

1932

The comparative validity of subject matter in certain scholastic aptitude tests

Alfred Harold Holway
University of Massachusetts Amherst

Follow this and additional works at: <https://scholarworks.umass.edu/theses>

Holway, Alfred Harold, "The comparative validity of subject matter in certain scholastic aptitude tests" (1932). *Masters Theses 1911 - February 2014*. 1617.
<https://doi.org/10.7275/6871160>

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

UMASS/AMHERST



312066013576944

COMPARATIVE VALIDITY OF SUBJECT MATTER
IN CERTAIN SCHOLASTIC APTITUDE TESTS

HOLWAY - 1932

LIBRARY
UNIVERSITY OF MASSACHUSETTS



DATE DUE			

UNIV. OF MASSACHUSETTS/AMHERST
LIBRARY

LD
3234
M268
1932
H759

THE COMPARATIVE VALIDITY OF SUBJECT MATTER
IN CERTAIN SCHOLASTIC APTITUDE TESTS.

BY
ALFRED HAROLD HOLWAY

"THESIS SUBMITTED FOR DEGREE OF MASTER OF SCIENCE."

"MASSACHUSETTS STATE COLLEGE, AMHERST."

JUNE, 1932.

CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
1. Statement of the problem.....	1
2. Development of the mental tests.....	2
3. Present status of the scholastic aptitude tests.....	7
II. COLLECTION OF DATA.....	14
1. Construction of the Massachusetts State College Scholastic Aptitude Test.....	14
2. Administfation of the tests.....	22
3. Collection of data.....	23
III. THE PRELIMINARY ANALYSIS.....	25
1. Nomenclature and methodology.....	25
2. Results of the preliminary analysis.....	33
3. Summary and interpretation.....	72
IV. SECONDARY ANALYSIS: FIRST TERM.....	75
1. Methodology.....	75
2. Composite scores and average marks.....	78
3. Gross scores and average marks.....	79
4. Gross scores and term marks.....	81
5. Individual scores and average marks.....	85
6. Individual scores and term marks: the Psychological Examination.....	91
7. Individual scores and term marks: the Massachusetts State College Scholastic Aptitude Test.....	100

8. Summary and interpretation.....	111
V. THE HIERARCHY OF INTELLIGENCE.....	115
1. The problem and the method.....	115
2. Results: the Psychological Examination..	116
3. Results: the Massachusetts State College Scholastic Aptitude Test.....	113
4. Summary and interpretation.....	134
VI. PERCENTILE INTERPRETATION.....	137
1. The problem and the method.....	137
2. First quartile comparison.....	138
3. Second quartile comparison.....	141
4. Summary and interpretation.....	150
VII. SUMMARY AND CONCLUSIONS.....	152
APPENDIX: SAMPLE COPY OF THE MASSACHUSETTS STATE COLLEGE SCHOLASTIC APTITUDE TEST.....	157
REFERENCES.....	171
ACKNOWLEDGMENTS.....	173

TABLES

Table	Page
1. Calculation of the median of a frequency distribution.....	29
2. Computations of the means of a frequency distribution.....	30
3. Calculation of the mode of a frequency distribution.....	31
4. Calculation of the standard deviation of a frequency distribution.....	32
5. Gross scores and first term marks.....	82
6. Individual scores and first term average marks (Psychological Examination).....	87
7. Individual scores and first term marks (Psychological Examination).....	93
8. Individual scores and first term marks (Massachusetts State College Scholastic Aptitude Test).....	103
9. Intercorrelations among individual scores (Psychological Examination).....	118
10. Intercorrelations among individual scores (Massachusetts State College Scholastic Aptitude Test).....	124
11. Quartile comparison showing degree of correspondence between first term average marks and mental test composite and gross scores.....	139
12. Quartile comparison between the first term average marks and the Army Group Examination Alpha gross scores.....	142
13. Quartile comparison between the first term average marks and the Massachusetts State College Scholastic Aptitude Test gross scores...	144
14. Quartile comparison between the first term average marks and the Psychological Examination gross scores.....	146

Table	Page
15. Quartile comparison between the first term average marks and the composite scores...	148
16. Quartile position of the twenty-five first- term failures.....	149

FIGURES

Figure	Page
1. Actual and theoretical frequency distributions of the first term average marks.....	34
2. Actual and theoretical frequency distributions of the first term Orientation marks.....	35
3. Actual and theoretical frequency distributions of the first term English marks.....	37
4. Actual and theoretical frequency distributions of the first term Mathematics marks.....	39
5. Actual and theoretical frequency distributions of the first term Chemistry marks.....	40
6. Actual and theoretical frequency distributions of the first term French marks.....	41
7. Actual and theoretical frequency distributions of the first term German marks.....	42
8. Actual and theoretical frequency distributions of the first term German and French marks.....	43
9. Actual and theoretical frequency distributions of the second term average marks.....	45
10. Actual and theoretical frequency distributions of the second term Orientation marks.....	46
11. Actual and theoretical frequency distributions of the second term English marks.....	47
12. Actual and theoretical frequency distributions of the second term Mathematics marks.....	48
13. Actual and theoretical frequency distributions of the second term Chemistry marks.....	49
14. Actual and theoretical frequency distributions of the second term French marks.....	51
15. Actual and theoretical frequency distributions of the second term German marks.....	52

16. Actual and theoretical frequency distributions of the Composite Scores.....	53
17. Actual and theoretical frequency distributions of the Army Alpha gross scores.....	54
18. Actual and theoretical frequency distributions of the Psychological Examination gross scores.	56
19. Actual and theoretical frequency distributions of the Massachusetts State College Scholastic Aptitude Test gross scores.....	57
20. Actual and theoretical frequency distributions of the Completion section individual scores (Psychological Examination).....	58
21. Actual and theoretical frequency distributions of the Opposites section individual scores (Psychological Examination).....	59
22. Actual and theoretical frequency distributions of the Artificial Language Section individual scores (Psychological Examination).....	60
23. Actual and theoretical frequency distributions of the Analogies section individual scores (Psychological Examination).....	61
24. Actual and theoretical frequency distributions of the Arithmetic section individual scores (Psychological Examination).....	62
25. Actual and theoretical frequency distributions of the Botany section individual scores (Massachusetts State College Scholastic Aptitude Test).....	65
26. Actual and theoretical frequency distributions of the Mathematics section individual scores (Massachusetts State College Scholastic Aptitude Test).....	66
27. Actual and theoretical frequency distributions of the Chemistry section individual scores (Massachusetts State College Scholastic Aptitude Test).....	68

28. Actual and theoretical frequency distributions
of the Power Reading section individual scores
(Massachusetts State College Scholastic
Aptitude Test)..... 69
29. Actual and theoretical frequency distributions
of the Memory Reading section individual scores
(Massachusetts State College Scholastic
Aptitude Test)..... 70
30. Actual and theoretical frequency distributions
of the Artificial Language Section individual
scores (Massachusetts State College Scholastic
Aptitude Test)..... 71

CHAPTER I

INTRODUCTION

The general purpose of this study is to investigate the relative validity of different types of subject matter in certain scholastic aptitude tests for college freshmen. More specifically, an attempt will be made to answer the following questions:

1. What is the comparative validity of (a) the Psychological Examination, (b) the Army Group Examination Alpha, and (c) the Massachusetts State College Scholastic Aptitude Test with respect to the criterion, academic success?

2. To what extent does a test based essentially on recall and recognition differ from one based on ability to adapt oneself to a novel situation as regards accomplishment?

3. Which type of subject matter contains the most valid and "extensive" items?

4. What is the practical validity of each test as a prognostic indicator of academic success?

5. Which type of subject matter will select the highest and the lowest average-mark quartiles with the greatest accuracy?

6. To what extent will these tests indicate the students least fitted to succeed in the work offered in the freshman curriculum?

The final answers to the above questions are subject to

many limitations and restrictions. The number of cases used in the study is but two hundred and eighty-five. The conditions under which the tests were administered were not ideal: the subjects were divided into two groups, two administrators were employed, and some twelve students took at least one of the tests a week late. The marks used in the work are limited to those obtained by the Class of 1935 during the first term of the 1931-1932 academic year. Moreover, in regard to questions 2 and 3, the investigation is restricted to the scores obtained on the Psychological Examination, the Massachusetts State College Scholastic Aptitude Test, and school marks. In spite of the many limitations, however, it is hoped that this minor study may furnish some data which will cast a pencil of light on a narrow phase of the most baffling problem now confronting the biological sciences; namely, prediction.

Development of the mental tests. Ever since Thales turned to nature for an explanation of nature, man has attempted to analyze and classify natural phenomena without reference to a deus ex machina; i.e., scientifically. The rise and development of experimental psychology is unquestionably a phase of the history of scientific effort. Curiously enough, it was in astronomy, the first science, that attempts were made to measure individual variations.

At Greenwich, in 1796, an astronomer's assistant was discharged from the Observatory because his reports differed

from those of his superior by a seemingly incredible length of time: 0.8 seconds. Fortunately, the event was recorded in an historical account of the Observatory in the Astronomical Observations at Greenwich which, in 1819, fell under the critical eye of Bessel, the astronomer-director of the Konigsberg Observatory. This man, a pioneer in the more exact measurements of modern astronomy, subjected the matter to an experimental investigation, and discovered what is now known as the "personal equation". This was the nativity of differential psychology--the scientific study of individual differences.

Later, c. 1850, the United States Coast Survey developed Repsold's chronograph which made easy the measurement of the "personal equation". The next two decades saw (1) the perfection of the Hipp chronoscope which measures reaction-time intervals in thousandths of a second, and (2) the culmination of the astronomer's work on the "personal equation". Thus did astronomy furnish modern psychology with a problem, a few essential facts, and some apparatus.

A very great stimulus to the study and consequent measurement of individual differences was presented by the English scientist, Sir Francis Galton. In 1869, ten years after the publication of Darwin's Origin of Species, Galton proposed an imaginary scale for the measurement of general ability. The scale was based on the principle of normal distribution, was divided into fourteen grades, and ranged from the idiot to the

genius. In 1884, Galton founded the Anthropometric Laboratory in London, and it was here that the mental test movement had its origin. However, in spite of the fact that he had posited general mental ability theoretically, Galton measured only simple capacities in the laboratory. In 1886, he introduced the method of statistical correlation which has contributed greatly to a more accurate classification of vital data.

Chronologically, the next significant personality in the development of the mental tests was Hermann Ebbinghaus. To him, psychologists owe the nonsense syllable -a simple and ingenious instrument of measuring associations, immediate memory span, etc.; the Ersparnismethode and the Erlernungsmethode - methods of savings and complete mastery; and the principle of the completion test - a medium now widely used by many educators. Yet, although Ebbinghaus' greatest influence on later psychological thought was his systematic attack on the mnemonic processes, the mental test workers remember the crucial importance of the completion test as a working tool, and it is for this contribution that he deserves mention.

The last decade of the nineteenth century is known to psychologists as the period of democratization, for two major reasons: (1) the control of the psychological laboratories passed from the few to the many, and (2) psychological effort suffered a shift in emphasis -the natural resultant of the first change-

from pure science to applied research.¹ From 1900 on, and with few exceptions, the materials and methods employed in mental testing became tools for practical research rather than objects for improvement and refinement. The outstanding exception to this rule² was the work of the French psychologist, Alfred Binet. In 1905, this man proposed a series of tests which was to measure general intelligence -in direct contrast to all that had gone before, and all the prevailing, contemporary theories. The fundamental idea of his measure was to secure data in regard to mental capacity (brightness) with direct reference to the chronological development of the individual. The idea has proved most fruitful from the standpoints of both pure science and educational technology.³ Three years later (1908), Binet's efforts culminated in what is now known as the Binet-

1. For example, in 1891, Bolton measured the immediate memory span for digits of certain school children for the purpose of predicting success in grade school work.--T.L.Bolton, "The Growth Of Memory In School Children"; American Journal of Psychology, 1891, (4), pp. 362-80.

2. Further attempts to measure general intelligence were made, and are still being made, all over the world. The most outstanding figures in this field are Yerkes and Thorndike in America, Spearman in England, and Stern in Germany. However, the great majority of the vast army of mental testers has rallied to that side of "no man's land" which flies the banner of practicality, attempts to measure specific capacities, or aptitudes, and is composed almost entirely of applied psychologists and teachers of pedagogy.

3. Professor Boring of Harvard University ventures to assert that during this decade the development of intelligence tests at the hands of Binet and under his influence is second in importance only to the development of the experimental psychology of thought. - E.G.Boring, A History of Experimental Psychology, The Century Publishing Co., 1929, p. 636.

Simon scale. This measure classified children according to mental age. It ranged from three to thirteen years inclusively, its unit was one year, and it was employed with much success. Nevertheless, the scale had one serious drawback: for each pupil to be tested, there had to be present an expert test administrator.¹ Obviously, such a limitation hindered universal usage.

Later, in 1917, faced by the immediate practical need for a test unhampered by any such restrictions, the American Psychological Association met the situation with the Army Group Examination Alpha. Then came the deluge: the unprecedented "success" of the test, the fact that it had been conscientiously administered to 1,729,966 incipient soldiers, was extremely popular, and entertained promising possibilities, gave the whole test movement an impetus from which it has not fully recovered. Almost over night, tests appeared by the dozen, each one more "promising" than its predecessor. Portentous claims were too often made for the tests by their promoters.² On the

1. This practical weakness still obtains, in spite of the several subsequent revisions of the scale (Goddard, 1911; Stanford 1913; Herring, 1922).

2. And the end is not yet. To imply, as does Professor Pintner of Columbia University, that the tests will soon be able to select such typical personages as "'The Immigrant'; 'The Voter'; 'The Applicant for a Marriage License'; 'The Candidate for Public Office'; 'The Civil Servant'; and so forth." (Cf. Preface, Intelligence Testing, R. Pintner; Henry Holt and Company, 1931) is hardly warranted by the available facts. New movements in science progress slowly and cautiously, in spite of the fact that discovery is the chief concern of science. In the opinion of the writer, advancement is most sure when it is the resultant of the force (say) of the new movement and the persistent inertia of the opposing criticism: intemperance often involves a needless waste of time and energy.

other hand, the Army Group Examination Alpha, the first group test, supposedly served its purpose well, and is still used by the United States Army and many civil and academic institutions.

The most recent development of the mental test movement is the attempt to predict specific abilities, aptitudes, or capacities. Instead of trying to measure general intelligence for purely scientific reasons, the mental testers are, in the main, now attempting to measure specific abilities in order to predict future accomplishment. A brief example of the method employed is as follows: Outstandingly expert and successful mechanics are selected and requested (1) to solve some mechanical puzzles, (2) to take a completion test, and (3) to answer certain definite questions. The results are analyzed and classified, and then on the basis of these data a mechanical aptitude test is formulated. The list of such non-academic aptitudes which have been studied is extensive and increasing. It ranges from ability tests for policemen and factory employees to those for musicians and interpretive readers. Even in the restricted field of academic study, numerous tests have been devised to forecast abilities in particular subjects, and with some success.

Present status of the scholastic aptitude tests. Ability to do well in academic tasks depends on a large number of factors among which intelligence, as measured by the tests, is but one. Scholastic success, requires (say) docility, regular hab-

its, sustained effort over prolonged periods, and so forth. To date, there is no evidence that such factors are encompassed at all by the so-called scholastic aptitude tests. In fact, gestalt psychologists vigorously assert that the nature of what is tested is not at all known, and for this reason they condemn the whole test movement. Kohler believes that only by first understanding the nature of behavior, association, reproduction, and insight can intelligence be understood. Quite in contrast to this viewpoint are the interpretations of Edward L. Thorndike and Charles Spearman. The former maintains that intelligence, as measured by the tests, is a composite of the individual's specific capacities; the latter that intelligence is general. Spearman,¹ from elaborate statistical studies induces a general (g) factor, and this, he infers, is more fundamental than the specific (s) factors. From similar (and often the same) investigations, Thorndike² finds only specific factors in evidence. In his opinion, the "g factor" is occasioned by the overlapping of the "s factors". Each theory has many ardent supporters, and each side has advanced seductive evidence to support its tenets. Interested psychiatrists, physiologists, psychologists, medical men, and

1. Spearman, C., "General Intelligence, Objectively Determined and Measured", American Journal of Psychology (15), 1904, pp. 201-293.

2. Thorndike, E.L., The Measure of Intelligence, New York, 1927.

philosophers appear to be equally divided. The writer is aligned by no particular group -for him, the tests measure responses to certain standardized formulae.

In practice, the responses to specific stimuli are compared with definite criteria. As regards the prediction of academic success, the utilized criteria are (1) general achievement, as measured by the student's average mark, and (2) achievement in a definite subject or group of subjects; the determining technic employed usually being the product-moment of Karl Pearson. Numerous studies are now being made to test the validity, reliability, and usefulness of each new prognostic measure. From all accounts, the most successful measure is the Thorndike Intelligence Examination (for high school graduates). It is based essentially on recall, has four sections, and requires two hours and fifty minutes on the part of the examined. In 1921, it was given to the entering freshmen at Columbia University. The resulting scores were compared with the marks secured by the participants over the four-year period of undergraduate work. The correlation coefficients between the university marks and the test scores are as follows:

Year	I	II	III	IV
Pearson Coefficient	.56	.43	.36	.38

Nevertheless, in spite of the probability that these unusually

high correlations may be accurate, the writer challenges their prognostic validity on ground not statistical; to wit, the fact that the test scores were distributed among the instructors.¹

Since the present study involves prediction in collegiate work, results of the tests most extensively used at this level will now be presented. The previously mentioned Army Group Examination Alpha predicts academic success as follows:

Arkansas University.....	.28-.65
Brown University.....	.44-.46
Carnegie Institute.....	.34
Dartmouth.....	.44
Hamlin University.....	.47
Illinois State.....	.48
University of Illinois.....	.37
Minnesota University.....	.15-.50
Ohio State.....	.15-.38
Southern Methodist University.....	.52
Stanford University.....	.31-.43
Syracuse University.....	.20-.40
Yale University.....	.37-.38

The most widely used test of college level is the Psycho-

1. The same test was used at the University of Chicago, where the test scores were not made known to instructors. The highest obtained correlation was .40. Moreover, a recent investigator reports that certain instructors' judgments were definitely affected by a knowledge of the mental test scores (Cf. "Instructor's Use of Mental Test Scores", Journal of Educational Research, January, 1932).

logical Examination.¹ The first edition of this test was issued in 1924. Since then, it has been administered to over 750,000 college freshmen, and revised seven times. Of the thirteen different sections utilized in previous editions but five -the most valid- have been retained. In a survey covering 43 American colleges and universities, MacPhail reports that the correlation between the results obtained on the Psychological Examination and college success is .29

In Canada, the test results are less promising. In an investigation relative to the comparative validity of matriculation examinations and mental tests in predicting scholastic achievement, Kellogg² finds a .173 coefficient for both tests, i.e., the Army Group Examination Alpha and the Thurstone IV. He believes that urban distractions and general lack of interest in course work account for the low degree of correspondence. Kellogg further concludes that if ability, application, and past experience are the three chief factors in scholastic success, application is the most important. And so it would seem. Without application, past experience and ability may be of little import.

An extensive study in regard to scholastic prognosis is

1. This test, sometimes called the Thurstone IV, is recommended and published by the American Council on Education

2. Kellogg, C.E., "Relative Values of Intelligence Tests and Matriculation Examinations as Means of Estimating Probable Success in College", School and Society, (26), October, 1926, pp. 501-502.

presented by Odell¹ of the University of Illinois. Having tested 1981 freshmen with the Otis Self-Administering Test, he reports that the correlation between intelligence test scores and the general average for the freshman year is .38. Professor Odell ascribes the low predictive value of this result to the following factors: (1) the test is too brief (requiring but 30 minutes), (2) many inexperienced administrators -high school principals, teachers, and even school committee men and women -introduced the tests, and in general (3) conditions were inadequately controlled.

Needless to say, the results of these representative surveys are hardly encouraging. The above figures indicate that the most popular "barometers" of mental ability are practically useless for the differential prognosis of scholastic success. On the one hand, in predicting achievement or aptitude of whatever kind, perfect correspondence is not expected. Yet practically speaking -predictions should be nice enough to have value in discriminating between individuals who may be expected to do well in academic work and individuals who are unlikely to do acceptable academic work. How may this be accomplished?

In an unpublished classroom discussion, Professor H. N. Glick stated that psychologists have been measuring recall,

1. Odell, C.W., "An Attempt at Predicting Success in the Freshman Year at College", School and Society, (25), June, 1927, pp. 702-706.

and predicting ability to succeed in an institution of learning. He also advanced the hypothesis that any indirect measurement must be as nearly as possible in harmony with the attribute to be predicted. This hypothesis implies that success in college rests more on the capacity of an individual to memorize, along with the attendant attributes, the ability to learn, retain, recall and recognize -than on the capacity to recall knowledge acquired in the past as, for example, it is measured by the Army Group Examination Alpha and the Psychological Examination. That this hypothesis (good guess) might be evaluated logically, there is no question. However, the conviction that its worth should be determined by experimental investigation gave rise to the present study, of which a major portion necessarily comprises the preparation of a test based essentially on ability to achieve through learning, retaining, recalling and recognizing new material (Cf. question 2, p.1).

CHAPTER II

COLLECTION OF DATA

Construction of the Massachusetts State College Scholastic Aptitude Test. Since there were no available tests which measured ability to learn, Professor H. N. Glick and the writer have prepared an empirical, learning test which, it is hoped, may serve ultimately to measure academic ability somewhat more successfully than have the other tests used in this experiment. The Massachusetts State College Scholastic Aptitude Test¹ comprises six sections and two "assignment" sheets, and is presented only tentatively. In part, and as a whole, the scale is regarded as a venture rather than an achievement. It is set forth as a provisional and imperfect, preliminary attempt, and -if the experimental results warrant such- as a point of departure for further research. Should this study prove it to be of impractical validity, it is to be remembered that the new scale is the studio model, and not the finished effigy. From such a preparatory cast, the scientist is no more exempt than the artist.

