between research and non-research groups. Unfortunately, most of the correlations were close to zero. The author felt this outcome may have been due to the small number of items in each test. Since this was intended to be only a preliminary study, he felt that many different types of tests should be used. In order to persuade men to take the test, the total time had to be kept fairly short, and each sub-test was therefore very brief.

The two tests which had the most significant results were formulation and scientific subject matter. Formulation was good in differentiating research from non-research personnel. Subject matter tests correlated highly with ratings on originality in physicists, chemists, and engineers.

---

[5] Milton M. Mandell, "Measuring Originality in the Physical Sciences," Educational and Psychological Measurement, X, (1950), pp. 380-85.

[6] Gottschaldt figures are representations in which familiar figures are concealed within more complex figures. The task is to find the hidden figure.

Another test for scientific aptitude which has been more widely used and better standardized than either of the two tests discussed is the Engineering and Physical Science Aptitude Test.[7] This test is made up of six individually timed sub-tests: mathematics, formulation, physical science comprehension, arithmetic reasoning, verbal comprehension, and mechanical comprehension. It is essentially a speed rather than a power test, intended for use at the college level.[8] The whole test requires approximately seventy-five minutes to take.

George Gregg gave the EPSAT to 352 freshmen entering the College of Engineering of the University of Colorado in the fall of 1948.[9] A year later, he found the correlations between the various sub-scores of the test and grades relating to science subjects. The directions of the manual suggest a corrective formula to cancel out the effects of guessing on the test.[10] In this study, one validation of the test was done using the corrective formula and another validation using just the raw scores. When the scores were corrected for guessing, the $\underline{r}$ coefficient was .63; when the scores were not corrected for guessing, it was .58. Though this difference is small, it does seem to justify the use of the corrective formula.

Griffin and Borow found the coefficient of correlation between scores on EPSAT and science grades to be .737 when they gave the test to 188 engineering students at Pennsylvania State College.[11]

[7]This test will hereafter be referred to as EPSAT.

[8]Bruce V. Moore, C. J. Lapp, and Charles H. Griffin, Manual for the Engineering and Physical Science Aptitude Test, p. 3.

[9]George W. Gregg, "An Investigation of the Reliability and Validity of the Engineering and Physical Science Aptitude Test," Journal of Educational Research, XLV, (December, 1951), pp. 299-305.

[10]Moore, Lapp, and Griffin, op. cit. p. 5.

[11]C. H. Griffin and H. Borow, "An Engineering and Physical Science Aptitude Test," Journal of Applied Psychology, XXVI, (October, 1944), pp. 376-387.

Not all studies done with EPSAT, however, have gotten such favorable results. Cooprider and Laslette did a study comparing science grades with scores on EPSAT, the Stanford Scientific Aptitude Test, the American Council on Education Psychological Examination,[12] and the Ohio State University Psychological Test.[13] For subjects they used 376 male engineering students who had completed two to six quarters of college work.

Brief descriptions of EPSAT and the Stanford Scientific Aptitude Test have already been given. The ACE examination is designed to appraise the scholastic aptitude of students entering college.[14] The manual claims that linguistic tests give high correlations with scholarship in the liberal arts colleges and that quantitative tests do a better job of predicting success in scientific and technical schools. The ACE examination has been designed to take care of both. It has three quantitative sub-tests, which combined give a Q-score; and it has three linguistic sub-tests, which give an L-score. The quantitative sub-tests are arithmetical reasoning, number series, and figure analogies. The linguistic sub-tests are same-opposite, completion, and verbal analogies. The scores of all six sub-tests can be combined to give a total or composite score. Each sub-test is timed separately, and the entire test requires approximately an hour to administer. The ACE examination is essentially a speed test rather than a power test.

---

[12]This will hereafter be referred to as the ACE examination.

[13]H. A. Cooprider and H. R. Laslette, "Predicitive Values of the Stanford Scientific and the Engineering and Physical Science Aptitude Tests," Educational and Psychological Measurement, VIII, (November, 1948), pp. 683-87.

[14]Manual of Instructions for American Council on Education Psychological Examination, p. 2.

The Ohio State University Psychological Test is made up of three sub-tests. They are same-opposites, word relationships, and reading comprehension. There is no time limit.[15]

Cooprider and Laslette's study showed that the Q-score and the T-score of the ACE examination, and scores from EPSAT, and the Stanford Scientific Aptitude Test are all about equal in their ability to predict grades in science.[16] The correlations were disappointingly low, however; the Q-score and the T-score had correlations with science grades of .40 and .39 respectively. EPSAT's correlation was .39, and the Stanford Scientific Aptitude Test correlated .37 with science grades. The L-score of the ACE examination and the Ohio State University Psychological Test, as might be expected, had much lower correlations with grades in science. The conclusion that Cooprider and Laslette draw is that the tests specifically designed to test scientific aptitude do not do any better job of it than do tests which are designed to measure general scholastic aptitude.

Although the correlations with science grades of EPSAT, ACE, and the Stanford Scientific Aptitude Test were low, it was nevertheless decided that these tests are the best predictors of scientific aptitude on the market. These instruments were therefore used as models for the new test.

---

[15]George A. Ferguson, The Fourth Mental Measurements Yearbook, p. 310.

[16]Cooprider and Laslette, op. cit., pp. 685-87.

III. Preparation of the New Test

With the old tests and the traits desirable for the ideal scientist in mind, (pages 4-5), the new test was written.

To test mathematical ability the mathematics test and the formulation test were selected from EPSAT. While both of these tests are related to mathematics, they measure two entirely different skills. The mathematics test is composed of simple algebraic problems and requires a knowledge of high school algebra. The formulation test does not require any knowledge of mathematics, but calls for an ability to express a formula in mathematical symbolism. The formulation test was felt to be especially important, as it is not so much influenced by training in mathematics as the mathematics test. It would be just about impossible to make up a test which would not be influenced by previous training, but it is certainly desirable that an aptitude test be as little influenced by educational training as possible.

To measure interest in science two other tests, also from EPSAT, were selected. They were physical science comprehension, a true-false quiz on well-known scientific facts, and verbal comprehension, a vocabulary test containing chiefly words used in connection with scientific matters. It might be asked why these two tests were chosen to measure interest in science, for they seem to be testing a <u>knowledge</u> of science rather than just interest. The answer is that they probably measure both. A person with scientific ability who is also interested in science will voluntarily expose himself to general information concerning science and will acquire a scientific vocabulary. It is also true that a person not interested in science but who has had sufficient training in it would also pick up this type of information and vocabulary. Therefore, a high score on these two tests would not necessarily be indicative of high interest in science; but a low score would be very significant. It would either mean that the person was not interested in science, or that, if he were interested, he lacked sufficient ability to be successful in it.

One test was chosen from the ACE examination. This was the number series test. It was felt that this test was appropriate for measuring the ability to see relationships. A spatial relationship test might have been desirable, too; but it was impractical to use a measure of that sort because of the difficulty in printing the elaborate drawings.

No tests were used from the Stanford Scientific Aptitude Test. The form of the test was such that it could not have a separate answer sheet which could be machine scored. Instead of the Stanford Scientific Aptitude Test, sub-tests from the Watson-Glaser Critical Thinking Appraisal were used. This test, although not primarily designed as a scientific aptitude measure, actually correlates very highly with the Stanford Scientific Aptitude Test.[17] The two instruments appear to be very similar in the type of questions asked. The Watson-Glaser test has the advantage of being a much more recent test[18] with more up-to-date questions, and it can be machine scored. The authors emphasize that their test is not in any way related to an intelligence measure. They say that many people with superior mental ability do poorly on this test, but that those who do well on it are also very likely to score well on an intelligence test.[19]

---

[17]Anastasi, op. cit., p. 515.

[18]The Stanford Scientific Aptitude Test was published in 1929 and 1930; the Watson-Glaser Critical Thinking Test was published in 1952.

[19]Goodwin Watson and Edward Maynard Glaser, Watson-Glaser Critical Thinking Appraisal, p. 9.

One point in favor of using parts of the Watson-Glaser Critical Thinking Appraisal is that it is very different from an achievement test. Its sub-tests are inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. Of these five sub-tests, inference and recognition of assumptions appeared to be the best for measuring critical thinking ability; and they were selected to be included in the new instrument.

These seven sub-tests, then, made up the battery for the new scientific aptitude test: inference, recognition of assumptions, mathematics, formulation, physical science comprehension, verbal comprehension, and number series. While all of the questions were taken directly from the sub-tests which have been previously mentioned, the new sub-tests are not identical with the sub-tests that appear in the parent tests. In order to get in several different types of measures in a short period of time, all of the original sub-tests were greatly shortened; and only some of their questions appear in the new instrument. This means that, although the sub-test items are not original, the sub-tests are. The new sub-tests will have to be validated just as though their items were entirely new, for the validation data on the original tests cannot possibly apply to certain isolated pieces from them.

Each of the seven sub-tests in individually timed. One unusual feature is that the reading of the directions for each sub-test is part of that test. The person administering the examination does not go over the directions with the group before each of the sub-tests; each person is responsible for understanding them himself. This is quite different from the procedure followed on other tests. Usually the examiner reads the directions aloud, and numerous sample problems are done before the person is allowed to proceed with the test. For the new test, that procedure was considered to be unnecessary. The scientist should be able to follow directions accurately; and if the printed directions are clear, there should be no need of drilling.

The entire test required forty-five minutes to take. This meant that it was suitable to give in one class period. A copy of this new test is included in the appendix.

IV. Sampling Procedure

The ideal way to validate a test of scientific aptitude would be to examine thousands of high school seniors or college freshmen and then wait about twenty years to see which ones actually become successful scientists. It would be simple to check whether a correlation then existed between scores on the test and success in science. Unfortunately, in a one-semester Honors Project that procedure is not possible.

The next best thing would be to find a group of persons who were already successful scientists and see if they made significantly higher scores than those who were not scientists. This, too, was not possible to do in the short time available. In order to select a group of ideal scientists, it would be necessary to do extensive research, studying the records of each person considered. Even if there were time to do this, there would still be the problem of getting them all together, for they would come from all parts of the country.

The third best thing to do was to use a group of people who were in the field of science but who were not necessarily "ideal" scientists. This method would eliminate the time-consuming case study of each person to see if he qualified. Merely holding a job in science could be the criterion for the scientist. The group which was most accessible were the science faculty members of Woman's College of the University of North Carolina. All of them probably would not meet the rigid requirements for the ideal scientist, and this circumstance would tend to make the results less impressive. But assuming there is such a thing as scientific aptitude which can be measured, it seemed logical to assume that science faculty members would possess more of it than groups outside the sciences and that there should be some overall differences between the two groups. The group that would be the best for comparison with science faculty members would be composed of non-science faculty members of the Woman's College of the University of North Carolina, inasmuch as both groups would be on approximately the same educational and intellectual level.

Not only would college faculty members be tested, but also students. The students would be treated very much like the faculty members. There would be two groups, one of science majors and one of non-science majors. Science-majoring students are certainly not ideal scientists; but as a group, they should have some scientific aptitude. It would be expected that the science majors would do better than the non-science majors and that the science faculty members would do better than the non-science faculty members. The students would also be compared with the faculty members; and it was quite uncertain as to how that would come out, especially the science students in comparison with non-science faculty members.

All juniors and seniors who were majoring in a science subject were asked to take the test. While not every student who was asked was able to take the test, all the science departments were represented in the sample. The biology and psychology departments, however, were more heavily represented in the sample than the other science departments. Students who were taking Psychology 221 (Introductory Psychology) were used for the non-science group of students. This is a course required for all students getting a teaching certificate, and all types of majors enroll in it. Science majors who were in this class were not asked to take the test over again if they had taken it with the science majors. However, if they had not taken the test before, they did take it; and their scores were, of course, counted in with the other science majors.

All the science faculty members at Woman's College of the University of North Carolina were asked to take the test. There were thirty-five of these in all. A selective random sample was drawn of the remaining faculty for the non-science group. It was necessary to make the random sample selective because, as some departments were larger than others, a purely random sample would have biased the sample towards those departments. After a random sample was drawn, some of the people from the larger departments were randomly marked off, thus making all departments as equally represented as possible.

One of the main problems connected with giving the test to faculty members was persuading them to take the test. Practically none would consent to take a test that required much time, and so the test had to be made short. The test, as was mentioned earlier, required only forty-five minutes and could easily be taken in one class period. A letter asking for the cooperation of the faculty members in taking the test was sent to all the members of the science faculty and to those selected in the random sample from non-science departments. There was no way of knowing how many would accept the invitation to take the test; but it seemed likely that a higher percentage of the science faculty would be willing than the non-science faculty, since the test was concerned with detecting scientific aptitude. Therefore, more letters were sent out to the non-science faculty than to the science faculty. This was done in order to make the two groups approximately equal in number. In all there were one hundred letters sent out, thirty-five to the science faculty and sixty-five to the non-science faculty.