Previous to any attempt at the initial figurine, however, an agreement was reached in regard to the materials to be used, the nature of the thing to be delineated, and the general structure of the piece. That is to say, apart from the assumptions that all the individuals to be tested could read and

1. Cf. Appendix for copy of the test.

write English, and could follow simple directions, it was agreed that the test should include no material which might be familiar to the majority of the entering freshmen; that the subject matter used should be in harmony with that of the freshmen curriculum; that the responses evoked should be analogous to those called forth by work in college; and that these responses should be directed -in order the better to eliminate personal bias in correction.

The Massachusetts State College Scholastic Aptitude Test differs from all other general, academic tests inasmuch as it requires the student to study subject matter during the time allotted for the examination. Just before the examination proper, each participant is given two study sheets, face down, which he is advised to study carefully¹ in order that he may answer questions on them later. After the students have spent seven minutes on the first sheet, they are told to turn to the second study sheet. No restriction is placed on the remaining seven minutes. At his pleasure and discretion, the subject may complete the second study sheet and revert to the first.

Each study sheet is composed of two parts, A and B. Part A of Section I, contains a life-sized diagram of an agaric

1. This reveals a distinct limitation of the group test. If the gods of chance be propitious, the individual score rises. Should the mood or attitude of the student be such as to render 100% cooperation impossible, the score suffers accordingly. An individual test, on the other hand, may be postponed, and administered under more favorable circumstances. To offset this difficulty is the chief raison d'etre of the "battery".

(mushroom).¹ Beside the drawing are names indicating sixteen portions of the agaric. To the right, and below, the first part is Part B comprising fourteen biological definitions. In each instance the key-word, to be deleted later, is under-scored. Part A of Section II (second study sheet) consists of a brief sketch pertaining to Sir Thomas Malory and his work Morte d' Arthur.² Part B presents George Santayana's sonnet, "O world, thou choosest not the better part!..." The subjects are cautioned not to commit the selections to memory. At the close of the allotted fourteen minutes, the study sheets are passed, face down, to the aisles, where they are collected by the proctors, and copies of the aptitude test are distributed. When the name, age, and school last attended has been written on the cover page of the test, orders are given to "turn to Section I and look at the directions" -the examination is begun.

At this juncture, a detailed description of the complete scale is pertinent. Section I of the test is divided into two parts, A and B. Part A might well be called an identification test. Psychologically, each identification-response entails

1. See Acknowledgements, p. 173.

2. Professor Greenlaw of Johns Hopkins University was the authority consulted.

reproduction (visual imagery) and association.^{1,2} The figure³ on the first study sheet is reproduced without alteration. In place of the botanical name (used on the study sheet to identify the various anatomical structures), numbers (1,2,3,etc.) have been substituted. To the right of the figure is a list containing the missing botanical terms -plus two additional (and in this case), meaningless, morphological terms; namely, stipule and mesodermis. It is assumed that the addition of these terms will lessen the chances of successfully identifying, say, the fourteenth and fifteenth parts of the structure by the process of elimination. Beside each term are parentheses in which the number adjacent to the corresponding part is to be introduced. One unit of credit is awarded for each of the fifteen numbers correctly inserted. As soon as the participant has completed Part A, he is instructed to turn immediately to the following page.

Part B, in the judgment of the writer, is an improved form of what is commonly denoted a completion test. For instance, instead of calling upon the student to complete the sentence (in this case, the biological definition) with the missing term, the improved technic requires the student to insert a

1. Cf. Kohler, op. cit.

2. To what extent these factors are accessed is beyond the scope of this research. In point of fact, neither one has been studied outside the laboratory.

3. See Acknowledgements, p. 173.

number in the parentheses beside that term thus eliminating errors in spelling, destroying the possibility of presenting an improvised noun (disguised to confound the examiner), and facilitating correction. In this part of Section I, meaningful recall is stressed. With the exception of the key-words, the fourteen sentences presented on the study sheet are copied verbatim. Below the definitions, the deleted terms with adjacent parentheses -and two extra words; namely, zygote and brachycephalic- are set forth in two columns. Credit of one point is given for each correctly introduced numeral. Exactly ten minutes is allowed for the subjects to attempt Section I, Parts A and B.

Section II of the test is entitled "Artificial Language". Here, the subject is confronted with a situation quite analogous to the grammar assignment presented in college first year language work. A novel vocabulary, definite rules accompanied by examples, and sample sentences are placed before the encumbents. In addition, there are twenty-five English sentences and twenty-five equivalent sentences in the Artificial Language.¹ It is assumed that all students can distinguish among nouns, pronouns and verbs, and know the difference between the objective and nominative cases. Among the twenty-five artificial sentences are twenty-eight "errors" in translation. One credit unit is given for each detected "error",

1. Cf. Appendix, Massachusetts State College Scholastic Aptitude Test, Section II, pp. 2-3.

and one unit is deducted for each mistake in judgment. fifteen minutes is allowed for this section.

Part A, Section III, contains the substance of Part A on the second study sheet with some words omitted. The process involved (completion) is the same as that in Section I, Part A, with two exceptions. The first is one of subject matter; the second, one of organization. Instead of fairly technical nouns, this part omits some common place names and dates, verbs, and exact qualifying adjectives. Section I, the sentences are discrete -in this part, they are resolved into a unitary, meaningful paragraph. Below Part A is Part B which contains a list of twelve sentences based on the sonnet typed on the second study sheet. Some of these statements are true, some are false, and some (two) contain ideas not at all presented in the poem. The three catagories extend the scope of the test, and reduce the possibility of correct "guessing". Now, although this section is called "Memory Reading", there is a factor other than memory in force; viz, retroactive inhibition. It is a minor limitation of this work that this factor not even qualitatively here taken into account. Future research may decide its influence quantitatively.¹ One point is

1. The following method suggests a possible resolution: (1) divide the class into two groups (X, the control group, and Y, the experimental group), (2) present Section III to X immediately after the study sheet has been removed, (3) give Section III to Y in the normal order, and (4) compare (a) averages and (b) with a criterion.

given for each correct response; the highest possible score being twenty seven. The time allotted to Section III is exactly seven minutes.

In Section IV,¹ there are four separate parts -A, B, C, and D. This section is itself an aptitude test in epitome, one which well might be called a theoretical, chemistry aptitude test.² Wheeler³ defines reasoning as "the adjustment made to a novel situation by employing a concept". In the light of this definition, Section IV seems to be the logical ideal for a test of an individual's reasoning power.⁴ In each of the four parts, there are stated some facts and principles (concepts) followed by a number of novel⁵ problems. Each part is independent of the others, although each has to do with a particular phase of chemistry. Part D is worthy of especial note. Following the facts and principles set forth in this part are two figures. The first is a presented explanation re. The

1. Cf. Acknowledgements, p. 173.

2. This is mere conjecture, however. In spite of the fact that Section IV proved more valid than the other sections when compared to the criterion, average marks, a distressing discrepancy occurs when compared with chemistry marks -(Cf. Results, p. 107 -ergo, the limiting adjective, theoretical.

3. Wheeler, R.H., The Science of Psychology, Crowell Co., New York, 1929, p. 148.

4. With limitations, the mathematical results seem to justify the inference.

5. In the opinion of the creator of this section, Professor Paul Serex, the principles, facts, and problems here presented are entirely new to the majority of subjects taking the test.

second, the problem, demands of the student (1) a knowledge of the printed facts and principles and (2) a cognitive grasp of the process involved. Here, if nowhere else, is taxed the student's ability to reason. In regard to the time spent on each part, no exact regulation is recommended, but thirteen minutes are allowed for the whole. Twenty-four points constitutes a perfect score.

Section V is similar to the preceeding section inasmuch as it also demands that the subject learn and apply concepts in order to solve definite problems. Immediately below the directions are three simple, general rules taken from the calculus.¹ The necessary symbols are "translated" into English, and two sample problems are solved. The ten problems which follow can be evaluated independent of previous training in algebra, although a knowledge of coefficients, etc., will admittedly and undoubtedly facilitate the process temporally. However, this is not a speed test, since all of twelve minutes are given to this section. Three points were allowed for each correctly evaluated integral. If the constant of integration be omitted, one point is discounted.

The last section of the test, VI, is denoted "Power Reading".² Here is a quite difficult passage³ for the average,

1. The authority here consulted is Love, C.E., Ph.D., The Calculus, Macmillan Co., 1926.

2. One has but to consult the text to appreciate the force of this title. Cf. Section VI.

3. The passage is quoted from Boring, E.G., A History of Experimental Psychology, The Century Co., 1929, p. 166.

entering freshman. Beneath the quotation are thirteen statements which are to be marked (1) true, (2) false, or (3) didn't say. No memory work is requisite. Unlike, Section III, Part B, which is based on the recall of reading matter, Section III includes the material to be read and the questions regarding it. Each correct response is accorded two points. The time permitted for this section is exactly ten minutes.

Administration of the tests. A regulation issued from Office of the Dean requires all entering freshmen to take the mental tests administered by the Department of Education. During the first week of the 1931-1932 academic year, Professor H. N. Glick and the writer administered three sets of tests to c. 300¹ incoming freshmen on three consecutive days in the following order: The Army Group Examination Alpha, the Massachusetts State College Scholastic Aptitude Test, and the Psychological Examination. Twenty-four hours elapsed from the time of beginning one examination to the start of the next. Because of space limitations, the class was split into two equal groups.² No pains were spared to keep conditions as constant as may be: (1) each group was seated at random, (2) care was taken to start the tests at the same time, (3) the same directions were read in each case, (4) and three proctors

1. No more than 15 students appeared later or were absent during the three examinations.

2. Cf. p. 2.

were at the disposal of each group.

The attitude of the subjects deserves some mention. From direct observation, the writer feels that a large majority did their best to cooperate. (In a broad sense, these tests measure the subjects' willingness to cooperate). When the signal to begin was given they went to work enthusiastically, and when they were told to stop, they obeyed orders. The impressions of the proctors coincided with that of the writer. No attempts to cheat were reported. There were, however, a few slackers who spent much of the time yawning and gazing about the auditorium.

Collection of data (cont.). In all, 912 tests were corrected, scored, and ranked. All of the scores made on the Psychological Examination and the Massachusetts State College Scholastic Aptitude Test were recorded by sections and in composite. A decided limitation of this experiment is the fact that the Army Group Examination Alpha scores are not divided into sections thus permitting a more critically statistical analysis of its predictive value. However, as a whole this test was also corrected, scored, and ranked. During the third week of the second term, the first term marks of the subjects used in this study were obtained by special permission from the Office of the Dean.

The following entities, then, comprise the essential data used in this research:

1. 285 subjects;
2. the Army Group Examination Alpha, the Psychological Examination, and the Massachusetts State College Scholastic Aptitude Test;
3. the gross scores¹ obtained on each of these tests (and composite);
4. the individual scores of the Psychological Examination and the Massachusetts State College Scholastic Aptitude Test;
5. the term marks and the average marks obtained by each freshman during the first and second terms of the 1931-1932 academic year.

1. For definitions of terms, cf. p. 24, Nomenclature, etc.

CHAPTER III

THE PRELIMINARY ANALYSIS

Nomenclature and methodology. To prevent ambiguity or confusion in regard to terminology, a mark received in a single academic course of study is designated a term mark, and the general average of an individual for a completed term, an average mark. A score obtained on a single section of a mental test is referred to as an individual score; the sum of the individual scores on a test are called the gross score; and a combination of two or more gross scores, a composite score. Moreover, the average mark is taken as the criterion of academic success for that term. In similar fashion, a term mark is the criterion of achievement in that particular course of study.

Since, in this work, the sole criteria of academic achievement are the obtained marks, it is expedient to investigate the general nature of these standards. Strictly speaking, a mark represents the final judgment of the college instructor -who, in the words of Professor Dearborn, "is thus set up as the arbiter elegantiarum of the academic world and of its intelligence."¹ Thus, when taken individually, a mark stands definitely for scholastic accomplishment. If marks be taken collectively, however, it becomes at once necessary to

1. Dearborn, W.F., Intelligence Tests, Houghton Mifflin Co., 1928, p. 47.

discover whether or not they are truly representative of the group in question. This step entails an acceptance of the fundamental assumption of biometric research; namely, that "all representative measures within a class are symmetrically distributed about or concentrated at the mid-point of that class."^{1,2} A.S. Otis carries the theory to its logical conclusion when he applies it to teachers' marks: "...if the judgments of the teacher were expressed in true numerical terms, the measures of any fairly large group, such as the pupils of a class, would be distributed approximately according to the law of normal distribution."³ From these statements, it seems logical and plausible to conclude that the marks attained by any group of (say) 100 or more students, which within the limits of probable error do not corroborate the law of normal distribution, suggest that the courses are either "too difficult"⁴

1. Odell, C.W., Educational Statistics, the Century Co., 1925, p. 23.

2. The results of empirical research fully justify the acceptance of this assumption. Convincing evidence, in regard to the continuity and normality of mental attributes is abundant (Cf. Memoirs of the National Academy of Science, 15; Terman, L.M., The Measurement of Intelligence, Houghton Mifflin Co., 1916, p. 66; Woodworth, R.S., Psychology, Holt and Co., 1931, p. 39; etc.).

3. Otis, A.S., Statistical Method in Education Measurement, World Book Co., 1926, p. 287.

4. In no way should "difficulty" be equated with a high "intellectual" level. The inference here implied is that a "too difficult" course (as indicated by an abnormally skewed distribution) is either (1) unnecessarily "padded" or (2) taught by an incompetent instructor.

or "too easy" -and such is the contention of both Starch and Otis.¹

Odell² states that one of the best methods to compare and contrast, clearly and effectively, two fairly large measures of paired facts (scores and marks, for instance) is to represent the frequency distribution of each by means of a graph. The three most common forms of such a surface are the frequency polygon, the histogram or column diagram, and the normal frequency curve. For this particular study -the comparison of mental test scores and college marks -the histogram is selected, because it "is probably the most readily understood of the three types of curves mentioned"³.

The first problem which arises in the construction of the histogram is the number and size of the intervals to be used. Monroe believes that no fixed rule can be laid down concerning the number of intervals employed, but "in general, it is not wise to have the number of intervals exceed 20 nor be less than 10."⁴ In this preliminary analysis, the intervals used are always less than 20 and at least equal to 10 in number. With respect to the size of the interval, his only recommenda-

1. Cf. Starch, D. Educational Psychology, Macmillan Co., 1924, p. 446.

2. Odell, C.W., op. cit., p. 36.

3. Ibid., p. 42.

4. Monroe, W. S., op. cit., pp. 303-304.

tion is that they be equal.¹ Each frequency distribution here represented has equal intervals.

The next question is twofold: (1) How may these distributions be best described (2) in relation to the problem at hand? Monroe is most clear in regard to question number one: "A frequency distribution may be described in terms of its (1) shape, (2) its central tendency, and (3) its variability or spread."² However, the major concern of the preliminary analysis is one of gross comparison, and not of minute, statistical niceties. Therefore, in spite of the fact that a few measures of central tendency are calculated,³ and that in many cases the standard deviation is worked out and presented for inspection, no attempt is made to contrast the frequency distributions quantitatively.⁴ That is to say, the nature of the initial comparisons is qualitative. Thus, the comparisons are limited to a descriptive evaluation of the abnormally skewed deviations. The second part of the question has already been discussed; scores and marks not approximating or tending to approximate the normal-frequency distribution are designated

1. Ibid., p. 303.

2. Monroe, W. S., op. cit., p. 307.

3. Cf. Tables 1.

4. Direct mathematical comparisons of the attributes measured are attempted by the method of correlation in subsequent chapters.

TABLE 1

Calculation of the median of a frequency distribution
the measures are the average marks for the
first term.

Scale Intervals	f	
90	2	
86	12	
82	14	
78	35	
74	39	
<hr/>		
70	65	
6		
<hr/>		
66	38	
62	42	
58	23	
54	12	
50	<u>3</u>	
N = 285		

$$\text{Md.} = 1 + \frac{\frac{1}{2} N - S}{f}$$

$$= 70 + \frac{\frac{1}{2} 285 - 118}{65} \quad (4)$$

$$= 70 + \frac{24.5}{65} \quad (4)$$

$$= 70 + .38 \quad (4)$$

$$\text{Md.} = 71.5$$

TABLE 2

Computations of the mean of a frequency distribution
the measures are the average for the
first term.

Scale Intervals	f	d	fd
90	2	5	10
86	12	4	48
82	14	3	42
78	35	2	70
74	39	1	39
<hr/>			
70	65	0	+209
<hr/>			
66	38	-1	-38
62	42	-2	-84
58	23	-3	-69
54	12	-4	-48
50	3	-5	-15
<hr/>			
N = 285			-254
			+209
			<hr/>
			285 - 45
			<hr/>
			c = - .158
			<hr/>
			i = 4.
			<hr/>
			ci - - .632
			Ass. M = <hr/> 72.0
			<hr/>
			M = 71.4

TABLE 3

Calculation of the mode of a frequency distribution
the measures are the average marks for the
first term.

Scale Intervals	f
90	2
86	12
82	14
78	35
74	39
<hr/>	
70	65
<hr/>	
66	38
62	42
58	23
54	12
50	3
<hr/>	

N = 285

$$\begin{aligned}
 Z(\text{mode}) &= 3M_d - 2M \\
 &= 3(71.5) - 2(71.4) \\
 Z &= 71.3
 \end{aligned}$$

TABLE 4

Calculation of the standard deviation of a frequency distribution. The measures are the average marks for the first term.

Scale Intervals	f	d	fd	fd ²
90	2	5	10	50
86	12	4	48	192
82	14	3	42	126
78	35	2	70	140
74	39	1	39	39
<hr/>				
70	65	0	+209	
<hr/>				
66	38	-1	-38	38
62	42	-2	-84	168
58	23	-3	-68	207
54	12	-4	-48	192
50	3	-5	-15	75
<hr/>				
N = 285			-254	1227

$$\begin{aligned}
 \text{Sigma} &= \frac{1227}{285} - .025^{\frac{1}{2}} \\
 &= (4.280)^{\frac{1}{2}} \\
 &= 2.07 \text{ (interval units)} \\
 &= \frac{4}{4} \\
 &= 8.28 \text{ (in scale units)}
 \end{aligned}$$

"irregular".

Results of the preliminary analysis. The results of the initial investigation are presented in the following sequence: (1) frequency distributions of the first and second term average and term marks, (2) the frequency distribution of the composite, gross, and individual scores, and (3) a comparison of (1) and (2). In Figure 1 is found a histogram representing the frequency distribution of the 285 average marks of the freshmen class for the first term. Although it is somewhat skewed to the right the shape of this graph fairly well approximates that of the normal frequency curve. The marks cluster about the middle, and slope gradually and symmetrically from the highest point, the center. The mean, or "average", of the marks for the first term is 71.4. The median, or middle term, is 72, and the mode, the mark of greatest frequency, is 70. The standard deviation, the measure of the degree to which the average marks deviate from the arithmetic mean, is 8.28¹ in terms of the items, average marks. The standard deviation of the theoretical frequency distribution is 7.51 which differs from that of the average mark central tendency by .77.²

Figure 2 shows the frequency distribution of the 285 term

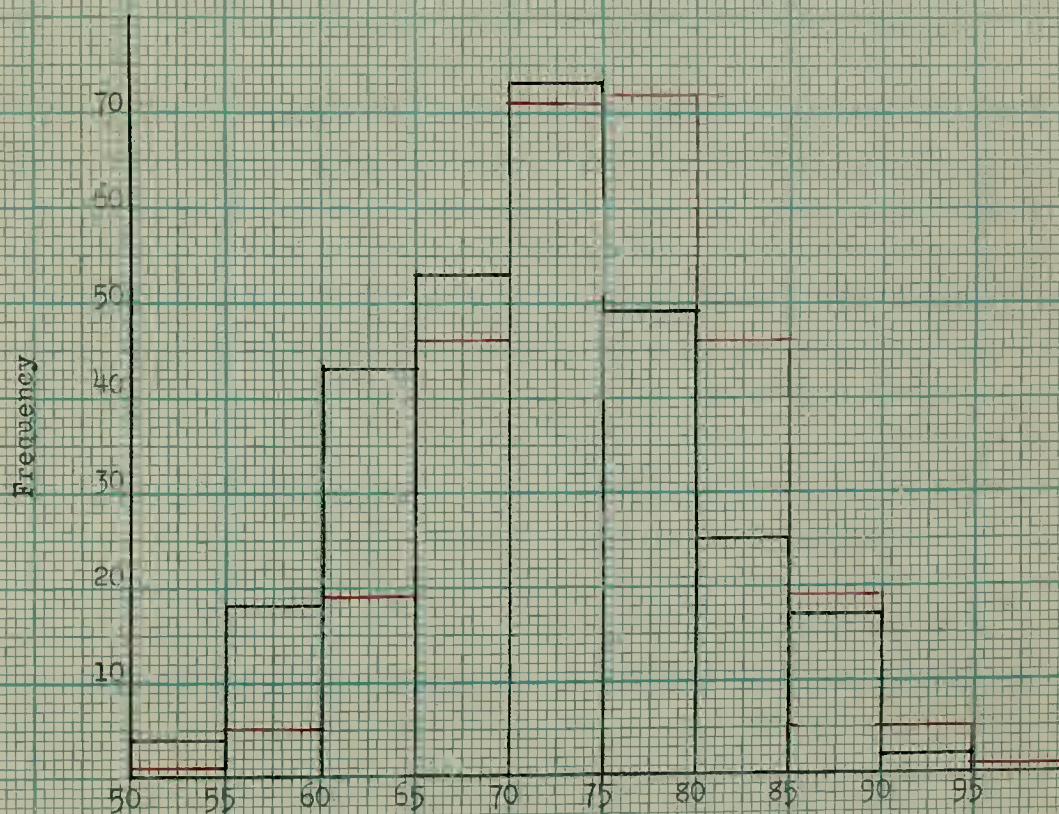
1. The methods used to obtain these results are taken from Mills, F.C., Statistical Methods, Henry Holt and Co., 1924, pp. 111-58; and Odell, C.W., op. cit., pp. 117-145

2. The significance of such close approximations for the immediate study is discussed later (Cf. Summary and interpretations, Chapter III).