The student science majors were urged by the heads of their departments to take the test. However, there was no great pressure put on them; and those who took the test did so entirely of their own will with the exception of the biology students. The biology students were asked by their department head to come to a meeting, the purpose of which was unknown to them. At the meeting they were given the test. This probably accounts for the large number of biology students taking the test.

The non-science students were the easiest of all to "induce" to take the test, as they did not have much choice about it. They were given the test during their regular Psychology 221 class periods.

V. Administering the Test

The test was extremely easy to administer. Each group was told: "You are about to take a scientific aptitude test. It will require only forty-five minutes. The test is made up of seven sub-tests, and each sub-test is timed separately. Use only the time allotted for each test. If you finish a test before time is called, do not look back to a test already taken or go on to a new one. If you have any figuring to do on paper, please use the answer sheet rather than the test booklet. When you get to a question that you do not know the answer to, guess if your guess is based on something; but do not guess randomly, for you will be penalized somewhat for your wrong answers. You are responsible for reading the directions for each test yourself. When I say 'Begin,' turn over your test booklets and read the directions for Test 1. After you have finished reading the directions, begin immediately working on the problems. Are there any questions?"

That was all the test administrator had to do except to time each test. Occasionally someone would have difficulty in reading a question due to poor printing of the test. In that case, the person giving the test would read the question for him. But if anyone had trouble understanding what the question meant, the test administrator was not allowed to interpret the question.

It was not possible for the same person to give all of the tests. In some cases a student administered the test, and in others a faculty member. However, the lack of uniformity in administrators should not have affected the results seriously since the test was virtually self-administering.

The printing was the biggest problem in administering this test. The tests were run off by hand on a duplicating machine which prints rather inconsistently. Although only the best copies were used, some of them are in places faint and "fuzzy." (Note page 2 of the test in the appendix.)

VI. Results

Fifty-four students majoring in science took the test. More than fifty-four non-science students took it; but, to keep the groups equal, only fifty-four tests from the non-science group were randomly selected to be checked. It is not necessary for the groups to be equal, but it is somewhat easier to compare then statistically if they are. Fifteen science faculty members took the test, and eighteen non-science faculty took it. Since both of these groups were so small, there was no attempt made to equate them.

The members of the science faculty were called Group 1, the members of the non-science faculty, Group 2; Group 3 were the science students, and Group 4, the non-science students. When there are four groups, it is possible to make six comparisons: Groups 1 and 2, Groups 1 and 3, Groups 1 and 4, Groups 2 and 3, Groups 2 and 4, and Groups 3 and 4. To compare the mean score of one group with that of another, t ratios were used.[20] The t value must at least meet the 5% level of significance in order for the difference between groups to be considered non-accidental.

The results of the comparisons are shown in Table I and in Figures 1-7. These data are discussed in the following section of this report.

---

[20] A t test is a statistical procedure designed especially for comparing small groups. Two-tailed tests were used instead of one-tailed tests, as it was not certain in which direction the differences in scores would fall.

TABLE I

Analysis of the differences between means

TEST	t VALUE	Significant?	If sig., is diff. in the expected direction?
TEST I Inference			
Groups 1 and 2	.27	no	
" 1 and 3	.0015	no	
" 1 and 4	.39	no	
" 2 and 3	.23	no	
" 2 and 4	.72	no	
" 3 and 4	.57	no	
TEST II Recognition of Assumptions			
Groups 1 and 2	.28	no	
" 1 and 3	1.53	no	
" 1 and 4	.84	no	
" 2 and 3	2.31	yes	no
" 2 and 4	1.49	no	
" 3 and 4	.90	no	
TEST III Mathematics			
Groups 1 and 2	2.33	yes	yes
" 1 and 3	2.06	yes	yes
" 1 and 4	5.33	yes*[21]	yes
" 2 and 3	.96	no	
" 2 and 4	2.30	yes	yes
" 3 and 4	4.67	yes*	yes
TEST IV Formulation			
Groups 1 and 2	1.13	no	
" 1 and 3	.75	no	
" 1 and 4	1.97	no	
" 2 and 3	.58	no	
" 2 and 4	.47	no	
" 3 and 4	1.42	no	

[21] The * indicates that the difference was great enough to meet the 1% level of significance.

TABLE I (Continued)

TEST	t VALUE	Significant?	If sig., is diff. in the expected direction?
TEST V			
Physical Science Comprehension			
Groups 1 and 2	.75	no	
" 1 and 3	3.22	yes	yes
" 1 and 4	6.14	yes*	yes
" 2 and 3	2.68	yes	no
" 2 and 4	5.87	yes*	yes
" 3 and 4	2.30	yes	yes
TEST VI			
Verbal Comprehension			
Groups 1 and 2	1.45	no	
" 1 and 3	5.77	yes*	yes
" 1 and 4	7.58	yes*	yes
" 2 and 3	5.02	yes*	no
" 2 and 4	6.17	yes*	yes
" 3 and 4	2.81	yes	yes
TEST VII			
Number Series			
Groups 1 and 2	1.19	no	
" 1 and 3	.45	no	
" 1 and 4	.85	no	
" 2 and 3	1.16	no	
" 2 and 4	2.86	yes	yes
" 3 and 4	2.14	yes	yes

Even though there was a significant difference between some of the groups, there was a great deal of overlapping of individual scores. This is made obvious by the frequency distributions of the raw scores on pages 24-30.

Figure 1

TEST I INFERENCE

Distribution of Scores by Groups, with Group Means ($\bar{X}$)

	$\bar{X}_1 = 5.766$	$\bar{X}_2 = 6.0$	$\bar{X}_3 = 5.81$	$\bar{X}_4 = 5.50$
0		✓		
.5				^
1.0			oo	^^^^^^
1.5				
2.0	+	✓	ooooo	
2.5				^
3.0	++		ooo	^^
3.5			ooooooo	^^^^
4.0		✓		^
4.5	+		oo	^
5.0	++	✓✓✓✓✓✓	oooooooooo	^^^^^^^^^
5.5	+		o	
6.0	++	✓✓✓	oooooo	^^^^^^^
6.5			o	^^
7.0	+++	✓	ooooo	^^^^^^^^^^ ^^
7.5		✓		
8.0	+			^^
8.5	+	✓	oo	^^^^
9.0	+		o	^
9.5				
10.0		✓✓	ooooo	^
10.5				
11.0		✓	o	^
11.5				
12.0			oo	
12.5				
13.0				
13.5				
14.0				
14.5				
15.0			o	

Figure 2

TEST II RECOGNITION OF ASSUMPTIONS

Distribution of Scores by Groups, with Group Means

	$\bar{X}_1$ 9.13	$\bar{X}_2$ 9.50	$\bar{X}_3$ 7.54	$\bar{X}_4$ 7.98
0	+		ooooooooooo	^^^
1				^
2	+		o	^^^^
3			oo	
4		✓✓	oo	^^^^^
5				
6	+	✓✓	oooo	^^^^
7	+		o	
8	++	✓✓✓✓✓	oooooooooo	^^^^^^^^^^^^
9			o	
10	++++	✓✓✓	ooooooooooooooo	^^^^^^^^^^^^^
11				
12	++	✓	oooo	^^^^^^^^^^
13		✓		
14	+++	✓✓✓✓	ooo	^^^

Figure 3

TEST III MATHEMATICS

Distribution of Scores by Groups, with Group Means

	$\bar{X}_1$ 8.43	$\bar{X}_2$ 5.56	$\bar{X}_3$ 6.43	$\bar{X}_4$ 3.72
0	+	✓	ooo	^^^^^^^
.5			o	^^
1.0		✓✓		^^^^^
1.5				^^
2.0			oooo	^^^
2.5				^^^
3.0	+	✓	o	^^
3.5				^^
4.0	+	✓✓✓✓	ooo	^^^^^^^
4.5			oooooo	^^
5.0		✓	oo	^^^^
5.5		✓	ooo	
6.0		✓✓	oo	^^^^^
6.5			oo	^^
7.0	+		ooooo	^^
7.5	++		ooo	^
8.0	+	✓✓	oooo	^^
8.5		✓	o	
9.0			ooo	^
9.5	+		ooo	^
10.0	+	✓✓		
10.5			oo	^
11.0	++		oo	
11.5				
12.0	++++	✓	oooo	

Figure 4

TEST IV FORMULATION

Distribution of Scores by Groups, with Group Means

	$\bar{X}_1$ 2.52	$\bar{X}_2$ 1.93	$\bar{X}_3$ 2.13	$\bar{X}_4$ 1.79
0.00	++	✓✓	oooooo	^^^^^
0.25			oo	^^
0.50			oooo	^^
0.75	+	✓✓	oooo	^^^
1.00			o	^^
1.25	+	✓✓	oooo	^^^^^
1.50	++	✓✓✓	o	^^^^^^
1.75	++		oo	^^^^^^^^^
2.00		✓✓	ooo	^
2.25				
2.50		✓✓	oooo	^^^^^^^
2.75	+	✓	ooo	^^^^
3.00		✓	oooo	^^
3.25			oo	^
3.50				
3.75	++	✓✓	oooooooo	^^
4.00	+	✓	ooo	^^^
4.25				
4.50				
4.75				
5.00	+++		ooo	^

Figure 5

TEST V PHYSICAL SCIENCE COMPREHENSION

Distribution of Scores by Groups, with Group Means

	$\bar{X}_1$ 9.40	$\bar{X}_2$ 8.11	$\bar{X}_3$ 4.48	$\bar{X}_4$ 2.59
0	+		oooooo oooooooooooo	^^^^^^^^^^^^^ ^^^^^^^^^^^^^
1			o	^^
2	+	✓	oooooooooo	^^^^
3		✓	oooo	^^^^^
4	++	✓✓✓	o	^^^^^^^^
5	+	✓✓		^^
6		✓	oooo	^^
7	+	✓		^
8	+	✓	oooo	^^^
9	+	✓✓		^
10		✓	ooo	
11				^
12	++		oo	^
13		✓✓	o	
14	++	✓	o	
15		✓✓		
16	++		oooo	
17				
18	+			

Figure 6

TEST VI VERBAL COMPREHENSION

Distribution of Scores by Groups, with Group Means

Score	$\bar{X}_1$ 18.70	$\bar{X}_2$ 17.47	$\bar{X}_3$ 12.35	$\bar{X}_4$ 10.13
0			o	
.5				
1.0				
1.5				
2.0				^^
2.5				
3.0				^^^
3.5				^
4.0			o	^^
4.5				
5.0			o	^
5.5				
6.0				^^
6.5				
7.0			oo	^
7.5				^^
8.0			oooo	^^^
8.5			o	^^
9.0			o	^^^
9.5			oooo	^^
10.0			oo	^^
10.5				^
11.0			ooooo	^^^^
11.5			o	
12.0			oo	^^^^^^^
12.5			o	
13.0		✓	ooooooo	^^^^
13.5		✓		^
14.0	+		ooo	
14.5	+		o	
15.0		✓✓	o	^^^^^^
15.5				
16.0		✓	oooooo	^^^
16.5	+			
17.0		✓✓	ooooooo	^^
17.5	++	✓✓		
18.0	++	✓✓	o	
18.5		✓✓	o	
19.0		✓✓		
19.5	+++		o	
20.0				
20.5				
21.0		✓		
21.5	++++	✓✓		
22.0	+			

Figure 7

TEST VII NUMBER SERIES

Distribution of Scores by Groups, with Group Means

	$\bar{X}_1$ 7.93	$\bar{X}_2$ 9.22	$\bar{X}_3$ 8.35	$\bar{X}_4$ 7.16
0	+		oo	^
.5				
1.0				
1.5	+			^^
2.0			o	^^^
2.5				
3.0	+		o	^
3.5				
4.0			o	^^
4.5		✓		^
5.0	+		o	^
5.5			oo	^^
6.0	+		oo	^^^^^
6.5			ooo	^
7.0	+	✓✓	oooo	^^^^^^
7.5		✓		^
8.0		✓✓✓✓✓	oooooooooooo	^^^^^^^^^^^
8.5				^
9.0	+	✓	oooo	^^^^^
9.5	+		o	
10.0	++	✓✓	ooo	^^^^^
10.5				
11.0	++	✓✓✓✓✓	ooooooooo	^^^^
11.5				
12.0	+++	✓✓	oooooooo	^^

VII. Interpretation of Results

The comparisons between Groups 1 and 2, the science faculty and the non-science faculty, were disappointing. Out of the seven comparisons between these two groups, only one was significant, that of Test III, the mathematics test. This is the test which would be most influenced by previous training, and it would be a strange test indeed if the science faculty did not do better on it than the non-science faculty. The other six tests came out with the differences between the two groups so slight that they had to be attributed to chance. In fact, on three of the tests the non-science faculty did slightly better than the science faculty.