FIGURE 1

Histogram showing the actual and theoretical frequency distributions of the first term average marks.

(285 cases)



First term average marks

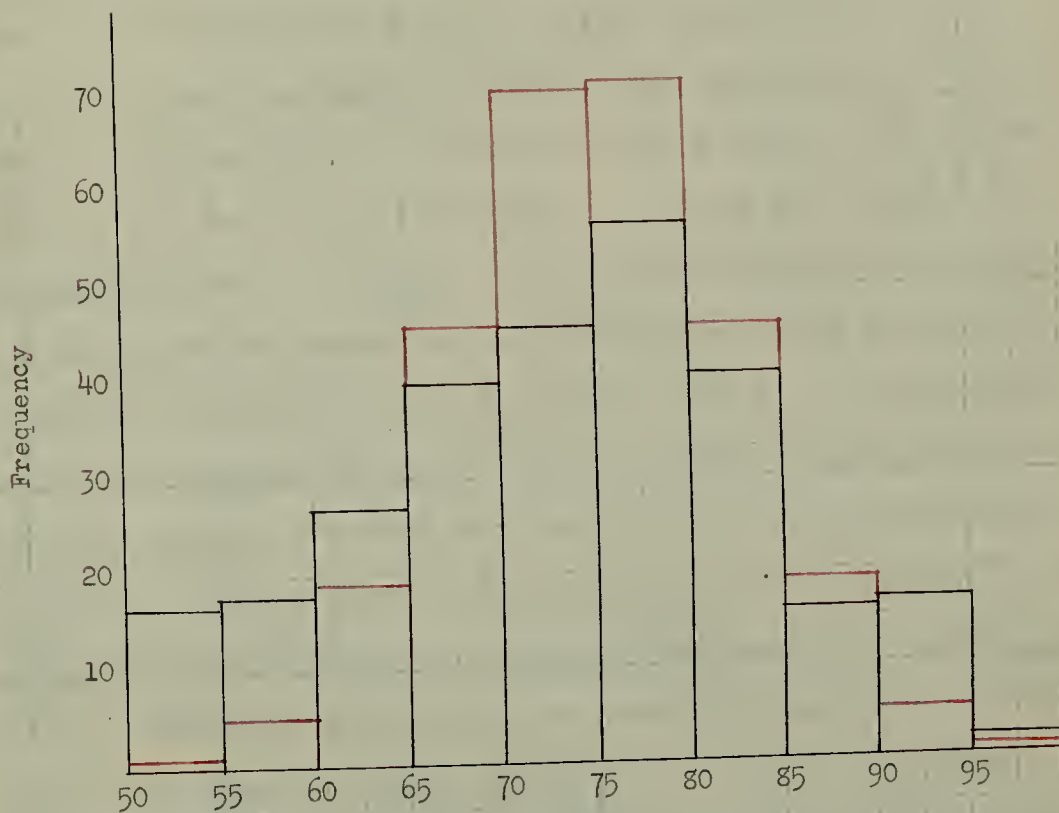
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 2

Histogram showing the actual and theoretial frequency distributions of the first term Orientation marks.

(285 cases)



First term Orientation marks

Represents actual frequency distribution

Represents theoretical distribution

marks for the first term in Orientation. The general contour is fairly symmetrical and only slightly skewed to the left. The number of failures and conditions (35) is rather high when compared with the number of marks above 90 (19 in all). This number (35) includes about 13% of the total marks, and falls 4% without the limit set by Starch¹. The five-division classification of Dr. Starch is 7% E, 24% D, 38% C, 24% B, and 7% A. Obviously, the Orientation marks approximate this classification with the exception of the lowest division.

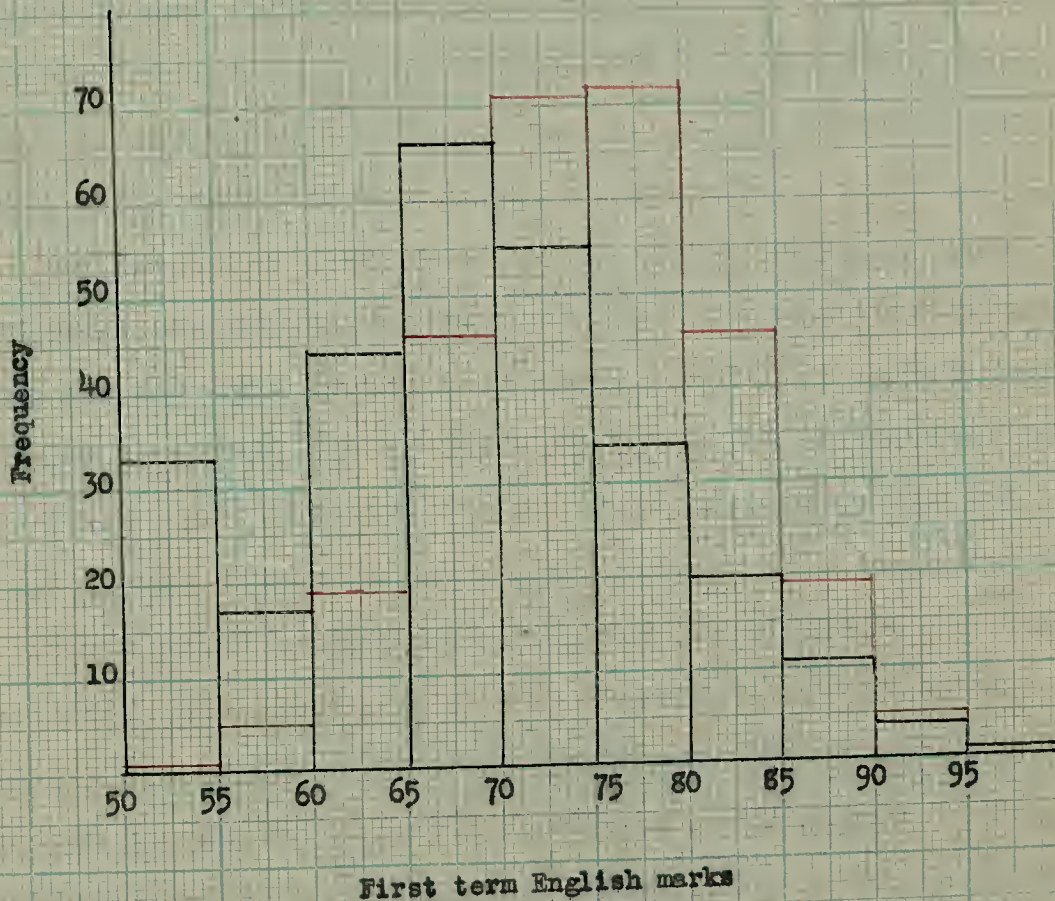
To a more noticeable extent, the English term marks (Figure 3) violate the law of normal distribution. The right-hand side of the curve slopes evenly away from the highest point. The other side of the curve drops quickly and rises abruptly. The standard deviation is 9.6, indicating that over 65% of the marks fall between 68.3 ± 9.6 ($68.3 = \text{mean}$). When compared with the standard deviation of the theoretical distribution for this group, a marked difference is at once perceptible. The actual mean departs 6.7 points from the theoretical mean of 75. If the freshman class is a representative group (as the distribution of the average marks for the same term would seem to indicate), the 33 failures and the 18 conditions can

1. Professor Starch of the University of Wisconsin asserts emphatically that for groups of college students exceeding 100 in number, the demands of the theoretical distribution should not be violated within a 25% deviation above or below the five-division scale. (Cf. Starch, D., Educational Psychology, The Macmillan Co., 1924).

FIGURE 3

Histogram showing the actual and theoretical frequency distributions of the first term English marks.

(285 cases)



Represents actual frequency distribution

Represents theoretical distribution

but suggest that the work is "too difficult."

The curve¹ presented by the Mathematics term marks hardly resembles a normal curve (Figure 4). Of the 259 term marks, 62 are conditions or failures, and approximately 54% fall below the 70 mark. The course is apparently quite difficult.

While not abnormally skewed, the Chemistry term marks (Figure 5) are distributed very unevenly. To offset the 42 failures and conditions, there are but 11 marks above 90. Possibly, the abnormal rise of the 50 and 85 columns accentuates the lowness of the 55 and 90 columns.

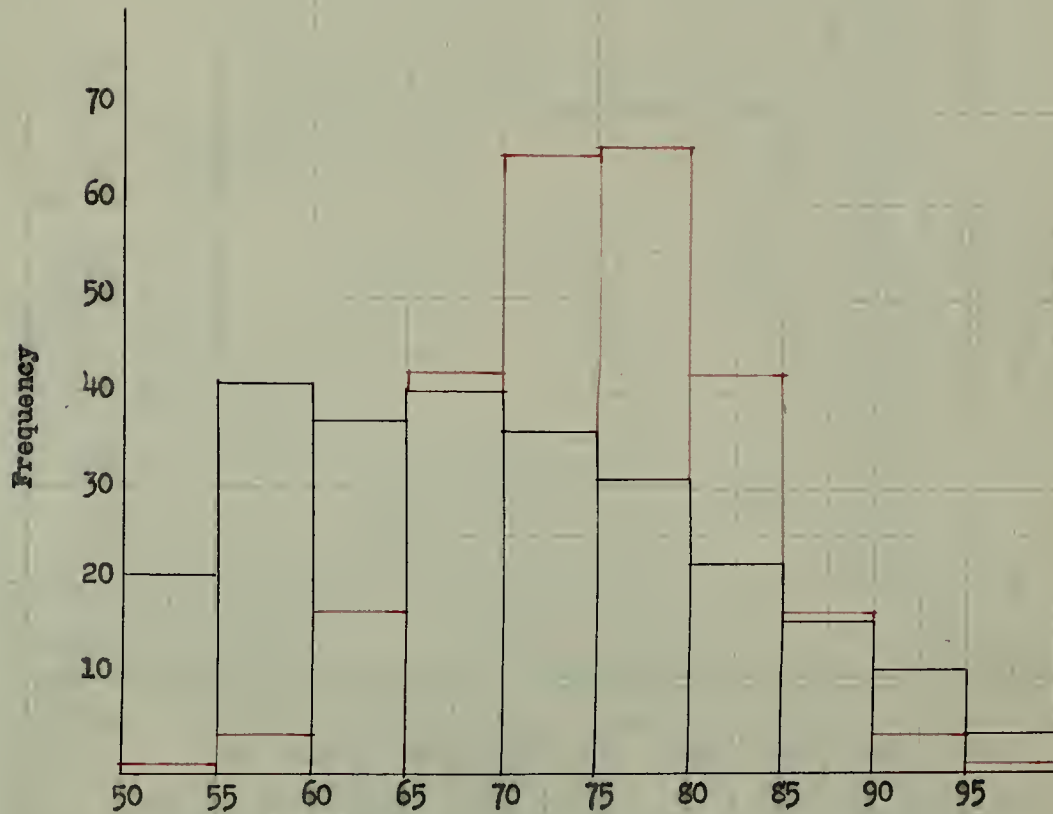
The frequency distributions for the French and German term marks are depicted separately and in composite (Figures 6, 7, and 8). The French term marks are fairly symmetrical, although the mean, 71, is somewhat below the theoretical mean, 75. There are 3 failures and 3 conditions, but these are offset by the 5 marks in the upper fifth. With reservations, Dr. Starch would probably approve the distribution. An inspection of the German term mark distribution reveals two distinct groupings: one about the 55-60 columns, and the other at the 80-85 columns. Examining the measures of central tendency, it is found that the majority of marks range from 70 to 95, while the remainder approximately 39%, fall below 70. In composite, however, the French and German term

1. Henceforth, the curves are described and interpreted qualitatively.

FIGURE 4

Histogram showing the actual and theoretical frequency distributions of the first term Mathematics marks.

(259 cases)



First term Mathematics marks

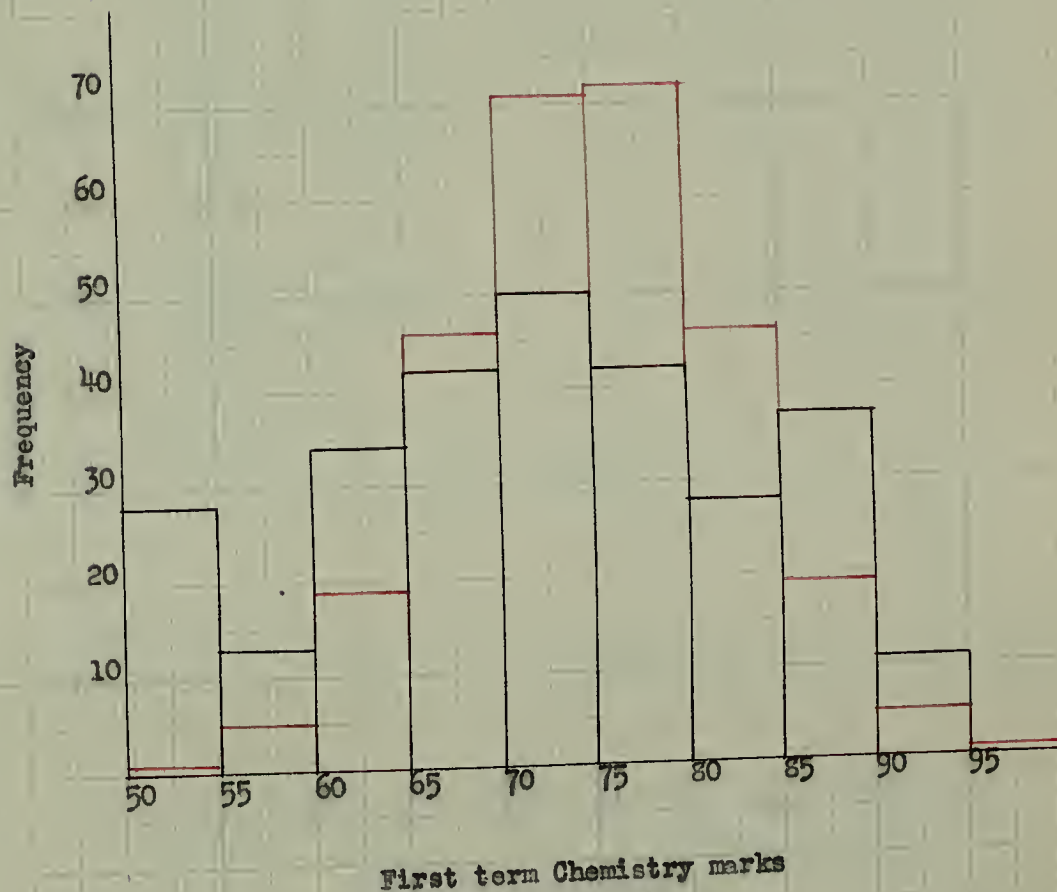
Represents actual frequency distribution.

Represents theoretical distribution.

FIGURE 5

Histogram showing the actual and theoretical frequency distributions of the first term Chemistry marks

(285 cases)



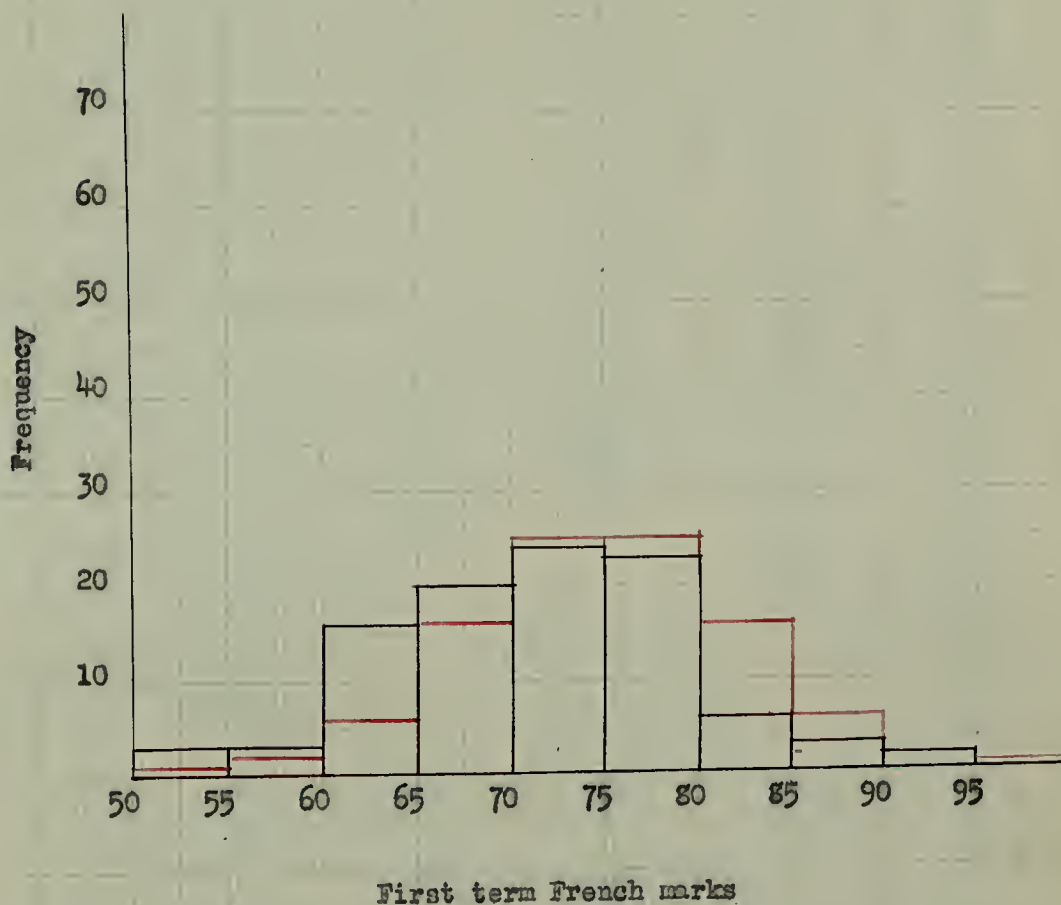
Represents actual frequency distribution.

Represents theoretical distribution.

FIGURE 6

Histogram showing the actual and theoretical frequency distributions of the first term French marks.

(100 cases)



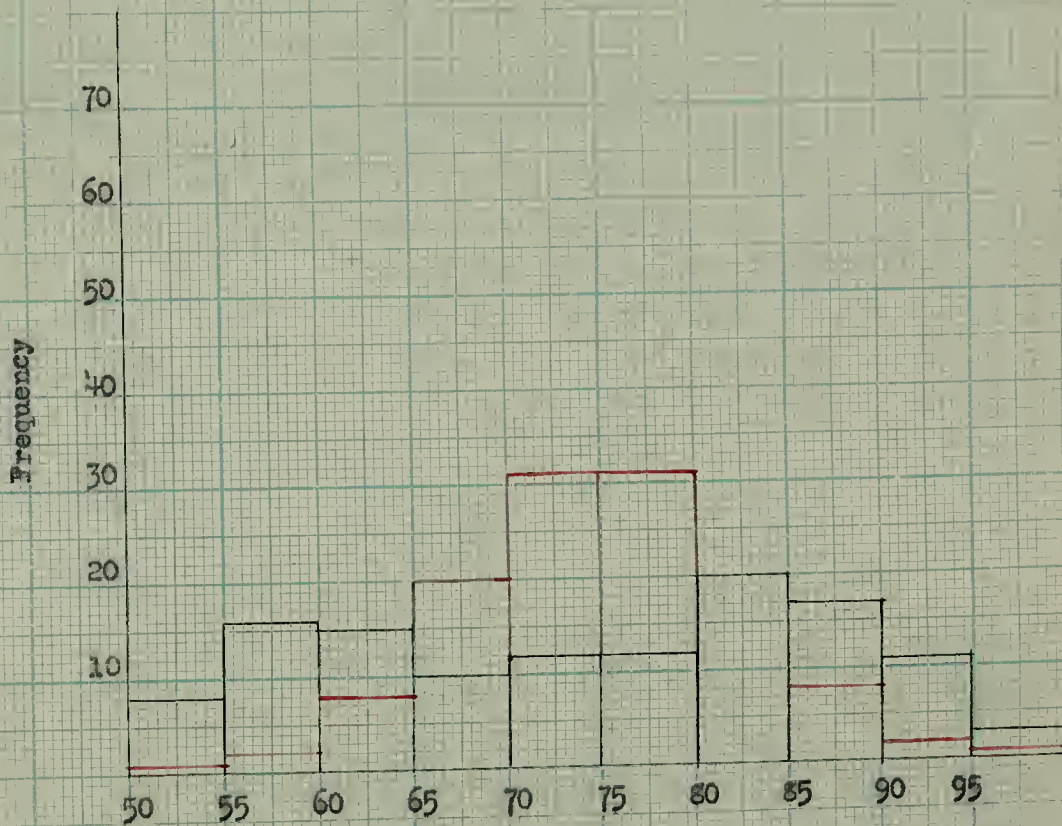
Represents actual frequency distribution.

Represents theoretical distribution.