The comparisons between Groups 3 and 4, science students and non-science students, are a little more encouraging. In four out of the seven tests, the science students made significantly better scores. These tests were mathematics, physical science comprehension, verbal comprehension, and number series. However, with the exception of number series, these tests are the ones which would be most likely to be affected by training as opposed to aptitude. The comparisons between these two groups, while more hopeful than between the faculty groups, are less significant than they might appear at first sight.

Groups 2 and 3, non-science faculty and science students, were compared mainly to see what made the most difference, additional education on the part of the faculty members or aptitude on the part of the students. Education seemed to influence the results the most. The non-science faculty members had higher mean scores on five of the tests, and the difference was a significant one on three of these tests. The science students had higher mean scores on two of the tests, and in neither case was the difference significant. This is discouraging, for it seems to indicate that the general educational level of the person or scholastic aptitude influences the results of the test more than scientific aptitude. If this be the case, then the significant differences between the science faculty and non-science students are pretty well accounted for.

Does this mean then that the whole project was a failure because the test seems to be just another scholastic aptitude test and not a scientific aptitude test? Not at all. Admittedly, the test in its present form is not much good for detecting scientific aptitude. But that does not mean that all the effort spent was wasted.

When a test is constructed and validated for the first time by an inexperienced person, mistakes are bound to be made. However, the person conducting the study should and usually does recognize many of his mistakes as such and gains new insight into the situation. Almost as soon as the project gets under way, he starts making plans for how it could be improved should he ever have the opportunity to do it again. Many do have the opportunity and do an experiment or a study several times before coming up with one which they feel is worthy of being published. While it is not likely that this project will be repeated any time soon, it seemed helpful to note some of the main shortcomings of it.

All the sub-tests were too short. They were deliberately made short so that as many different types of tests could be included in as short a testing period as possible. However, this meant sacrificing some of the tests' validity. The science faculty made a higher mean score on the formulation test than did the non-science faculty. The science students did better on it than the non-science students. This test, however, had only five questions; and the differences between any two groups were too slight to be statistically significant. The test on formulation was the shortest in the whole group, but none of the tests were more than half as long as the original tests from which they were taken.

The sample of people tested, especially the faculty members, was too small. There were only fifteen science faculty members in the sample and eighteen non-science faculty members. Despite the fact that the science group did have in common the fact that they were all in the field of science, it was still a very heterogeneous group. In those fifteen people were represented both sexes, all the different sciences, and wide age variation. To offset the differences within such a varied group, a sample of several hundred should have been used. The eighteen non-science faculty members were, on the other hand, not as heterogeneous as might have been desired. Efforts were made to invite about the same number from each department so that they would be equally represented. However, the faculty members did not accept in the same proportions in which they were invited. For some reason, the history department was very heavily represented; at least five out of the eighteen were from the history department.

It would not have hurt for the student group to have been larger, too; but the size of this group is not nearly so inadequate as that of the faculty group. Fifty-four is not a bad number considering the fact that they were all of the same sex and within two years of being the same age.

There were a few typographical errors in the test which had to be corrected, and in some cases the printing was distractingly poor. It is difficult to say what effect this had on the results of the test, but the test would certainly be better standardized if the printing were uniform on all the copies and if there were no mistakes to stumble over.

The criterion for the science group was a fair one for a preliminary study. Yet, just because a person is in the field of science is really no indication that it is what he is best suited for. In validating a test for scientific aptitude, one should do it with people that are known to possess it. In this case, we had a test which we thought would test scientific aptitude; and we tested a group which we thought would possess scientific aptitude. If the results did not come out favorably, as they did not, it could not be determined whether the test lacked ability to detect scientific aptitude or the group tested lacked the aptitude. In the case of the students, many were sophomores who had only selected their majors tentatively and might possibly change their minds or go into some other field after college. There is then even less certainty that the science students had scientific aptitude than the faculty.

If the project could be done again, then, these are the ways in which it would be changed:

There would be three groups of people tested: experts in the field of science, experts in fields other than science, and ordinary people chosen randomly from the population. There would be a definite criterion by which to judge whether a person qualified as an expert or not, and this criterion would be a rather rigid one. As it would be difficult to get all the eminent people together, the test would be mailed to them. With a timer, the test could be self-administered. This would mean that people from remote sections of the country would not be passed up because of their inaccessibility. There would be at least 500 people in each group.

The test itself would be greatly changed. The sub-test on inference and the sub-test on recognition of assumptions would be dropped. These two tests look as though they might correlate with scientific aptitude; but in this project, they did not come even close. In fact, the non-science faculty made slightly better scores on these two tests than did the science faculty. The mathematics test and the physical science comprehension test would also be omitted, because they are too much influenced by previous training. A vocabulary test would be used, but it would be a different one from the one in this test. There would be no attempt to have this test slanted towards scientific words. Instead, it would be just a regular vocabulary test that would have nothing directly to do with scientific aptitude. Vocabulary tests which have high correlations with intelligence tests are frequently used in the place of the latter to measure general intellectual ability. By use of a vocabulary test, it might be possible to establish cut-off points showing at which levels a person could operate as a scientist if his aptitude be in that direction. The formulation test would be left in, but it would be made much longer. Instead of having only five questions, the new test would have twenty or twenty-five. The number series test would also be left in; and it, too, would be lengthened. On that test, there would be an effort made to have some of the items much more difficult than any of those appearing on the test used in this project. Too many people made perfect scores for it to be a good measure of the limits in that ability. A test in spatial relations would be added. By having only four sub-tests instead of seven, it would be possible to make each test much more thorough than were the tests in this battery. Then, too, since each person would be taking the test at his own leisure, it would not be as essential to make the whole examination so brief.

VIII. Conclusions

Considering the elimination of falsely promising sub-tests, a more efficient instrument becomes increasingly possible; and specific operations toward that end are indicated. In view of this state of affairs, some progress can be said to have occurred as a result of this Honors Project.

BIBLIOGRAPHY

Anastasi, Anne, Psychological Testing. New York: Macmillan Company, 1955.

Coopreder, H. A. and Laslette, H. R., "Predictive Values of the Standard Scientific and the Engineering and Physical Science Aptitude Tests," Educational and Psychological Measurement, VIII, (November, 1948), pp. 683-87.

Ferguson, George A., The Fourth Mental Measurements Yearbook. (O. K. Buros, ed.) New Jersey: Gryphon Press, 1953.

Gregg, George W. "An Investigation of the Reliability and Validity of the Engineering and Physical Science Aptitude Test," Journal of Educational Research, XLV, (December, 1951), pp. 299-305.

Griffin, C. H. and Borow, H., "An Engineering and Physical Science Aptitude Test," Journal of Applied Psychology, XXVI, (October, 1944), pp. 376-387.

Mandell, Milton M., "Measuring Originality in the Physical Sciences," Educational and Psychological Measurement, X, (1950), pp. 380-85.

Manual of Instructions for American Council on Education Psychological Examination. Princeton, New Jersey: Cooperative Test Division, 1950.

Moore, Bruce V., Lapp, C. J., and Griffin, C. H., Manual for the Engineering and Physical Science Aptitude Test. New York: Psychological Corporation, 1951.

Roe, Anne, "A Psychologist Examines 64 Eminent Scientists," Scientific American, CLXXXVII (November, 1952), pp. 21-25.

Terman, Lewis M., "Are Scientists Different?" Scientific American, CXCII (January, 1955), pp. 25-29.

Watson, Goodwin, and Glaser, Edward Maynard, Watson-Glaser Critical Thinking Appraisal. New York: World Book Company.

Zyve, David L., Explanatory Booklet for the Stanford Scientific Aptitude Test, California: Stanford University Press, 1930.

TEST I. INFERENCE 7 Mins.

DIRECTIONS: An inference is a conclusion which a person draws from certain observed or supposed facts. Thus, from the electric light visible behind the window shades and from the sound of piano music in a house, a person might infer that someone is at home. But this inference may or may not be correct. Possibly the people in the house went out leaving the lights on, and the piano music could be coming from a radio or phonograph they left playing.

In this test each exercise begins with a statement of facts which you are to regard as true. After each statement of facts you will find several possible inferences -- that is, inferences which some persons might make from the stated facts. Examine each inference separately, and make a decision as to its degree of truth or falsity. On the Answer Sheet you will find for each inference spaces marked with the letters T, PT, ID, PF, and F. For each inference make an X on the Answer Sheet through the appropriate letter as follows:

T -- if you think the inference is definitely true; that it properly follows from the statement of facts given.

PT -- if, in the light of the facts given, you think the inference is probably true; that there is better than an even chance that it is true.

ID -- if you decide that there are insufficient data; that you cannot tell from the facts given whether the inference is likely to be true or false; if the facts provide no basis for judging one way or the other.

PF -- if, in the light of the facts given, you think the inference is probably false; that there is better than an even chance that it is false.

F -- if you think the inference is definitely false; that it is wrong, either because it misinterprets the facts given, or because it contradicts the facts or necessary inferences from those facts.

An English teacher arranged for the students in one of her classes to see the movie Great Expectations, while the students in other classes studied the book itself, without seeing the picture. Tests to measure appreciation and understanding of the story were administered immediately upon completion of each type of instruction. On all tests the class which was taught with the aid of the movie did better. The class which saw the movie became so interested that before the semester was over, most of those students read the book, entirely on their own initiative.

1. The test to measure appreciation and understanding of the story were administered both to the students who saw the picture and to those who only studied the book.

2. The children who were taught with the aid of the motion picture were required to read the book before the end of the semester.

3. Pupils who see movies instead of reading books lose interest in reading.

4. Most of the children in the class which saw the picture would have preferred to study the book Great Expectations in the usual way without the aid of the movie.

5. The teacher who conducted the experiment will hereafter try to use motion pictures when they are available, as an aid in teaching literary appreciation.

6. Pupils can learn more about any given subject from motion pictures than they can from books.

Some time ago a crowd gathered in Middletown, Miss., to hear the new president of the local Chamber of Commerce speak. He said, "I am not asking, but demanding, that labor unions accept their full share of responsibility for civic betterment and community interests. I am not asking, but demanding, that they join the Chamber of Commerce." The listening representatives of the Central Labor Unions applauded enthusiastically. Three months later all the labor unions in Middle town were represented in the Chamber of Commerce, where they served enthusiastically on commitees, spoke their minds, and participated actively in the civic betterment projects.

7. Both the labor union representatives and the other members of the Chamber of Commerce came to recognize one another's problems and viewpoints better through their Chamber of Commerce contact.

8. Labor union participation in the Middletown Chamber of Commerce has largely eliminated worker-management disputes in that town.

9. The active participation of the labor unions caused friction at the meetings of the Chamber of Commerce.

10. The union representatives soon regretted having accepted the invitation to participate in the Chamber of Commerce.

11. Many of the Chamber of Commerce members came to feel that their president had been unwise in asking the union representatives to join the Chamber.

12. The representatives of the Central Labor Unions joined the Chamber of Commerce against the desires of the great majority of their membership.

Studies have shown that there is relatively much more tuberculosis among Negroes in the United States than among Whites. There is no difference, however, in rate of tuberculosis between Negroes and Whites who have the same level of income. The average income of Whites in the United States is considerably higher than the average income of Negroes.

13. Tuberculosis can be cured.

14. Raising the economic level of Negroes would reduce tuberculosis.

15. Tuberculosis is less prevalent among Negroes with relatively high income than among Negroes with relatively low income.

16. Whether a White person is rich or poor makes no difference in the likelihood of his getting tuberculosis.

STOP! DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

6 Min.

TEST 2. RECOGNITION OF ASSUMPTIONS

DIRECTIONS. An assumption is something supposed or taken for granted. When someone states, "I'll graduate in June," he take for granted or assumes that he will be alive in June, that he will remain in school until that time, that he will pass his courses, and similar things. Below are a number of statements. Each statement i followed by several proposed assumptions. You are to decide for each assumpti n whether it necessarily is taken for granted in the statement.