FIGURE 7

Histogram showing the actual and theoretical frequency distributions of the first term German marks.

(124 cases)



First term German Marks.

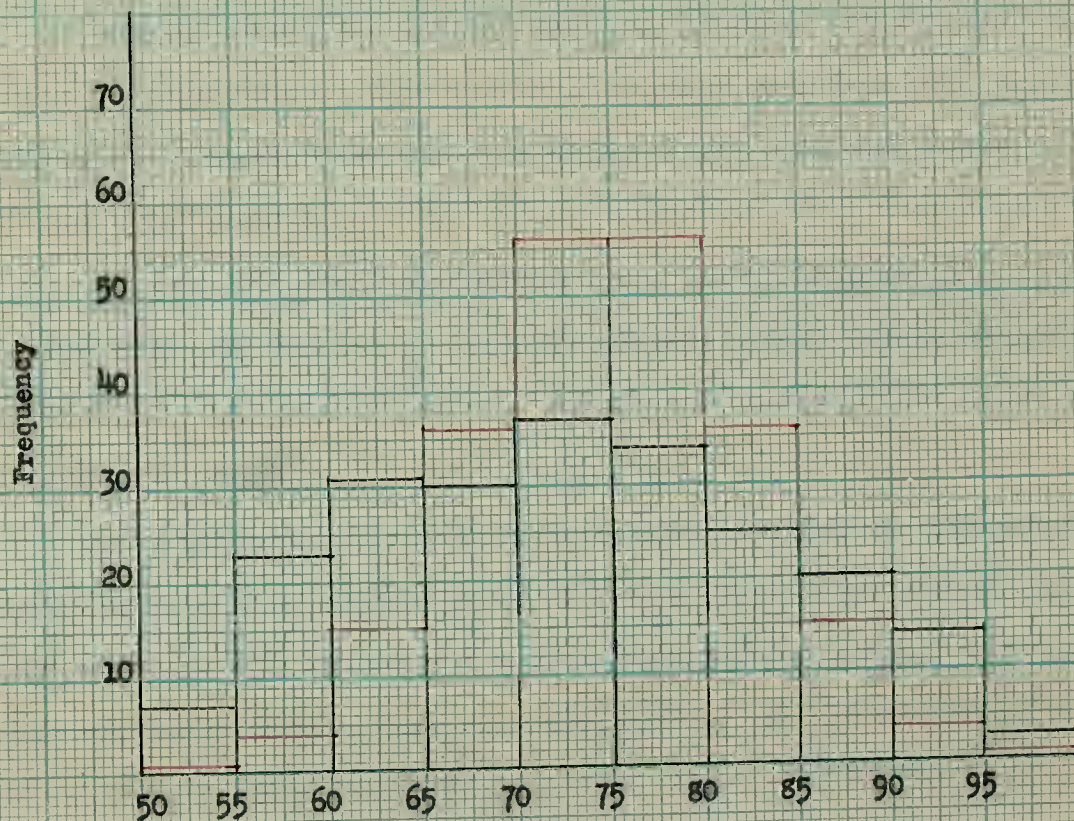
Represents actual frequency distribution.

Represents theoretical distribution.

FIGURE 8

Histogram showing the actual and theoretical frequency distributions of the first term German and French marks

(224 cases)



First term German and French marks

Represents actual frequency distribution.

Represents theoretical distribution.

marks assume a fairly symmetrical frequency distribution which would probably meet the approbation of the majority of the schoolmen who understand and adhere to the demands set up by the law of normal distribution.

The frequency distribution for the second term average marks (Figure 9) is skewed to the right. The majority of the marks lie between 65 and 80, the two end classes contain no measures, and but 11 average marks are below 60.

The Orientation second term marks tend "to spread" more than is necessary. The central columns are too low, while those on either side are too high; e.g., columns 2 and 3, and 8 and 9 (Figure 10).

Of the 253 English second term marks (Figure 11), 39 are conditions or failures. In the upper fifth, i.e., above 90, there are only 2 term marks. Freshman English is apparently quite difficult.

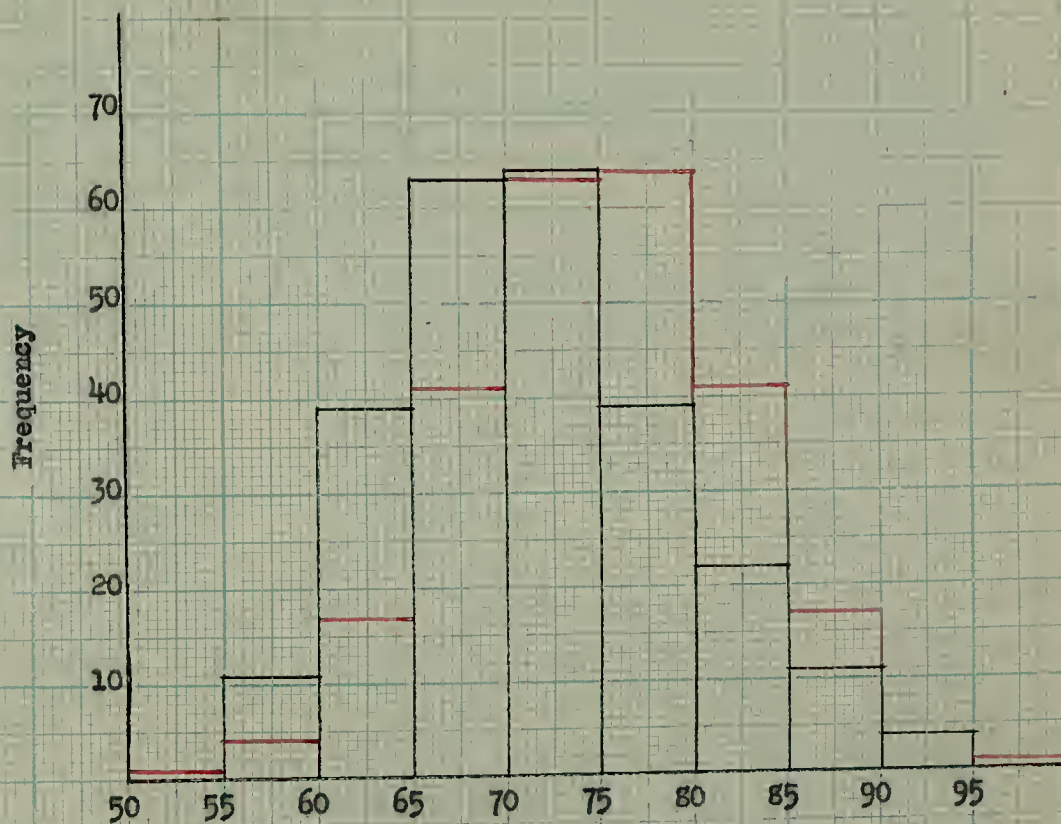
In Figure 12, the frequency distribution of the second term Mathematics marks shows the same tendency toward abnormal skewness as did the first term Mathematics marks. The majority of these marks, 126, are below 70. Fifty are conditions or failures (the lower fifth), and the upper fifth division (90-100) contains about 9% of the 234 measures.

The second term marks for Chemistry vary considerably from the theoretical curve of symmetry (Figure 13). There are but 20 marks above 84, while 77 of the total (245) fall below 65.

FIGURE 9

Histogram showing the actual and theoretical frequency distributions of the second term average marks.

(253 cases)



Second term average marks.

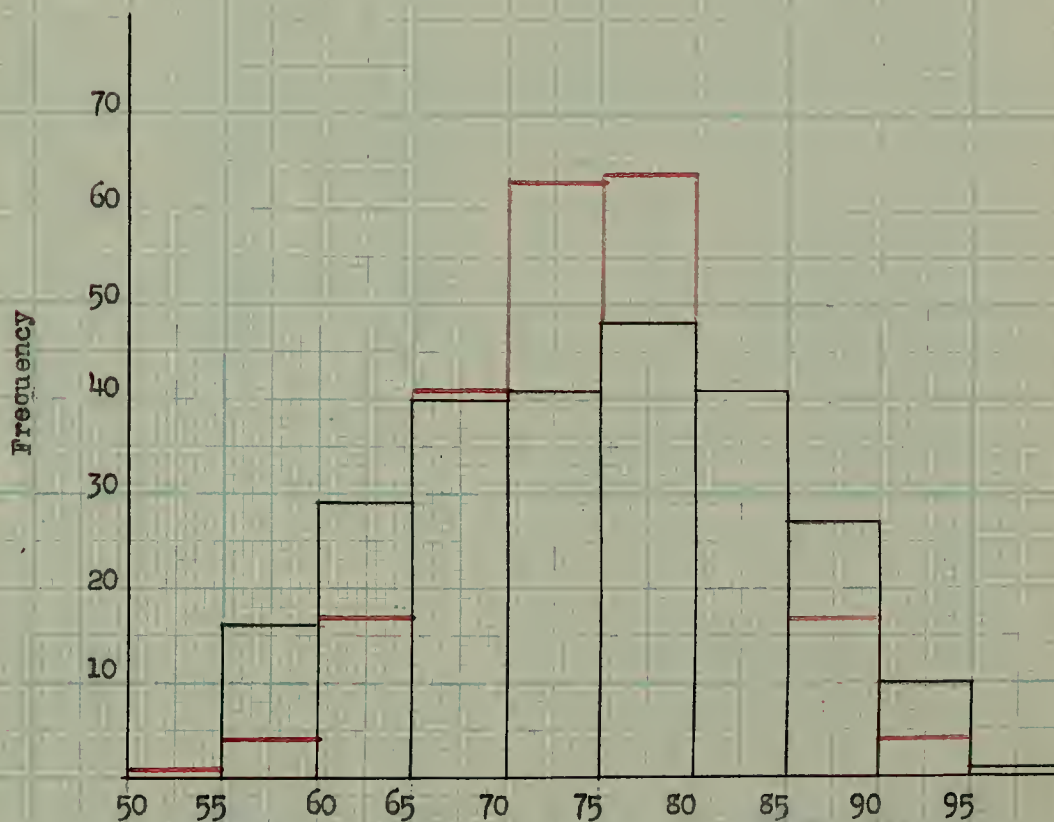
Represents actual frequency distribution.

Represents theoretical distribution.

FIGURE 10

Histogram showing the actual and theoretical frequency distributions of the second term Orientation marks.

(253 cases)



Second term Orientation marks

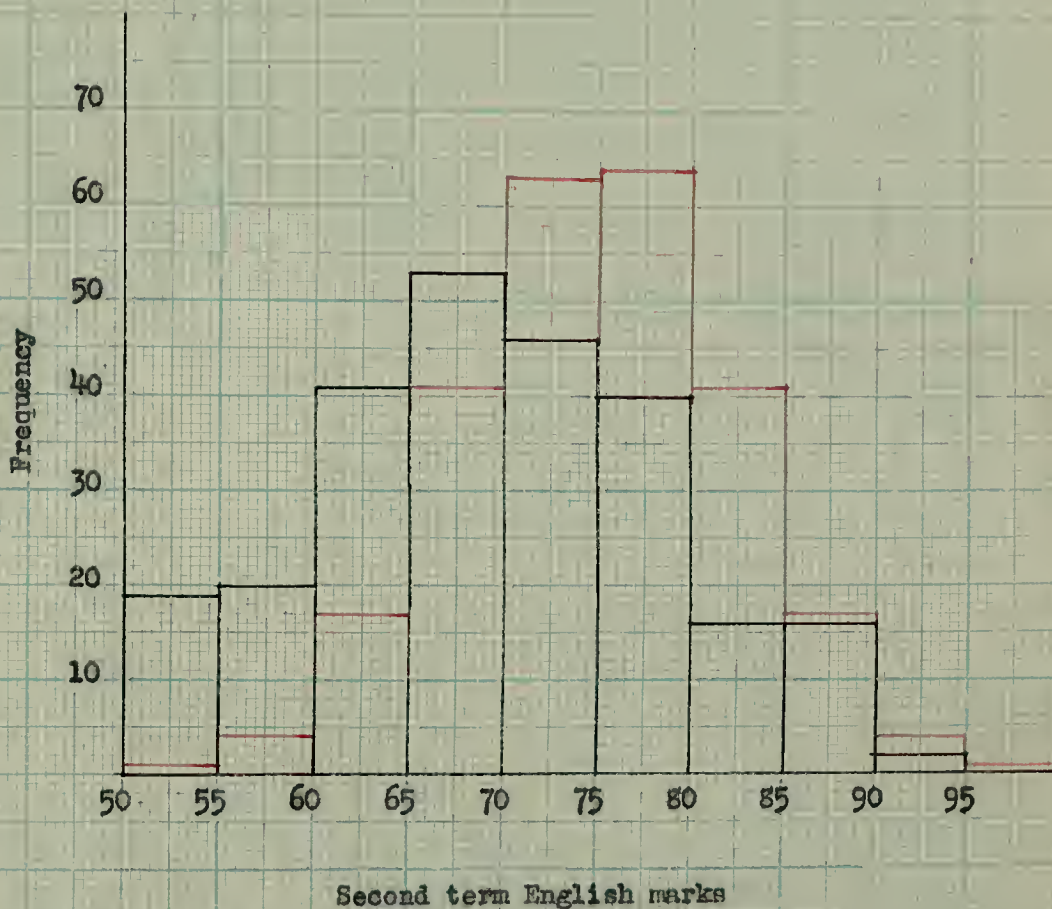
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 11

Histogram showing the actual and theoretical frequency distributions of the second term English marks.

(253 cases)



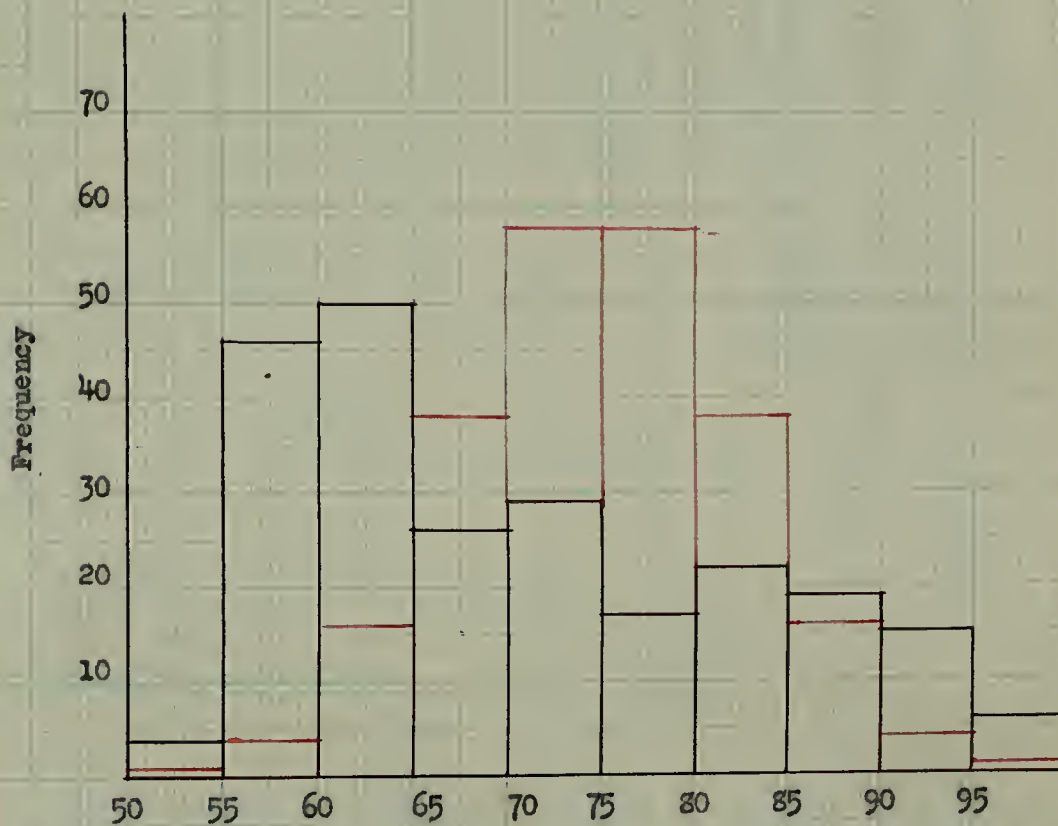
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 12

Histogram showing the actual and theoretical frequency distributions of the second term Mathematics marks.

(234 cases)



Second term Mathematics marks

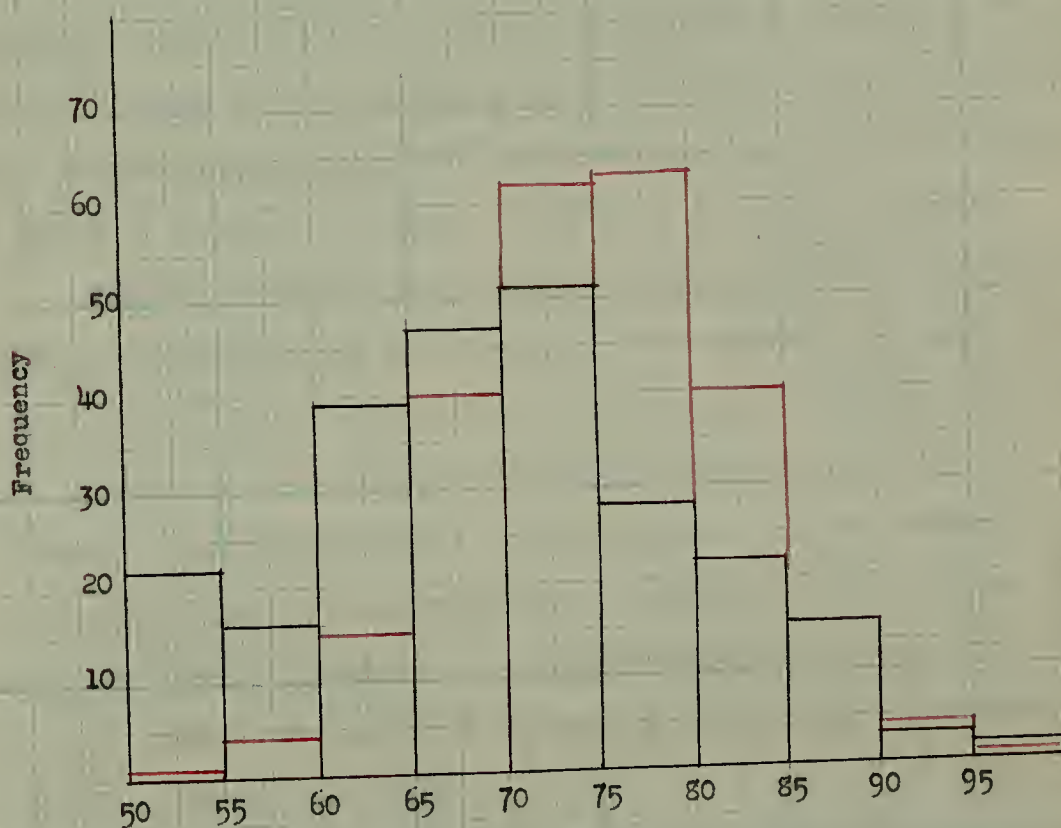
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 13

Histogram showing the actual and theoretical frequency distributions of the second term Chemistry marks.

(245 cases)



Second term Chemistry marks

Represents actual frequency distribution

Represents theoretical distribution

A superficial examination of this frequency distribution, seems to indicate that here is another "too difficult" course of study.

The German and French mark frequency distributions are presented in Figures 14 and 15 respectively. The German marks are distributed irregularly. Although for so few cases, 103, the curve of error is not "abnormally skewed". Again, the 91 French marks (with the exception of the upper-fifth class) exhibit a fairly normal distribution. This is the only distribution which reveals any consistent degree of conformity. All others vary substantially from the theoretical, normal - frequency curve -directly, from term to term, or both.

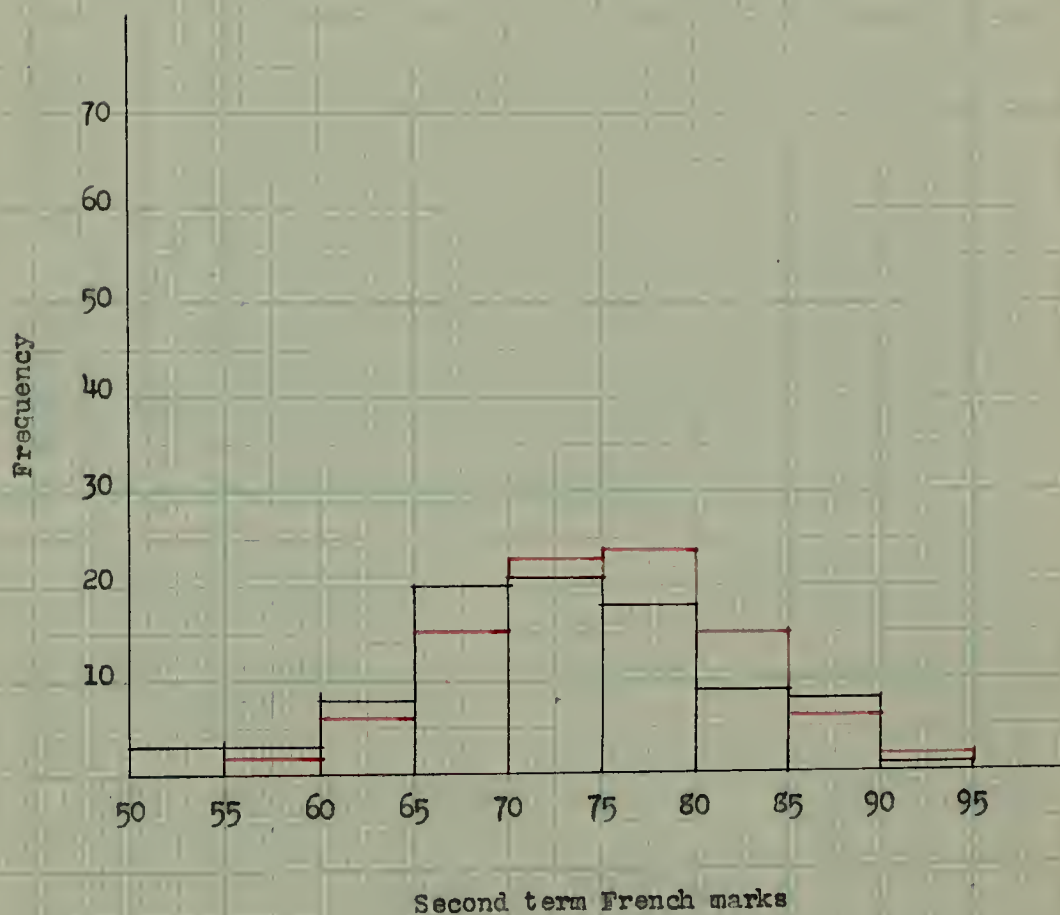
Figure 16 shows the actual and the theoretical normal frequency distributions of the Composite scores. It may be seen immediately that, while the actual curve is fairly symmetrical, the scores do not cluster so fully about the center as do those of the hypothetical distribution. It is probable that an elimination of, or an easier substitution for, a few of the more difficult portions of the three tests, along with increase in the number of problems in each would render the curve of the actual scores more normal.