If you think that the given assumption is taken for granted in the statement, make a heavy X through the word YES in the proper place on the Answer Sheet. If you think the assumption is not necessarily taken for granted in the statement, make a heavy X through the word NO on the Answer Sheet. In some cases more than one of the given assumptions is necessarily made; in other cases none of the given assumptions is made.

STATEMENT: "Let us immediately build superior armed forces and thus keep peace and prosperity."

PROPOSED ASSUMPTIONS:

17. If we have superior armed forces, that will insure the maintenance of peace and prosperity.

18. Unless we increase our armaments immediately, we shall have war.

19. We now have peace and prosperity.

STATEMENT: "Even if all the wealth in the country suddenly were to be distributed equally, some people soon would again become rich and others poor."

PROPOSED ASSUMPTIONS:

20. The real causes of wealth and poverty would not be much affected by such Socialism.

21. Our present economic system is better than such Socialism.

STATEMENT: "Mary isn't going to invite John to her party."

PROPOSED ASSUMPTIONS:

22. Mary hasn't yet had her party.

23. Mary now doesn't like John.

24. The party will be at Mary's house.

STATEMENT: "Live in the city of Zenith -- lowest taxes."

PROPOSED ASSUMPTIONS:

25. Efficient management of a city implies lower taxes.

26. An important consideration in deciding where to live is avoidance of high taxes.

27. The people of Zenith are content with their present city government.

STATEMENT: "Our school is fortunate in having all American pupils, so we have no race problems."

PROPOSED ASSUMPTIONS:

28. American pupils do not present any race problems.

29. If we practiced democracy, there would be no race problems.

30. A school is unfortunate if its pupils are of varied nationalities.

STOP!

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO!

$7\frac{1}{2}$ Mins.

TEST 3. MATHEMATICS

DIRECTIONS: The exercises in this test represent commonly used arithmetical and algebraic skills in first year college math and science courses. Solve each of the problems, find the answer among the five choices, and record your answer by putting an X through the letter of the correct choice on the answer sheet.

1. $ax = 8$; what does x equal?
(A) $\frac{8}{a}$ (B) $8a$ (C) $\frac{a}{8}$ (D) $-\frac{8}{a}$ (E) $-\frac{a}{8}$

2. $\frac{s}{t} = v$; what does s equal?
(A) $-vt$ (B) tv (C) $\frac{v}{t}$ (D) $\frac{t}{v}$ (E) $\frac{v}{-t}$

3. $m + n = a - b$; what does n equal?
(A) $a + b + m$ (B) $a - b - m$ (C) $a - b + m$ (D) $a + b - m$ (E) $-a - b - m$

4. Solve for x: $\frac{18.3}{12.2} = \frac{12.9}{x}$
(A) 0.1 (B) .05 (C) 10.2 (D) 14.6 (E) 21.9

5. Solve for y: $\frac{y}{2} + 3x^2$
(A) $\frac{3x}{2}$ (B) $\frac{3x^2}{-2}$ (C) $6x^2$ (D) $\frac{2}{3x^2}$ (E) $\frac{3x^2}{2}$

6. Simplify: $.36p^2$
(A) $72p$ (B) $.6p$ (C) $.06p$ (D) $9p$ (E) $18p$

7. $\frac{a}{b} = y$; what does b equal?
(A) $-ya$ (B) ya (C) $\frac{y}{a}$ (D) $\frac{a}{y}$ (E) $-\frac{y}{a}$

8. Combine: $a - (a - 4) =$
(A) 4 (B) $a^2 - 4$ (C) $2a - 4$ (D) $2a + 4$ (E) $a^2 + 4$

9. If $c = p + q$, how much greater than q is c?
(A) p (B) $p - q$ (C) $q - p$ (D) $-p$ (E) $p + q$

10. If $\frac{W}{W_2} = \frac{W_1}{W}$, what is the value of W?
(A) $W_1 W_2$ (B) $(W_1 W_2)^2$ (C) $\frac{W_1}{W_2}$ (D) $\frac{W_2}{W_1}$ (E) $\frac{W_1}{2W_2}$

11. A man judged a distance of 50 yds. to be 65 yds. What is his % of error
(A) 15% (B) 30% (C) 15/50% (D) 15/65% (E) 50/65%

12. Find the numerical value of $\frac{5 \times 10^6 \times 3 \times 10^{-4}}{.15 \times 1000}$
(A) 1000 (B) .01 (C) .1 (D) 100 (E) 10

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO!

5½ Mins.

TEST 4. FORMULATION

DIRECTIONS: In this part you are to read each statement or short paragraph and do what it tells you to do. In most cases this involves writing an algebraic expression for what the statement says. As in Part I, select the correct answer from among the five choices and record your answer by putting an X through the letter of the correct choice on the Answer Sheet.

1. The larger of two weights is four pounds less than twice the smaller. How heavy is the larger weight if the smaller weighs x pounds?

(A) $2x + 4$ (B) $\frac{x}{2} + 4$ (C) $2(x - 4)$ (D) $x - 4$ (E) $2x - 4$

2. The altitude of a rectangle is h units long. The base is three units longer than the altitude. Write an algebraic expression which represents the area of the rectangle.

(A) $h^2 + h^3$ (B) $2h + h^3$ (C) $2h + 3$ (D) $2 + 3h^2$ (E) $h^2 + 3h$

3. A submarine propelled by Diesel and electric motors crosses a large body of water. The Diesel motors drive the craft y miles and the electric the remainder of the distance. If the electric motors operate 500 miles less than the Diesels, how wide is the body of water?

(A) $500y + y^2$ (B) $2y + 500$ (C) $2y - 500$ (D) $y^2 + 500$ (E) $y + 500y$

4. If a body is to move uniformly in a circular path, it must be pulled toward the center with a force F. The magnitude of this force (F) is equal to the ratio of the product of the mass m of the parts and the square of the velocity V to the radius of rotation r. Express this as a formula.

(A) $F = Vr$ (B) $F + \frac{mr^2}{V}$ (C) $\frac{mV^2}{r}$ (D) $V = \frac{F}{r^2}$ (E) $F = \frac{2}{A} f$

5. How many kilowatt hours of electrical energy can be bought for D dollars if each kilowatt hour costs c cents?

(A) $100\,Dc$ (B) $\frac{100\,D}{c}$ (C) $\frac{D}{100\,c}$ (D) $\frac{D}{c}$ (E) $\frac{c}{D}$

STOP!