The gross scores of the Army Group Examination Alpha (Figure 17) are definitely skewed to the left. Instead of forming at the middle, the majority of the scores are spread too freely to the right. Obviously, this test is "too easy" for the

FIGURE 14

Histogram showing the actual and theoretical frequency distributions of the second term French marks.

(91 cases)



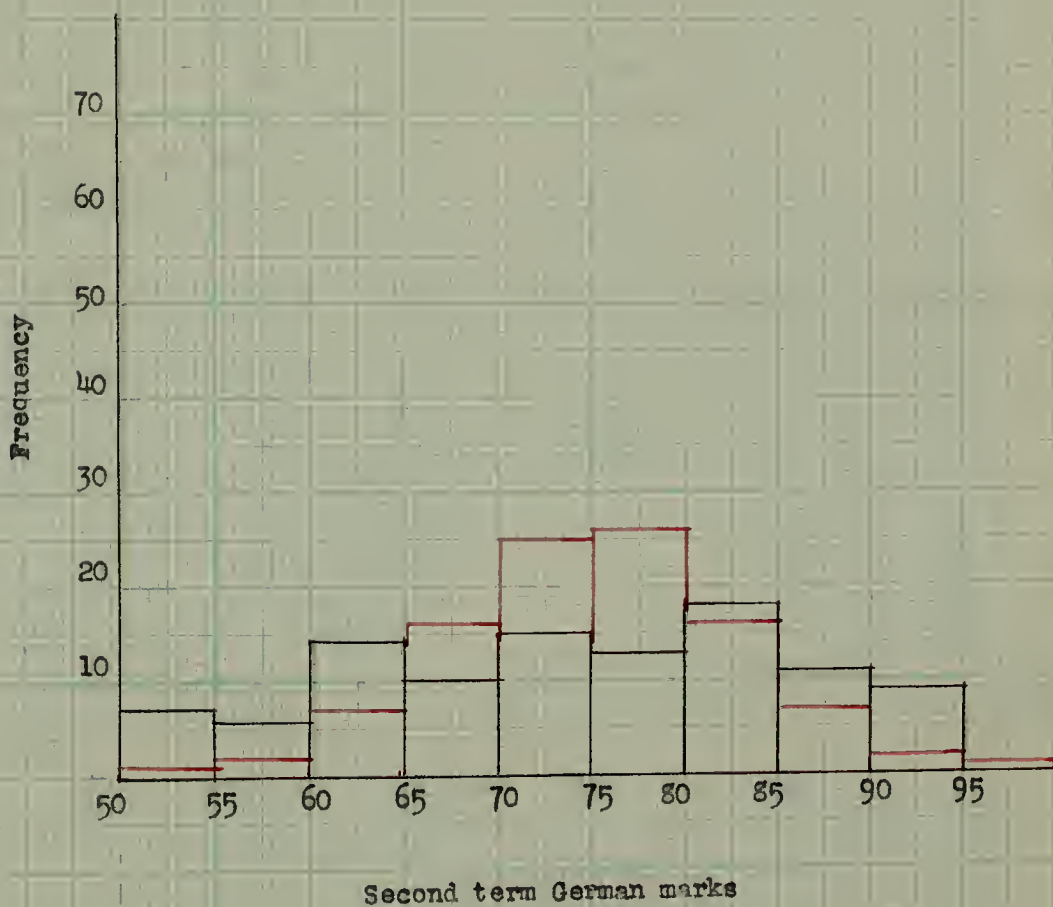
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 15

Histogram showing the actual and theoretical frequency distributions of the second term German marks.

(103 cases)



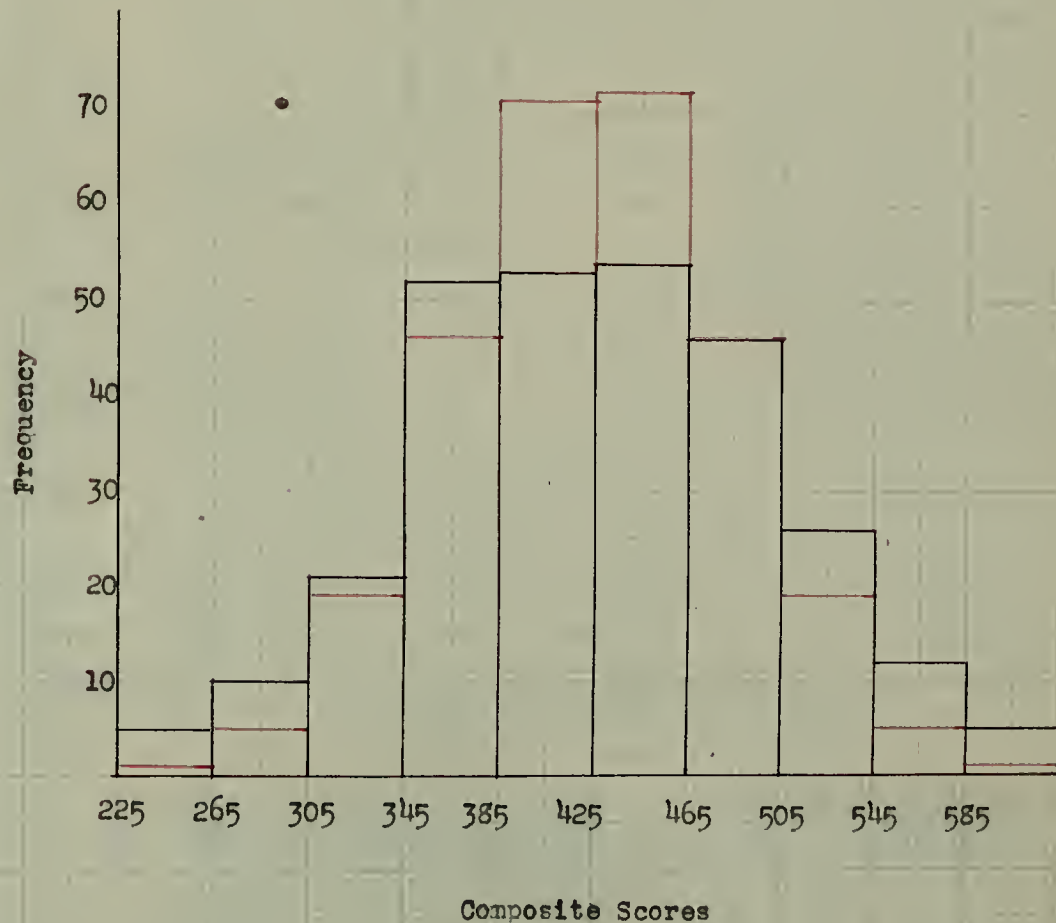
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 16

Histogram showing the actual and theoretical frequency distributions of the Composite Scores.

(285 cases)



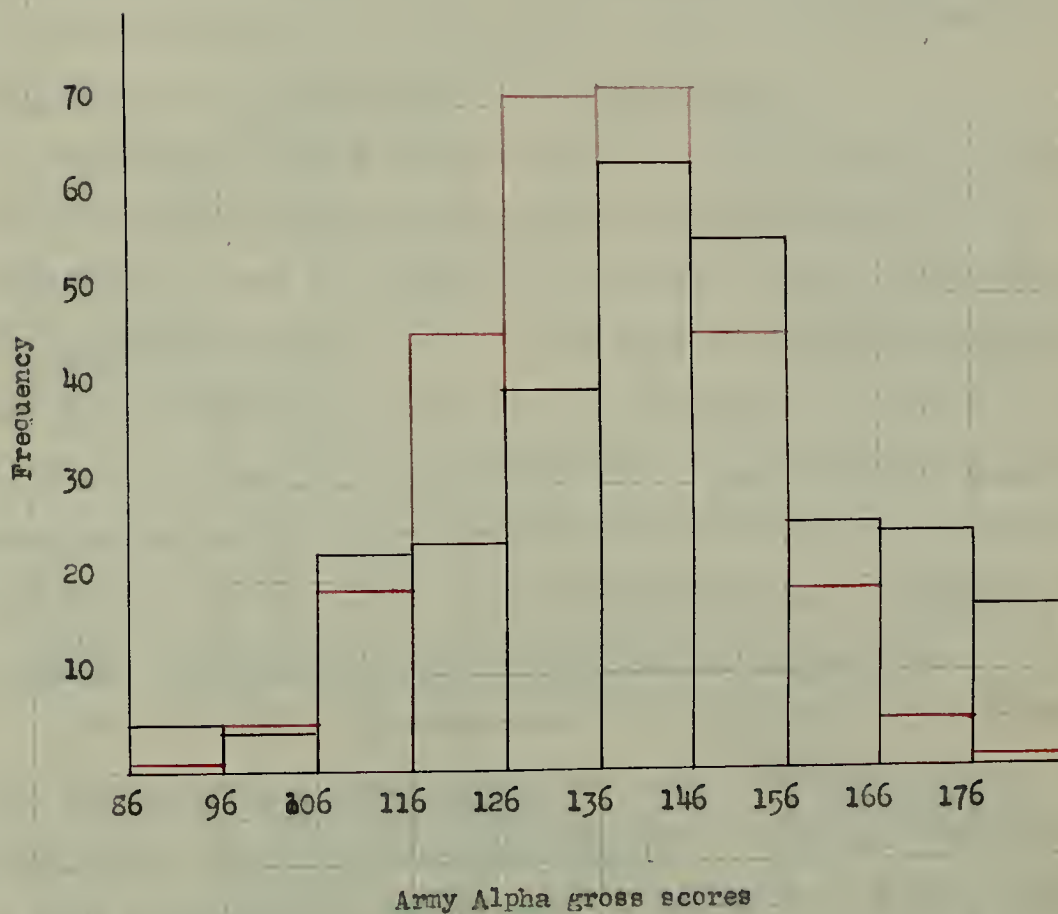
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 17

Histogram showing the actual and theoretical frequency distributions of the Army Alpha gross scores.

(285 cases)



Represents actual frequency distribution

Represents theoretical distribution

entering freshmen.

Generally speaking, the Psychological Examination gross scores tend to cluster about the mean (Figure 18). The curve is fairly symmetrical, although somewhat "flat" at the center. However, since the same number of cases is used to determine the actual and theoretical distributions of these gross scores, and since there is a rather marked separation in existence, the conclusion is inevitable: the Psychological Examination should be so revised as to foster a more normal distribution, or to fit better the average mark distribution.

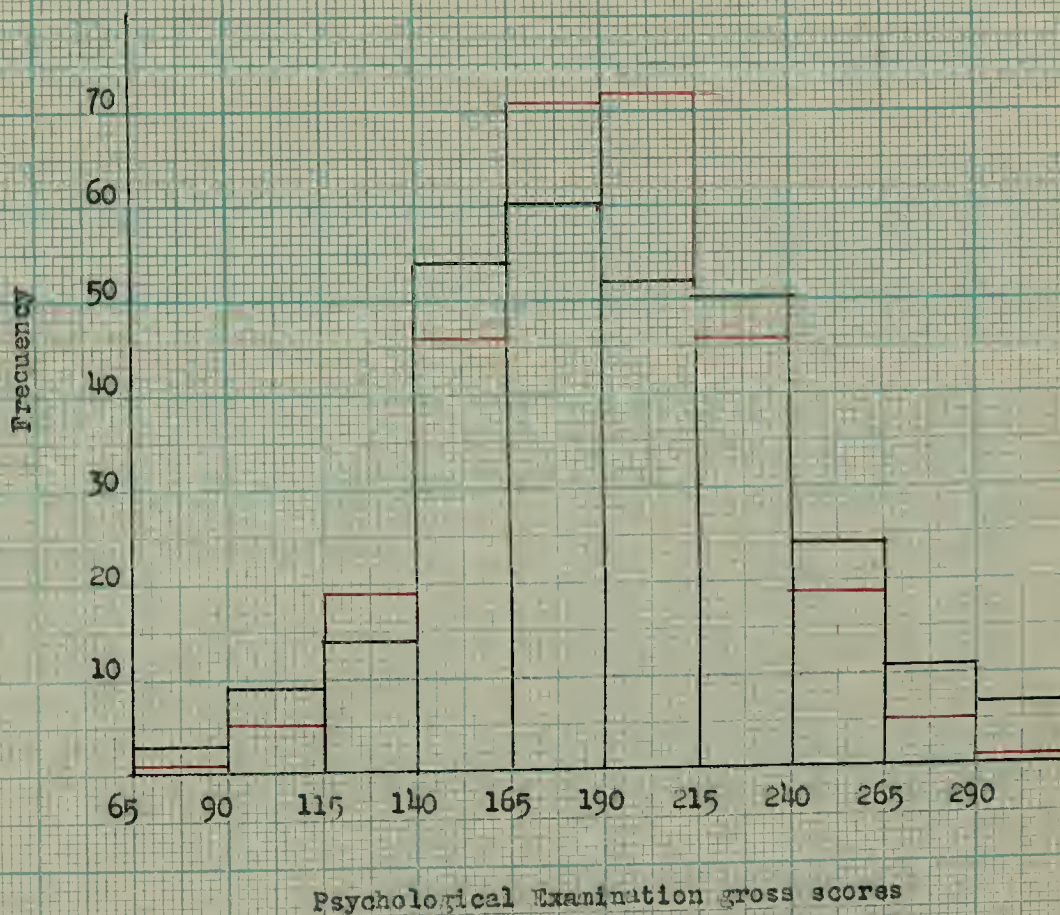
Figure 19 shows the frequency distribution of the Massachusetts State College Scholastic Aptitude Test gross scores. Here, too, since the scores are somewhat more in evidence on the right-hand side of the curve than is desirable, the test may be considered a trifle easy. Moreover, it should be noted that the frequency distribution of the gross scores of this test approximates coincidence with the theoretical curve of error less closely than does the Psychological Examination, and more so than does the Army Group Examination Alpha.

The frequency distributions of the individual scores for the Psychological Examination are shown in Figures 20-24. The individual scores of the Completion test (Figure 20) evidence a good distribution to the left of the center, while on the right of the highest column a marked discrepancy occurs. The actual curve may represent a bi-modal distribution.

FIGURE 18

Histogram showing the actual and theoretical frequency distributions of the Psychological Examination gross scores,

(235 cases)



Psychological Examination gross scores

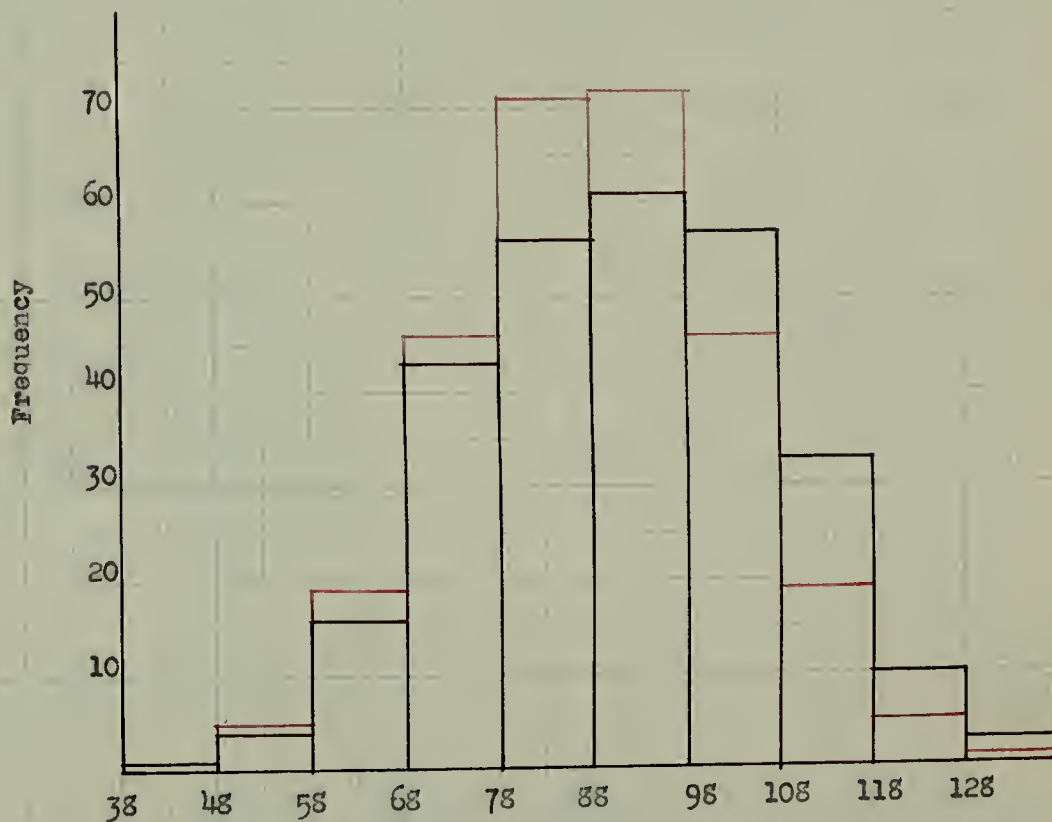
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 19

Histogram showing the actual and theoretical frequency distributions of the Massachusetts State College Scholastic Aptitude Test gross scores.

(235 cases)



Massachusetts State College Scholastic
Aptitude Test gross scores

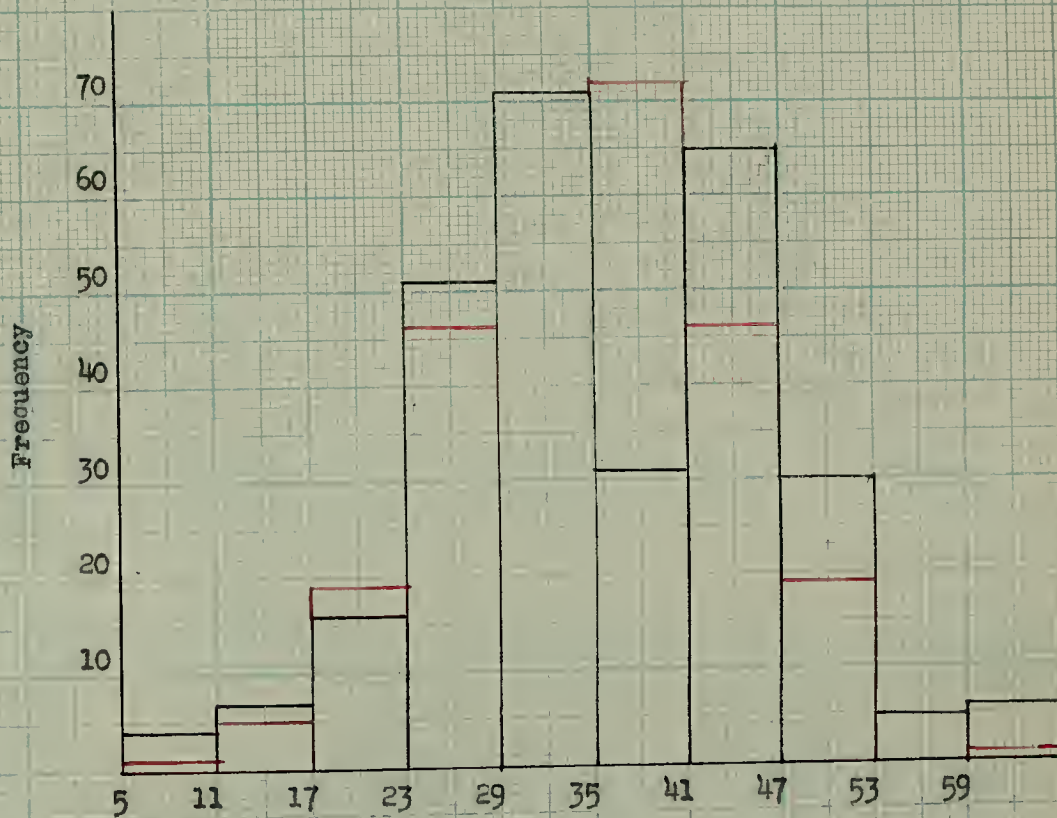
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 20

Histogram showing the actual and theoretical frequency distributions of the Completion section individual scores (Psychological Examination)

(285 cases)



Completion section individual scores
(Psychological Examination)

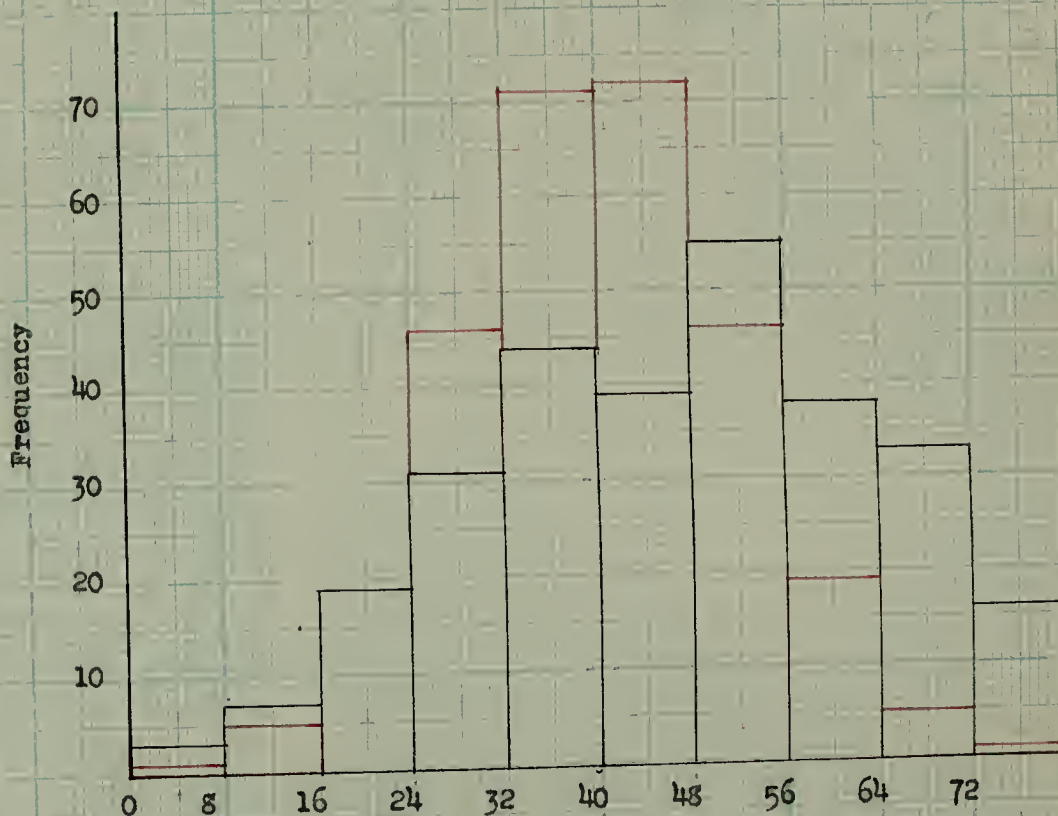
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 21

Histogram showing the actual and theoretical frequency distribution of the Opposites section individual scores (Psychological Examination)

(285 cases)



Opposites section individual scores
(Psychological Examination)

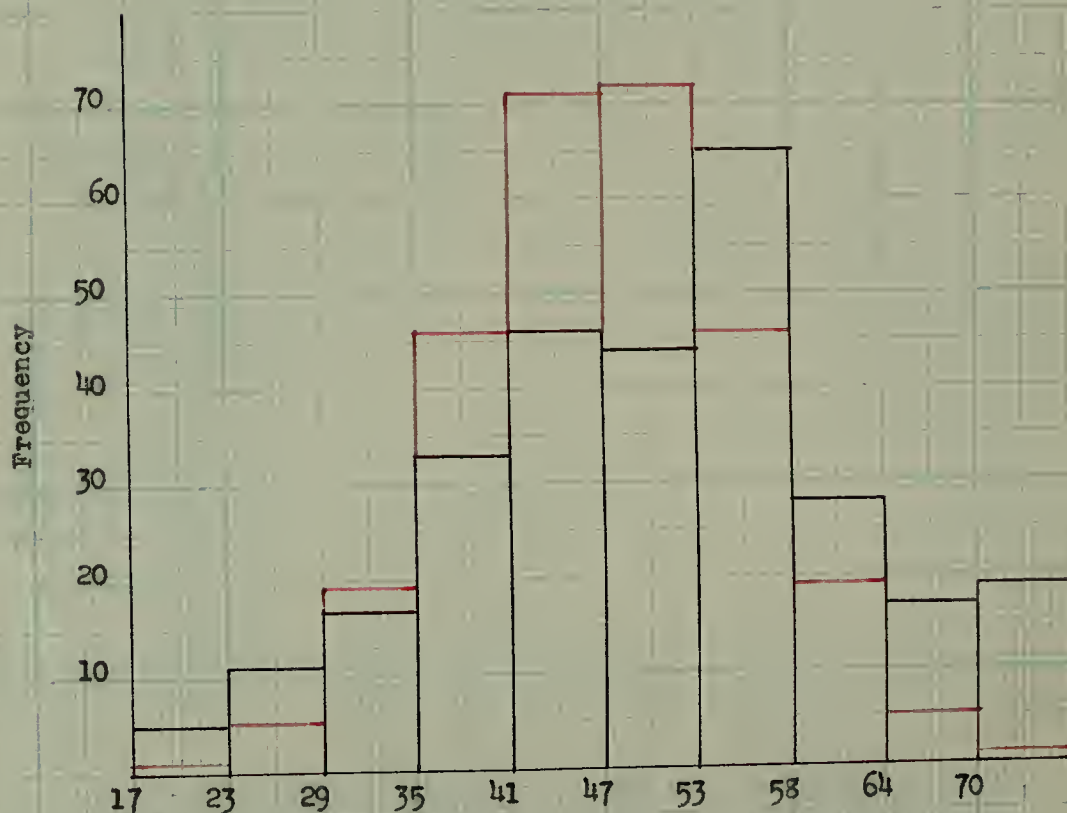
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 22

Histogram showing the actual and theoretical frequency distributions of the Artificial Language Section individual scores (Psychological Examination)

(285 cases)



Artificial Language Section individual scores
(Psychological Examination)

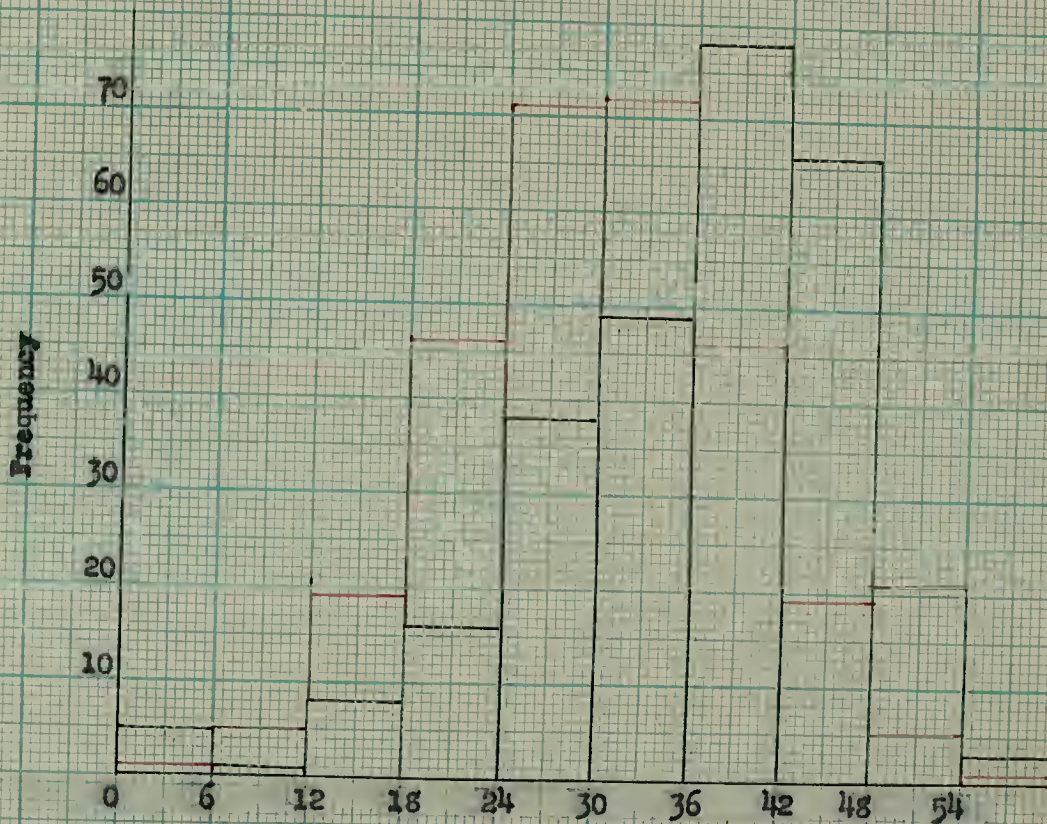
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 23

Histograms showing the actual and theoretical frequency distributions of the Analogies section individual scores (Psychological Examination)

(285 cases)



Analogies section individual scores
(Psychological Examination)

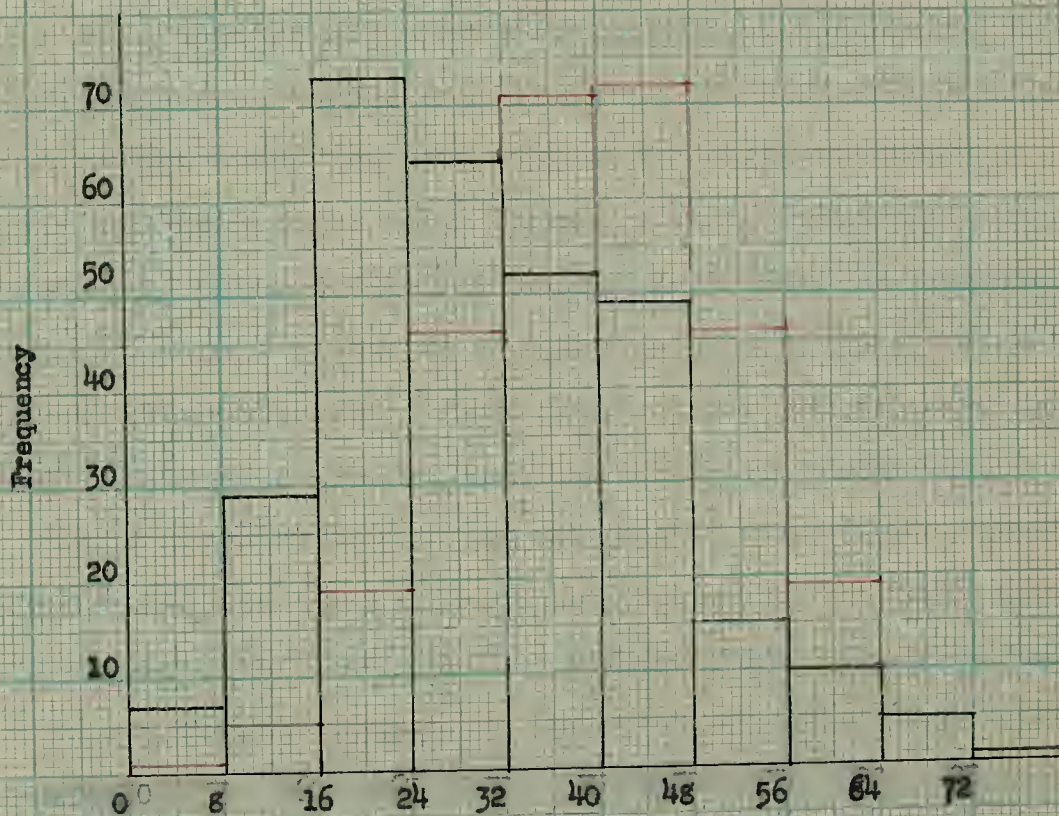
Represents actual frequency distribution.

Represents theoretical distribution.

FIGURE 24

Histograms showing the actual and theoretical frequency distributions of the Arithmetic section individual scores (Psychological Examination).

(285 cases)



Arithmetic section individual scores
(Psychological Examination)

Represents actual frequency distribution.

Represents theoretical distribution.

Figure 21 reveals a left-skewed curve which deviates considerably from the normal pattern. It is decidedly too easy. The unusually large number of cases occurring after the sixth column suggests that the Opposites section could well be extended in length (number of items).

The shape of the Artificial Language section individual scores distribution (Figure 22) indicates that the items could profitably be increased in number and (slightly) in difficulty.

In a similar way, the distribution of the Analogies section individual scores are skewed to the left (Figure 23). The curve rises evenly to the seventh and eighth columns and then drops precipitately. The unusually large number of cases appearing in columns seven and eight indicates, in general, that the test should either be rendered more difficult, or the number of items extended. Regardless of the method used, the possibility of a more symmetrical curve is increased.

The distribution of the Arithmetic individual scores (Figure 24), on the contrary, is shunted to the left -indicating at once that the test is too difficult. The fact that the third and fourth columns of the actual distributions are approximately as tall as the fifth and sixth columns of the hypothetical curve definitely suggests its difficulty. A remedy could here be effected by adding a few more problems similar in nature to those most successfully completed by the majority of the students.

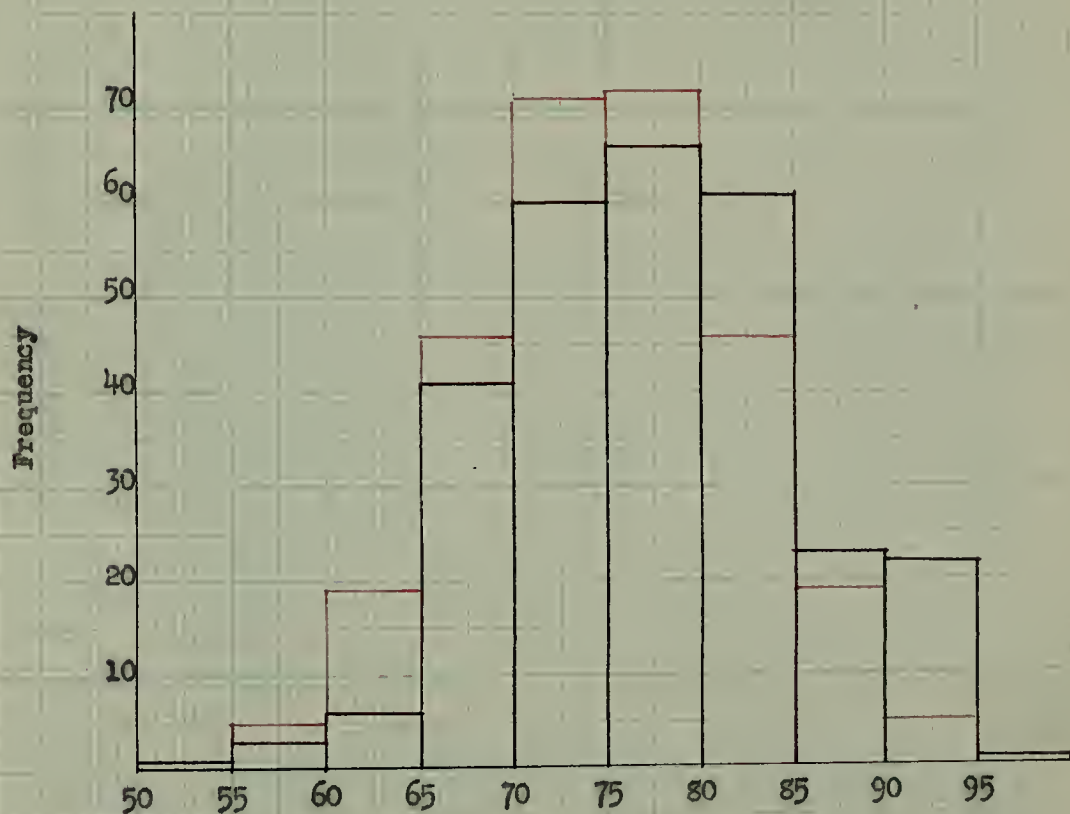
A study of the frequency distribution curve for the individual scores obtained on the Botany section of the Massachusetts State College Scholastic Aptitude Test, (Figure 25) reveals that it deviates from the normal. It is noticeably skewed to the left, and is thus rendered too easy. The second and third columns contain too few cases, while the seventh and ninth are quite in excess of the number permitted to each by the law of normal distribution. It is possible that an extension in regard to the number of items included in this section would serve to restore the shape of the histogram to normalcy. There is an alternative here, however. A minute or two can be deducted from the time allotted to either the study period or the period of examination. A brief trial-and-error experiment would possibly give the indication as to which method should be used with optimal results.

No detailed scrutiny is necessary to note that the frequency distribution of the Mathematics section individual scores (Figure 26) is abnormally "warped". The section is positively too hard for entering freshmen. The first two columns contain 48 actual measures; to these columns, the theoretical normal curve permits only 6. The third and four columns evidence discrepancies of the same kind and in the same order. The curve slopes from the first to the second, and similarly, from the third column to the fourth. Then, too, there is the fact that the third column is taller than the

FIGURE 25

Histogram showing the actual and theoretical frequency distributions of the Botany section individual scores (Massachusetts State College Scholastic Aptitude Test)

(285 cases)



Botany section individual scores (Massachusetts State College Scholastic Aptitude Test)

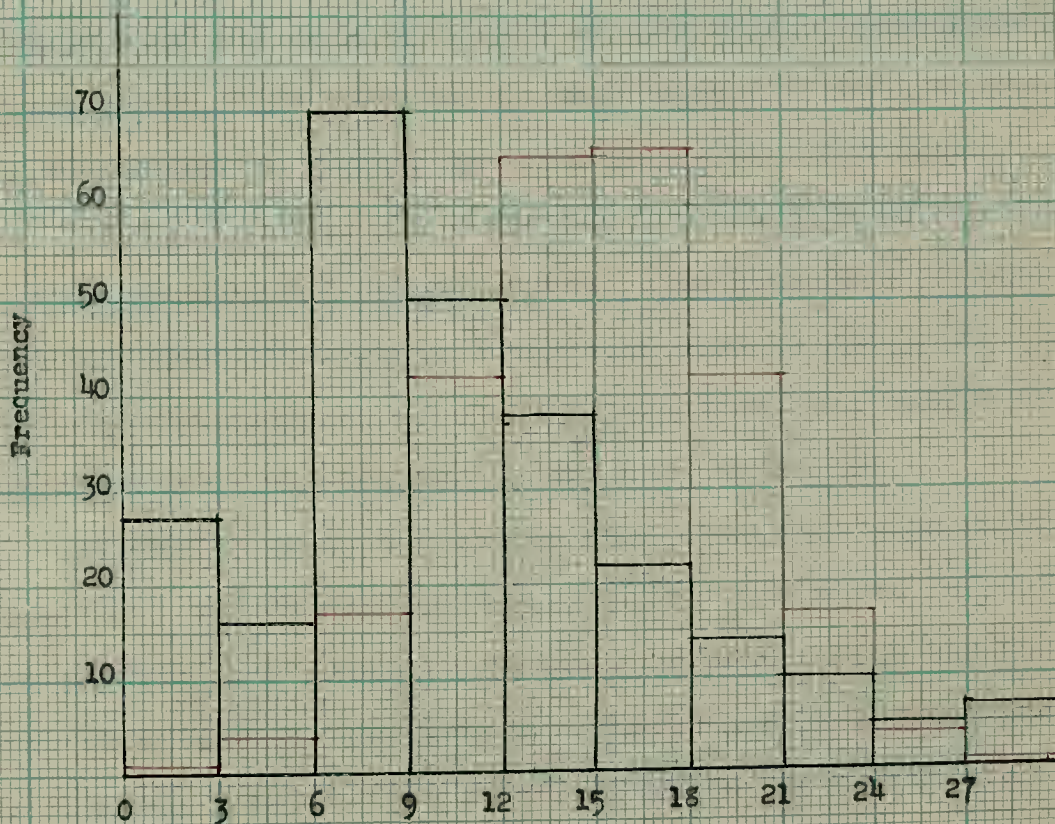
Represents actual frequency distribution.

Represents theoretical distribution.

FIGURE 26

Histogram showing the actual and theoretical frequency distributions of the Mathematics section individual scores (Massachusetts State College Scholastic Aptitude Test).

(259 cases)



Mathematics section individual scores (Massachusetts State College Scholastic Aptitude Test)

Represents actual frequency distribution

Represents theoretical distribution

middle columns of the normal distribution. If desirable, the difficulty of the section can be alleviated by an extension of time or an increase in the number of less difficult problems.

The opposite situation holds concerning the frequency distribution of the Chemistry section individual scores (Figure 27). Here, the curve rises very slowly for the first three bars, and then very rapidly reaches the maximum in column number 7. The remaining three bars indicate that the measures represented are much in excess of the normal apportionment. If necessary, this test can be made more difficult by increasing the number of parts, and decreasing the time limit.

Figure 28 shows the frequency distribution of the Power Reading section individual scores. The histogram indicates that the test is slightly "easy" and that it should be altered to force the measures to cluster more about the center -thus raising the center columns to the desired height.