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

$7\frac{1}{2}$ Mins.

TEST 5 - PHYSICAL SCIENCE COMPREHENSION

DIRECTIONS: Examine each statement below and decide whether it is true or false. If the statement is true, mark an X through the T on the answer sheet. If the statement is false, mark an X through the F.

1. A boat will sink deeper into the water as it passes from a river into the ocean.
2. The bubbles emerging from a diver's suit becomes smaller as they approach the surface.
3. An opaque substance obstructs the passage of light.
4. Cream has less density than skim milk.
5. Sliding friction is always greater than starting friction.
6. A vacuum is a good conductor of heat.
7. As a body is raised above the surface of the earth the force of gravity pulling it downward becomes smaller.
8. If all forces were eliminated from a moving body, it would gradually come to rest.
9. The temperature of the human body is about 37° C.
10. A bimetal strip can be used to actuate a thermostat.
11. Light travels 100,000 miles per second.
12. Increasing the pressure lowers the boiling point of water.
13. Air under extreme pressure liquifies at - 140° C.
14. The boiling point of water depends upon the atmospheric pressure.
15. The voltage of a group of cells in parallel is the same as the voltage of one cell.
16. A non-compensated pendulum of a clock should be lengthened in cold temperatures.
17. Meteorites become luminous as they near the earth chiefly because they are more clearly visible.
18. If thunder follows lightening at an interval of 10 seconds, the flash must have been 10 miles away.
19. Pure water can be cooled below 0° C without freezing.
20. Pumping more helium into a fully expanded balloon would decrease its lifting power.
21. A thermometer measure the quantity of heat in a substance
22. A siphon will work in a vacuum.
23. A kilogram is equal to 10,000 grams.
24. All known gases have been changed into liquids.
25. Mercury freezes at about - 40° F.
26. Rolling friction is usually greater than sliding friction.
27. The water in a steam locomotive boils at 100° C.
28. Bodies weigh more at the pole than they do at the equator.
29. A newly formed cumulus cloud is part of an ascending air column.
30. Moist air is lighter per unit of volume than dry air.

STOP!

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

5½ Mins.

TEST 6 - VERBAL COMPREHENSION

DIRECTIONS: After each word in CAPITAL LETTERS there are five choices of other words, one of which is most nearly equivalent in meaning to the word in capital letters. Select the most appropriate answer and put an X through its letter on the answer sheet.

1. REFRACTION (A) increase (B) refutation (C) bending (D) uniting (E) acclaiming

2. CONDUIT (A) easy (B) behavior (C) channel (D) uniting (E) concealed

3. CATALYSIS (A) color (B) catacomb (C) change (D) fumigation (E) activation

4. INCREMENT (A) accusation (B) assessment (C) expense (D) addition (E) discrepancy

5. SUBLIMATE (A) cool (B) subdue (C) elevate (D) regulate (E) combine

6. HYPOTHESIS (A) supposition (B) relation (C) provision (D) proof (E) unknown

7. APPARITION (A) storm (B) noise (C) phantom (D) threat (E) danger

8. OSMOSIS (A) combining (B) diffusion (C) ossification (D) incantation (E) clarification

9. COLLOIDAL (A) thin (B) mucinous (C) powdered (D) hairy (E) beautiful

10. PRETEXT (A) ritual (B) fictitious reason (C) sermon (D) truthful motive (E) context

11. DOGMATISM (A) dramatism (B) positiveness (D) doubtful (D) tentativeness (E) realism

12. ELECTRODE (A) officer (B) electrolyte (C) terminal (D) positive (E) election

13. EMIT (A) remain (B) return (C) enter (D) omit (E) discharge

14. PROFICIENCY (A) vocation (B) competency (C) repugnancy (D) prominence (E) urgency

15. BIBLIOGRAPHY (A) description (B) stenography (C) photograph (D) compilation of books (E) typographical sample

16. FIDELITY (A) belief (B) treachery (C) strength (D) loyalty (E) futility

17. SILHOUETTE (A) cloth (B) garnet (C) shadow (D) streak (E) price

18. MENISCUS (A) bottom (B) crescent (C) weight (D) size (E) color

19. ISOTROPIC (A) changed (B) identical (C) transferred (D) opposite (E) isolated

20. AMELIORATE (A) assimilate (B) improve (C) contaminate (D) abuse (E) imperil

21. EMANATING (A) important (B) abundance (C) cooling (D) radiating (E) destroying

22. TACIT (A) tactful (B) loud (C) implied (D) stupid (E) quick

STOP! DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO

TEST 7. NUMBER SERIES

DIRECTIONS: The numbers in each series proceed according to some rule. For each series you are to find the next number.

EXAMPLE: 11 15 14 18 17 21 20 (a)19 (b)21 (c)23 (d)24 (e)27

The pattern is: up four numbers, then down one number. It keeps repeating itself; up four, down one, up four, etc. As you can see, after 21 drops down one to 20, it is time to go up four again. The correct number is 24. You would make an X through the letter d on your answer sheet. Do the others the same way as this one was done.

1.	7	11	15	19	23	27	31	(a)34	(b)35	(c)36	(d)37	(e)38
2.	25	28	24	27	23	26	22	(a)18	(b)19	(c)22	(d)25	(e)26
3.	68	72	36	40	20	24	12	(a) 6	(b)16	(c)20	(d)24	(e)28
4.	86	78	70	62	54	46	38	(a)28	(b)30	(c)32	(d)34	(e)36
5.	94	92	46	44	22	20	10	(a) 4	(b) 5	(c) 8	(d)12	(e)14
6.	25	22	11	33	30	15	45	(a)15	(b)41	(c)42	(d)48	(e)135
7.	4	7	8	7	10	11	10	(a) 6	(b) 9	(c)11	(d)13	(e)14
8.	10	12	14	12	14	16	14	(a)12	(b)14	(c)16	(d)18	(e)20
9.	35	28	4	11	77	70	10	(a) 4	(b)17	(c)63	(d)70	(e)77
10.	4	5	7	4	8	13	7	(a) 0	(b)13	(c)14	(d)15	(e)16
11.	49	51	54	27	9	11	14	(a) 7	(b)16	(c)17	(d)18	(e)28
12.	4	8	16	8	16	32	24	(a)12	(b)16	(c)24	(d)32	(e)48

STOP! END OF TEST