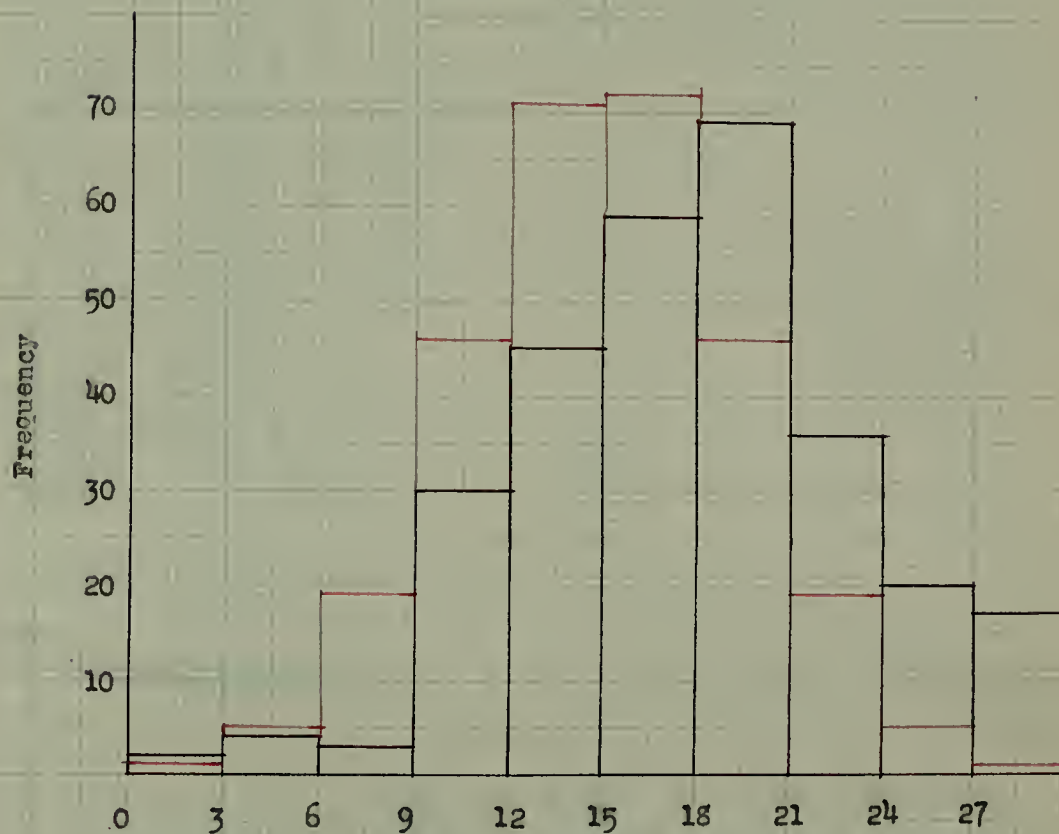
Figure 29 shows the actual distribution of the individual scores for the Memory Reading section. The curve is roughly symmetrical and, while it deviates in places from the theoretical distribution, it assumes a fairly normal shape. The test is possibly a shade "too easy". Perhaps the addition of a single item to each part would aid in bringing about a better approximation to the theoretical curve of error.

The histogram representing the frequency distribution of the Artificial Language individual scores (Figure 30) indicates

FIGURE 27

Histogram showing the actual and theoretical frequency distributions of the Chemistry section individual scores (Massachusetts State College Scholastic Aptitude Test)

(285 cases)



Chemistry section individual scores (Massachusetts State College Scholastic Aptitude Test)

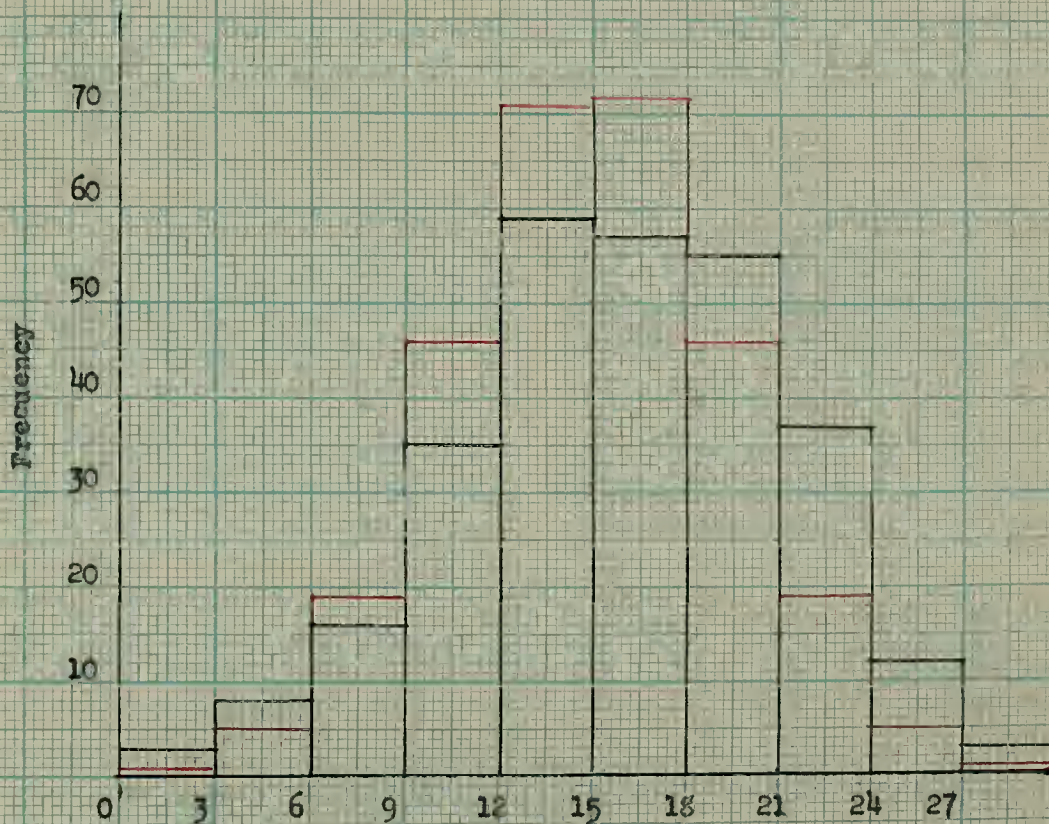
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 28

Histogram showing the actual and theoretical frequency distributions of the Power Reading section individual scores (Massachusetts State College Scholastic Aptitude Test)

(285 cases)



Power Reading section individual scores (Massachusetts State College Scholastic Aptitude Test)

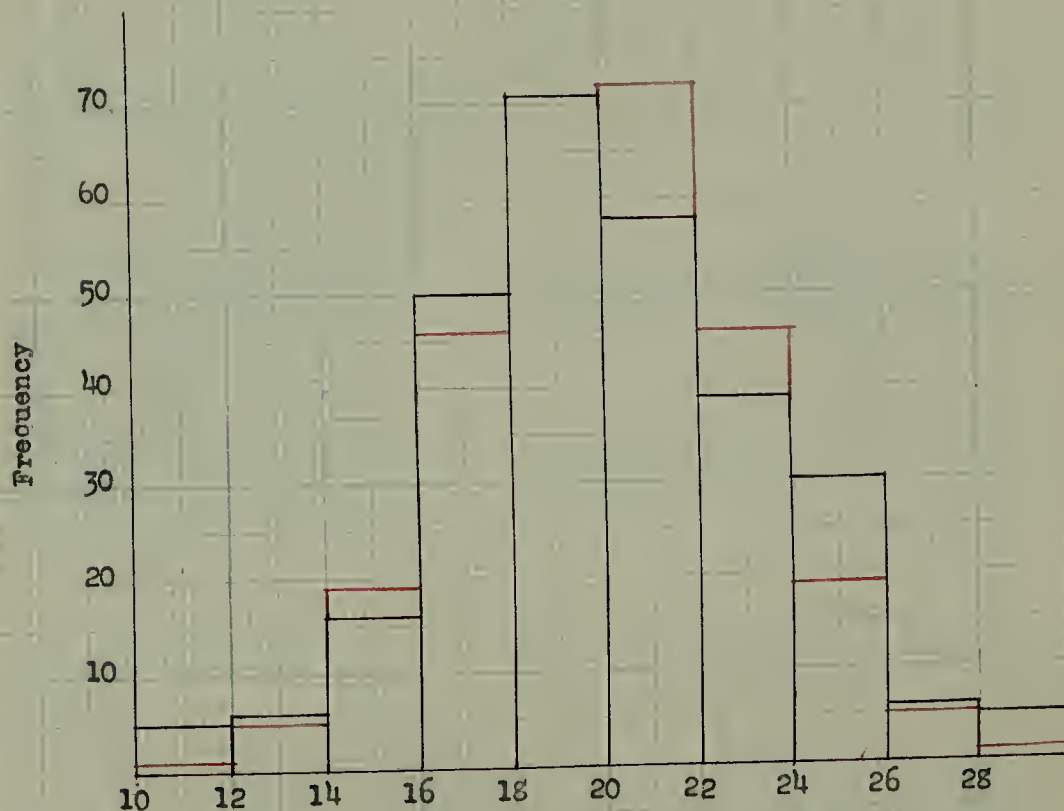
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 29

Histogram showing the actual and theoretical frequency distributions of the Memory Reading section individual scores (Massachusetts State College Scholastic Aptitude Test)

(285 cases)



Memory Reading section individual scores (Massachusetts State College Scholastic Aptitude Test)

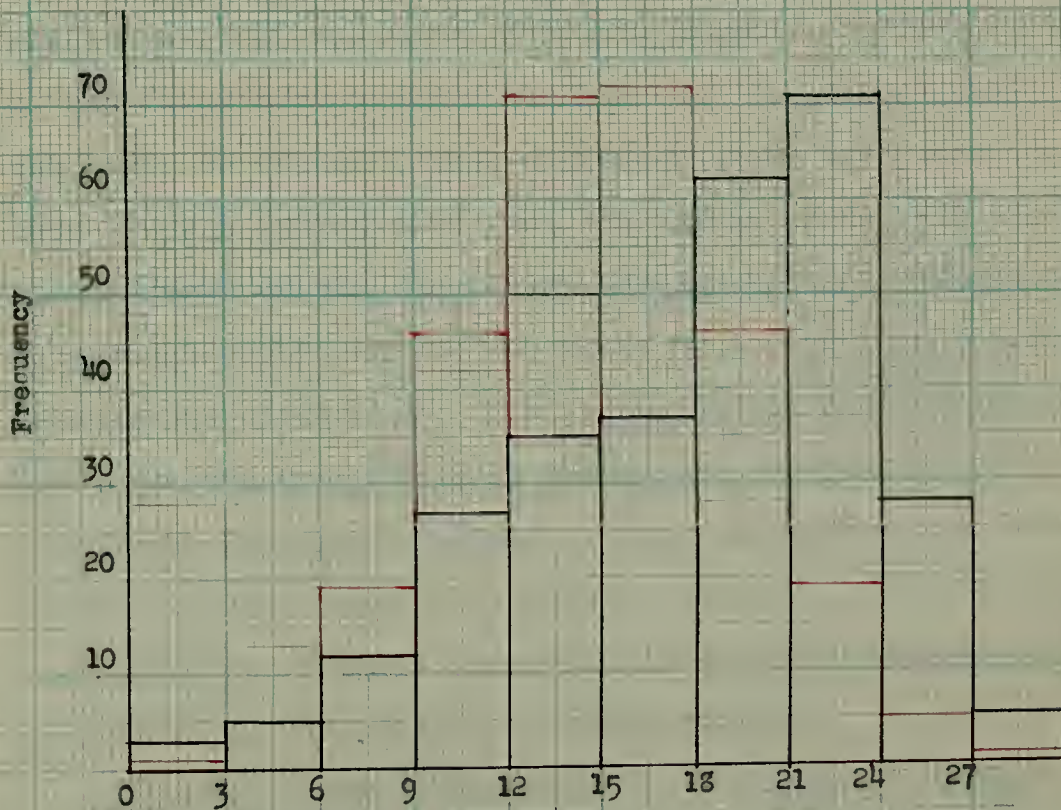
Represents actual frequency distribution

Represents theoretical distribution

FIGURE 30

Histogram showing the actual and theoretical frequency distributions of the Artificial Language section individual scores (Massachusetts State College Scholastic Aptitude Test)

(285 cases)



Artificial Language section individual scores
(Massachusetts State College Scholastic Aptitude Test)

Represents actual frequency distribution

Represents theoretical distribution

to a marked degree that the section is too easy. The measures rise gradually through the eighth column, and then drop rapidly in the two remaining columns. A substantial increase in the number of presented items is quite necessary in order to reform the skewed shape of the curve.

A general consideration of the test gross scores indicates that for the purpose of mathematical comparison, several changes should be made. First, no single test is of sufficient difficulty. If the curve shown in Figure 1 be representative of all first term average mark curves, it is recommended that the tests be of sufficient difficulty that the gross scores more closely approximate, not the normal curve, but the right-skewed curve of the first term average mark distribution. The adjustment should affect the tests in the following order of increase in difficulty: the Psychological Examination, the Massachusetts State College Scholastic Aptitude Test, and the Army Group Examination Alpha. While the extent of deviation can thus be evaluated theoretically, and the tests revised accordingly, the crucial determinants are best secured only by persistent and intelligent experimentation.

Summary and interpretation. A qualitative account of the frequency distributions of instructors' marks for the first and second terms shows (1) that the first term average marks tend to approximate the normal-frequency curve, (2) that the second term average marks deviate somewhat from the theoret-

ical curve of error, and (3) that with few exceptions the distributions of the term marks are "irregular" or abnormally skewed. A similar inspection of the composite, gross and individual score frequency distributions reveals that here, too, are discrepancies. In the latter instance, the matter is easily remedied -the tests are wholly amenable to change, to revision. To attempt a change in regard to the marks of the college instructor -is another matter.

The thesis and raison d'etre of the preliminary analysis is this. In order to predict with any degree of accuracy the future achievement of a group, it is absolutely essential to know the general nature of the criterion of success. In this case, the criterion of success is "teachers' marks". Were these marks in harmony with the normal-frequency distribution, test construction could so be directed as to secure likewise normal-frequency distributions. This accomplishment would establish, at least theoretically, the possibility of a highly accurate prognostic indicator. The results of this study indicate that the frequency distributions of instructors' marks for freshman in the first two terms of the 1931-1932 academic year deviate markedly from the Gaussian curve. Hence, the results of the preliminary investigation suggest questions which cannot be side stepped: (1) Can this obstacle be overcome and, (2) if so, how?

The writer believes that this contingency can be met in

one of two ways. The first is the ideal: that the freshman instructors be required to observe the law of normal distribution when assigning marks to classes of more than 100 students. The second is more pragmatic, since it is more probable. A study¹ can be made of the actual frequency distributions in each course for a period of (say) five years. From the obtained results, an approximation of the mark frequency-distributions can be attempted for the test scores -thus increasing the possibility of an accurate indicator of future achievement. Unfortunately, however, there are several millstones attached to this procedure. The detailed process of securing the warped curves of the college instructors' marks is one of extreme tedium. Then, too, there is the uncomfortable fact that, when once secured, the frequency distributions are not wholly stable -since college instructors are subject to geographical change. Nevertheless, although the first method is the ideal, and in spite of the seemingly insurmountable "blocks" in the pathway of the second, the results here presented indicate definitely that it is with the second possibility which the test investigators now have to deal.

1. Under the direction of Dr. H. N. Glick, such a study is now in progress. It is the hope of this writer that the publication of the results of that study -in comparison with the results of similar studies in other American colleges -will at least partially remedy the situation.

CHAPTER IV

. SECONDARY ANALYSIS: FIRST TERM

Methodology. The technic employed to determine the relative validity of the Psychological Examination, the Army Group Examination Alpha, and the Massachusetts State College Scholastic Aptitude Test as prognostic indicators of the success achieved scholastically by the members of the Class of 1935 during the first term of the 1931-1932 academic year is the product-moment method of correlation advocated by the English statistician, Karl Pearson.¹ Using the expression in its general sense, the coefficient of correlation may be defined as "a numerical index of the relationship between two sets of paired facts."² More definitely, the coefficient of correlation between two variables may be regarded as a measure of the degree to which a change in one tends to be accompanied by a change in the other. It ranges in value from +1, a perfect positive correlation; through zero, no correlation at all; to -1, a perfect negative correlation. In regard to the interpretation of the obtained fraction, certain definite regulations have been presented. F. H. Harper³ quotes King as suggesting these rules:

1. For an illustration of the product-movement method of correlation, see Figure

2. Monroe, W. S., op. cit., p. 342.

3. Harper, F.H., Elements of Practical Statistics, The Macmillan Co., 1930, pp. 197-8.

- "1. 'If r is less than the probable error there is no evidence whatever of correlation.'
2. 'If r is more than six times the size of the probable error the existence of correlation is a practical certainty.'
3. When the probable error is relatively small, 'if r is less than .30 the correlation cannot be considered at all marked.'
4. If the probable error is relatively small, a coefficient 'above .50 indicates decided correlation.'"

Professor H. D. Rugg¹ insists that for educational purposes a correlation less than .15 or .20 is "negligible"; one ranging from .15 or .20 to .35 or .40 is "present but low"; from .35 or .40 to .50 or .60, it is "markedly present" or "marked"; and from .60 or .70 up, the correlation is "high".

Another standard for interpretation is that set forth by Professor Trow².

"	.80 to .95	very high
	.60 to .80	high
	.40 to .60	substantial
	.20 to .40	low
	.05 to .20	very low"

Obviously, here is an instance where experts disagree. Yet, for this study, some kind of arbitrary standard must obtain. The decision of the writer is that general classifications are unsatisfactory unless tempered by common sense, in

1. Rugg, H.D., Statistical Methods Applied to Education, Houghton Mifflin Co., 1917, pp. 256-7.

2. Trow, W.C., Educational Psychology, Houghton Mifflin Co., 1931, pp. 170-1.

this case, by reference to the correlations secured in representative studies similar to the one at hand. For example, a correlation may be "low" with respect to perfection ($r = +1$), and yet be "high" when compared with the existing correlations between the same two attributes. Hence, for this investigation, a dual standard is established: (1) the percent of forecasting efficiency (100% forecasting efficiency = 1.00 coefficient of correlation) as determined by Kelley's formula for the coefficient of alienation, $K = (1 - r^2)^{\frac{1}{2}}$, and (2) the figure most commonly secured when similar data are correlated. The first is the standard, or goal. The second is the normal, or the level of success thus far achieved.

The alienation coefficient, K , reveals lack of relationship and consequently the inaccuracy of any prognosis based on the relationship. Therefore, the complimentary value of K , expressed in percent, gives the prognostic efficiency of the correlation coefficient.¹ As regards the second standard, the normal, the results of the survey presented in Chapter I indicate that any index exceeding .40 is above average for the prediction of success in the first year of college work.

The second test for a satisfactory coefficient is its reliability. The greater the number of cases, the more representative and reliable is the correlation coefficient. The

1. Hull, C.L., "The Correlation Coefficient and Its Prognostic Significance", Journal of Educational Research, (15), May, 1927, pp. 327-338.

numerical index expressing the extent of dependability is called the probable error and is denoted the P.E._r. Speaking in general of the ratio which must exist between the coefficient and its probable error before the existence of a relationship can be asserted, Monroe¹ states that the experts differ: "By at least one writer the ratio is placed at 6. Another places it as low as 2 or 3. A conservative rule is that the coefficient must be four times its probable error before the existence of a relationship can be assumed." However, in order to meet the approval of all critics, it seems more advisable to project the results against the most exacting standards. Thus, in the present study, the criterion of reliability -the degree to which a test is consistent in measuring what it purports to measure -is that the coefficient of correlation must be at least 6 times as great as the probable error.

Composite scores and average marks. The coefficient of correlation obtained by the Pearsonian product-moment method in the study of the composite scores and average marks for the first term is .53; the probable error $\pm .03$. This means that the chances are even that the true correlation lies between .50 and .56.² There is no question as to the existence of a positive relationship between the two kinds of responses here measured -r being 13 points higher than the "average" index of

1. Monroe, W.S., op. cit., p. 346-7.

2. Monroe, W.S., op. cit., p. 346.

these responses, and over 6 times as large as the P.E. The chances are 21 to 1 that if these tests were given again next year, other factors being equal, the correlation between the composite scores and average marks would be at least 4 points higher than the average of the correlation coefficients heretofore obtained in similar studies. This assured margin is small, yet it is indicative of substantial progress.

On the other hand, when the index is interpreted in terms of exact forecasting efficiency, the result is less satisfactory. The complimentary value of the alienation coefficient is little better than 15, indicating that future academic success can be predicted with only 15% accuracy. As low as this expression is, however, it represents an advance of more than 85% in prognostic efficiency over the "average" indices.

Gross scores and average marks. The correlation between the gross scores of the Psychological Examination and the first term average marks is $.347 \pm .0347$. Since the index is 10 times as great as the P.E., it can be said that the coefficient is reliable. That is to say, the chances are 21 to 1 that, other factors being equal, a correlation between the scores on this test and the average marks for the first term will be at least .243. While this figure is but 5.6 points lower than the normal, it is practically useless for predictive purposes being less than 7% efficient.

The index secured between the average marks and the Army

Group Examination Alpha gross scores is. $387 \pm .034$. Here, again, the coefficient is over 6 times the probable error - suggesting the existence of a definite relationship. When compared with the goal, 100% prognostic accuracy, however, it is not quite 8% efficient. On the other hand, it is less than 2 points below the normal in validity.

The gross scores of the Massachusetts State College Scholastic Aptitude Test, the measures presumably based on ability to learn, when compared with the average marks give a correlation coefficient of $.488 \pm .035$. In this instance, the index is well over 6 times the probable error. According to Monroe¹, the chances are even that the true correlation lies between .453 and .523, and 21 to 1 that it falls between .37 and .59. The results also show that the test is not quite 13% accurate in forecasting efficiency. Yet, with respect to the normal, it indicates a significant advance. The same advance holds over and above the other two tests used in this study. The "learning" test is more valid than the Army Group Examination Alpha by 10 points, and more valid than the Psychological Examination by better than 14 points. Hence, the results of the first comparison between the test based primarily on ability to learn and the tests essentially measuring past experience seem to substantiate the hypotheses provisionally set forth in an ear-

1. Monroe, W.S., op. cit., p. 346.

lier section¹.

Gross scores and term marks. Table 5 shows the correlation coefficients secured between the gross scores and the first term marks. The obtained index between Army Group Examination Alpha gross scores and the Orientation first term marks is $.30 \pm .039$ ($N = 282$). There is a small, positive relationship between the test and the term marks. As a prognostic test, however, it is impractical -being only 5% efficient.

The Psychological Examination gross scores and the Orientation term marks ($N = 282$) produce a coefficient which is 4 points higher and more reliable than the Army Group Examination Alpha "index" $r = .34 \pm .035$. Although somewhat closer to the normal, the index indicates only a slight advance in forecasting efficiency, from 5% to c. 6%.

A marked increase in reliability and predictive value is apparent from an examination of the correlation coefficient secured by comparing the 282 gross scores of the Massachusetts State College Scholastic Aptitude Test and the Orientation term marks: $r = .588 \pm .026$. This means that the chances are 1 to 1 that the true correlation is at least .56, at best .61 - showing an increase of no less than 16 points over the normal. Moreover, when used to predict Orientation term marks, this test is almost 20% accurate.

1. Cf. This work, Chapter I, p.

TABLE 5

Correlations among gross scores and first term marks.

Term Marks	Army Alpha Gross Scores		Psychological Gross Scores		M. S. C. Gross Scores		Number of Cases
	Coef. of Cor.	P.E.	Coef. of Cor.	P.E.	Coef. of Cor.	P.E.	
Orientation	.30	±.036	.34	±.035	.588	±.026	282
Chemistry	.167	±.039	.21	±.038	.197	±.038	283
Mathematics	.24	±.04	.30	±.04	.35	±.037	259
English	.21	±.038	.17	±.039	.217	±.038	285
German	.28	±.057	.21	±.058	.247	±.057	124
French	.27	±.063	.24	±.064	.61	±.043	100
Average Index	.24		.25		.37		

An examination of the indices existing among the gross scores of the three mental tests and the Chemistry term marks reveals that none is significant. In spite of the fact that 283 cases are used, in no case is the coefficient sufficiently in excess of the probable error to warrant the assumption of a relationship.

The index existing between the Army Group Examination Alpha gross scores and the Mathematics term marks is $.24 \pm .038$ ($N = 259$). This figure is little better than one half the normal and is useless for prediction.

An increase of 6 points is noted in the coefficient of correlation between Mathematics term marks and the gross scores of the Psychological Examination. While a positive relationship may be assumed, the index, $.30 \pm .035$, is only three fourths as large as the normal, and is but 5% accurate as regards forecasting efficiency.

The coefficient, $.35 \pm .032$, obtained by comparing the Massachusetts State College Scholastic Aptitude Test gross scores with the Mathematics first term marks represents a gain of 11 points over the Army Group Examination Alpha index and a gain of 5 points over the second "recall" test. Although this figure is 5 units below the normal, it is 11 times greater than its probable error -indicating a rather definite relationship. With respect to forecasting accuracy, however, it is hardly better than the other two tests -being less than 7% ef-

ficient.

The first two correlation indices existing among the mental test gross scores and the English first term marks are probably insignificant. Neither figure is 6 times as great as its respective probable error -and if one were, as is the third case, the most lenient of statisticians would claim but a "low" relationship. The predictive value of all three tests for English marks is of little use.

Practically the same adverse results hold in the case of the German term marks. The coefficients among the Army Group Examination Alpha, the Psychological Examination, the Massachusetts State College Scholastic Aptitude Test gross scores, and German term marks are respectively: $.26 \pm .048$, $.21 \pm .055$, and $.247 \pm .054$. These results are based on 124 subjects, which probably accounts for the relatively high probable errors. The mental test gross scores are hardly more efficient prognostic indicators of success in German than they were in English. In each case the results are far below the normal.

The correlations among the tests gross scores and the French term marks represent a slight increase over the indices secured among the other modern languages and the test gross scores. The index between the Army Group Examination Alpha gross scores and the term marks in French is $.27 \pm .06$, that between the Psychological Examination and the French term marks is $.24 \pm .051$ and the coefficient secured by comparing

the Massachusetts State College Scholastic Aptitude Test and the French term marks is $.61 \pm .049$. It is possible that the number of cases ($N = 100$) renders the first two indices invalid. In the third case, however, it is quite safe to assume a definite positive relationship between the two attributes measured. The coefficient, $.61$, is more than 12 times the probable error, and in forecasting accuracy the test is slightly better than 20% efficient.

The correlations among the mental test gross scores and the separate term marks indicate that the Massachusetts State College Scholastic Aptitude Test is superior to the Army Group Examination Alpha and the Psychological Examination as a prognostic indicator of school success. The average indices of the Army Group Examination Alpha and the Psychological Examination are $.24$ and $.25$ respectively. Each is much less than the normal. For the Massachusetts State College Scholastic Aptitude Test gross scores and the term marks, the average coefficient is $.37$ -but 3 points below normal, and at least 15 points higher than either of the other two coefficients.

Individual scores and average marks. The study of the relationships existing among the mental test individual scores and the average marks is limited to the Psychological Examination and the Massachusetts State College Scholastic Aptitude Test. There are two principal reasons for this restriction: (1) the time required to carry three tests throughout the work

was prohibitive, and (2) the status of the Psychological Examination¹ was such as to occasion its selection in preference to the Army Group Examination Alpha.

Table 6 presents the various coefficients obtained by correlating the individual scores and the average marks for the first term. The index, $.327 \pm .036$, represents the relationship existing between the average marks for the first term and the Completion section individual scores of the Psychological Examination. The probable error is less than one-sixth of the coefficient, showing that the relationship is significant and dependable. The compliment of the coefficient of alienation expressed in percent is less than .06, thus denoting impractical efficiency. In regard to the average, the section is below normal by almost 8 points.

The coefficient of correlation existing between the Artificial Language section individual scores of the Psychological Examination and the first term average marks is $.31 \pm .036$. This figure is 9 points below the normal, is about 5% efficient prognostically, and is of little if any practical use for prediction. Though small, the coefficient is reliable, since it is more than 6 times as large as the probable error.

1. Previous studies seemed to indicate that the Psychological Examination was the more valid test for predicting success in the first year at this college (Cf. Glick, H.W., "The 1929 Psychological Examination, " Educational Record, April, 1930 pp. 120-1).

TABLE 6

Correlations among individual scores and first term average marks.

Psychological Examination
Individual Scores

	Completion		Artificial		Arithmetic		Analogies		Opposites	
	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.
Average Marks	.327	$\pm .036$.31	$\pm .036$.45	$\pm .03$.259	$\pm .037$.14	$\pm .039$

(N = 284)

Massachusetts State College Scholastic Aptitude Test
Individual Scores.

	Botany		Chemistry		Artificial Language		Mathematics		Power Reading		Memory Reading	
	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.	Coef. of	P.E. Cor.
Average Marks	.267	$\pm .037$.339	$\pm .035$.315	$\pm .036$.257	$\pm .037$.19	$\pm .038$.307	$\pm .036$

(N = 289)

The index obtained by correlating the average marks for the first term with the Arithmetic Section individual scores of the Psychological Examination is $.45 \pm .03$. That the coefficient is reliable is indicated by the low probable error which is about one-fifteenth of the index. Generally speaking, the index $.45 \pm .03$ is unusual for a single part of a test. Although the writer is acquainted with no exhaustive account of such studies, he believes from experience that a correlation above $.40$ between the criterion and a special section of a test is unusually high. In fact, not only is this figure high for a single section, but it is higher than that obtained by the gross scores of the whole test. Thus it happens that the Mathematics section of the Psychological Examination is a more valid measure of academic success for the first term of the 1931-1932 school year than the gross test. It may be¹ that more factors are encompassed by the test as a whole. Yet, for the purpose of forecasting scholastic achievement, the technologist assumes the right to exclude invalid parts of any test regardless of the theoretical implications.

The index, $.259 \pm .037$, shows the extent to which a change in the Analogies section individual scores of the Psychological Examination tends to be accompanied by a change in the first term average marks. The mathematical expression of this tendency is almost 15 points below the normal, and although it is

1. Cf. This work, Chapter V.

reliable (\bar{r} is more than 6 times the P.E.), it is less than 4% efficient with respect to predictive accuracy.

The coefficient of correlation existing between the first term average marks and the Opposites section individual scores is insignificant ($\bar{r} = .14 \pm .039$); the index being less than five times the probable error. However, since Monroe¹ states that a conservative rule is to accept a coefficient which is four times the probable error, it may be well to examine the coefficient more closely. In the first place, it is roughly but one-third as great as the normal. With respect to the standard, it is quite far removed. Lastly, the prognostic efficiency of this figure is less than 1% better than a guess. Hence, from all points of view, the test under consideration is of little import as regards prediction.

In measuring a complex variable by means of somewhat different scales, it is assumed that the greater the number of factors measured the more truly is that function represent. Consequently, when a fairly accurate measure of a complex function is obtained by means of separate measurements, it can usually be anticipated that each measurement will correlate less with the complex variable -here, the criterion -than will a composite of these measurements. Hence, the correlation coefficients among the individual scores and the average marks are necessarily lower than those of the gross scores and the aver-

1. Monroe, W.S., op. cit., p.

age marks -unless a particular section (as in the case of the Mathematics section of the Psychological Examination) is more valid than the composite.

In Table 6 are also found the coefficients (and the respective probable errors) existing among the individual scores of the Massachusetts State College Scholastic Aptitude Test and the criterion, average marks. The first index, $.267 \pm .037$, represents the relationship existing between the average marks and the individual scores of the Botany section. The probable error is less than one-sixth the index, showing that the index is dependable. Expressed in per cent, the compliment of the coefficient of alienation is less than .05, denoting impractical, prognostic efficiency. It is also a fact that the coefficient is below the normal by almost 14 points.

The coefficient of correlation between the Chemistry individual scores and the first term average marks is $.339 \pm .035$. This figure is practically 6 points below the normal, is approximately 6% accurate, and consequently is practically useless for efficient prediction. The relationship here in existence is reliable, since the index is more than 6 times the probable error.

The correlation coefficient obtained by comparing the average marks with the Artificial Language section is $.315 \pm .036$. That the index is reliable is indicated by the probable error which is but one-ninth the coefficient. This correlation is

8.5 points below the normal, and is only 5% efficient -indicating impractical accuracy as regards prediction of academic success.

The figure, $.257 \pm .037$, shows the relationship between the two sets of paired facts-the Mathematics section individual scores and the average marks for the first term. The coefficient, .257 is well over 6 times the probable error. Hence, the relationship is reliable and probably positive. Yet, it is almost 15 points below the normal, and for prediction it is about 4% efficient.

The coefficient, $.307 \pm .036$, measures the degree to which a change in the average marks tends to be accompanied by a change in the individual scores of the Memory Reading section. The numerical index of this tendency is less than 10 points below the "average", and though it is reliable (r is over 6 times as great as the probable error), it is only 5% efficient prognostically -i.e., the section is useless for practical forecasting.

The individual scores of the last section, Power Reading, correlate $.19 \pm .038$ with the average marks. This means that the relationship is of questionable dependability. Even Professor Monroe would hesitate to assume a relationship here. The coefficient is 21 points below the "average", indicating that the section is but 2% accurate, and useless for prediction.

Individual scores and term marks: The Psychological Exam-

ination. In comparing the individual scores of the mental test sections with the term marks, no normal, or "average", of what has thus far been accomplished is available. However, Professor Trow¹ suggests that for this purpose a correlation from .05 to .20 is "very low"; one ranging from .20 to .40 indicates that a relationship is "present but low"; and for a coefficient above .40 that is a "substantial" relationship, etc. Table 7 shows the coefficients obtained by comparing the Individual scores of the Psychological Examination with the first term marks. The index (and probable error) existing between the Completion section individual scores and the Orientation marks is $.147 \pm .039$. This means that it is unsafe to assume a relationship between the variables measured. The index is very "low", and is not dependable. With respect to predicting Orientation term marks, the Completion section is less than 1% efficient, i.e., practically useless.

The index, $.21 \pm .038$, represents the relationship existing between the Chemistry term marks and the individual scores of the Completion section. The compliment of the coefficient of alienation expressed in per cent is about .02 -denoting impractical efficiency as regards the prediction of Chemistry marks. The coefficient is not 6 times the probable, but it is more that 5 times as great. If a relationship be assumed, it is "very low".

1. Trow, W.C., op. cit., pp. 171-172.

TABLE 7

Correlations among individual scores and first term marks

Psychological Examination

Term Marks	Completion	Artificial Language		Arithmetic		Analogies		Opposites		Number of Cases
	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.			
Orientation	.147 \pm .039	.225 \pm .037	.17 \pm .039	.24 \pm .037	.188 \pm .038				282	
Chemistry	.21 \pm .038	.11 \pm .04	.307 \pm .036	.097 \pm .04	.127 \pm .039				283	
Mathematics	.07 \pm .04	.19 \pm .04	.297 \pm .039	.165 \pm .04	.07 \pm .04				259	
English	.12 \pm .04	.35 \pm .035	.077 \pm .04	.215 \pm .038	.145 \pm .039				285	
German	.136 \pm .059	.14 \pm .059							124	
French		.28 \pm .06							100	

The coefficient of correlation existing between the Mathematics term marks and the individual scores of the Completion section is $.07 \pm .04$. The test is less than one-half per cent accurate as regards prognosis, and as indicated by the probable error it is not reliable. The Completion section of the Psychological Examination is not recommended to predict Mathematics term marks.

The index found by correlating the Completion test individual scores with the English term marks is $.12 \pm .04$. That the index is not reliable is signified by the probable error which is but one-third of the coefficient. Again, since the per cent of forecasting efficiency is less than 1, the test is hardly accurate as a prognostic indicator. If a relationship be assumed, it is very low.

The index, $.225 \pm .037$, shows that extent to which a change in the Artificial Language section individual scores tends to be accompanied by a change in the Orientation first term marks. The mathematical expression of this tendency is more than 6 times the probable error, and is therefore reliable. Professor Trow would here assume a relationship which he would designate "present but low". However, the prognostic efficiency of the Artificial Language test for predicting success in Orientation is slightly better than 2%, signifying non-practicality.

The second coefficient in the Artificial Language column

is $.11 \pm .04$. This index expresses the relationship existing between the Chemistry term marks and the Artificial Language section individual scores of the Psychological Examination. The coefficient is less than 3 times the probable error signifying that the index is not dependable. It is unsafe to assume a positive relationship between the two attributes measured.

The index existing between the Artificial Language test individual scores and the Mathematics term marks is $.19 \pm .04$. Should the index be taken as a reliable one, the relationship existing between the mental test section and the Mathematics course is very low. In regard to prognosis, the section is practically useless -since it is slightly less than 2% accurate.

The coefficient, $.35 \pm .035$, represents the relationship existing between the individual scores of the Artificial Language section and the English first term marks. That the index is reliable is apparent from the size of the probable error which is one-tenth of the index. This fact signifies that the coefficient is dependable. Although this figure, $.35 \pm .035$, is the most significant and valid index recorded in Table 7, it is none the less impractical for prediction -the per cent of forecasting being less than 7%.

The next column contains the coefficients obtained by comparing the Arithmetic section with the various first term

marks. The first index, $.17 \pm .039$, is that secured by correlating the individual scores of the Arithmetic section with the first term Orientation marks. If a relationship be assumed to exist between these two sets of paired facts, it is "very low" -- according to Professor Trow. Whether or not this assumption be valid is of little import, however, since the compliment of the coefficient of alienation expressed in per cent is less than .02. The Arithmetic section is not recommended as a prognostic test for Orientation term marks.

The coefficient of correlation existing between the first term Chemistry marks and the Arithmetic section individual scores is $.307 \pm .036$. The probable error indicates that the coefficient is dependable. In regard to the validity of the Arithmetic section, Professor Trow would assume a relationship, and designate it "present but low". Taken alone, the Arithmetic section is not a valid instrument for predicting Chemistry first term marks -- as indicated by the per cent of forecasting efficiency, .05.

The index found by correlating the Mathematics first term marks with the Arithmetic section individual scores is $.297 \pm .039$. The probable error is less than one-sixth the index, denoting reliability. While Professor Trow would assume a relationship ("present but low"), the per cent of forecasting efficiency is but .05 -- indicating that the Mathematics section is practically useless for predicting Mathematics first term marks.

The last figure, $.077 \pm .04$, in the Arithmetic column represents the relationship existing between the Arithmetic section individual scores and the English first term marks. This index is not dependable since it is less than 2 times as great as the probable error. Nor is it safe to assume that a positive relationship exists between the two variables here measured. The test is useless for the prediction of English first term marks.

The coefficient obtained by comparing the Analogies section individual scores with the Orientation first term marks is $.24 \pm .037$. That the index is dependable is signified by the probable error which is less than one-sixth the coefficient. If a relationship be assumed to exist between the attributes measured, it is "present but low". The Analogies section is not an accurate indicator of success in Orientation, since it is but 3% efficient.

The only negative coefficient of correlation found in the complete study is that denoting the relationship between the Analogies section of the Psychological Examination and the Chemistry first term marks, $r = -.097 \pm .04$. However, since this index is but little more than twice the probable error, no inverse relationship can be assumed. On the other hand, were such a relationship assumed, it would necessarily be interpreted as "very low" and of impractical significance for prognostic purposes.

The next obtained coefficient, $.165 \pm .04$, is that existing between the first term Mathematics marks and the individual scores of the Analogies section. Since the coefficient is only slightly over 4 times as great as the probable error, it is hardly safe to assume a positive relationship which, if assumed, would be "very low". As regards the prediction of Mathematics first term marks, the Analogies section is less than 2% efficient.

The index, $.215 \pm .038$, shows the extent to which a change in the individual scores of the Analogies section tends to be accompanied by a change in the first term English marks. The index is about 6 times the probable error -indicating dependability. It is safe to assume a relationship between the variables here measured. It is "present but low". The forecasting accuracy of the section less than 2% -denoting impractical efficiency.

In the column labelled Opposites are found the coefficients (and the probable errors) existing between the Opposites section individual scores and various term marks. The first index, $.188 \pm .038$, is that secured by comparing the individual scores of the Opposites section with the first term Orientation marks. This index is slightly better than 4 times the probable error. Since Professor Monroe would assume a relationship in this instance, it is well to examine the index more closely. If a relationship does exist between the attributes

measured, it is indeed "very low." Moreover, since the test is less than 2% accurate -it is practically useless for the purpose of forecasting first term Orientation marks.

The next coefficient, $.127 \pm .039$, indicates the extent to which a change in the individual scores of the Opposites section tends to be accompanied by a change in the first term Chemistry marks. That the section is unreliable is shown by the probable error which is over one-third the index. The test is less than 1% efficient in predicting Chemistry first term marks -and is therefore impractical.

The last two indices in the Opposites column, $.07 \pm .04$ and $.145 \pm .039$, indicate the relationship existing between the Analogies section individual scores and the first term (1) Mathematics and (2) English marks respectively. Neither section is dependable nor valid. In each case the index is less than 4 times the probable error, and the per cent of forecasting efficiency is less than 1.

It is to be regretted that a complete list of the German and French individual score-term mark indices is not presented.¹ However, of the three coefficients available, the first, $.136 \pm .059$, indicating the extent of relationship existing between the individual scores of the Completion section and the

1. Want of time was the determining factor. A choice was afforded in regard to (1) the complete table and (2) a further analysis of the data here treated (Cf. Chapter VI). The writer chose the latter.

German first term marks, denotes that the test is not reliable, and that it is impractical for the purpose of forecasting first term German marks.

The next two coefficients are obtained by comparing the Artificial Language section individual scores with the German first term marks ($r = .14 \pm .059$) and the first term French marks ($r = .28 \pm .06$). The first index denotes that the Artificial Language section is unreliable and practically useless for predicting German first term marks. The second likewise indicates that the section is unreliable. Although Professor Monroe would here assume a relationship which Professor Trow would stamp "present but low". The Artificial Language section is not efficient for predicting success in French -the forecasting accuracy being less than 5%.

Individual scores and term marks: The Massachusetts State College Scholastic Aptitude Test. The first index presented is $.31 \pm .036$. This figure expresses the degree to which a change in the individual scores of the Botany section tends to be accompanied by a change in the Orientation first term marks. That the test is reliable is signified by the probable error which is about one-ninth the index. The relationship existing here is "present but low". However, since the Botany section is scarcely more than 5% accurate -its prognostic efficiency is hardly practical.

The index, $.11 \pm .04$, represents the relationship existing

between the Chemistry first term marks and the individual scores of the Botany section. In the first place, the probable error denotes that the coefficient is not reliable -since it is less than 4 times the probable error. Next, if assumed, the relationship existing between the attributes measured is indeed very low. Lastly, the efficiency of the Botany section in regard to predicting Chemistry first term marks is about one-half of one per cent efficiency, designating non-practicality.

The coefficient found by correlating the first term Mathematics marks with the individual scores of the Botany section is $.22 \pm .04$. That the index is probably reliable is evidenced from the relative size of the probable error which is almost one-sixth the coefficient. The assumed relationship is "present but low". In regard to predictive accuracy, the Botany section is slightly more than 2% efficient in forecasting Mathematics first term marks.

The index obtained by comparing the Botany section individual scores with the first term English marks is $.217 \pm .038$. The index is less than 6 but more than 5 times the probable error. Professor Monroe would here assume a relationship between the variables measured. Professor Trow would label that relationship "present but low". Be this as it may, the Botany section is of impractical accuracy for predicting success in first-term English, since it is but 2% efficient.

The next column, entitled Chemistry, contains the correlation coefficients indicating the degree of relationship existing between the Chemistry section individual scores and the various first term marks. The first, .21 -.038, denotes the extent to which a change in the individual scores of the Chemistry section tends to be accompanied by a change in the Orientation term marks. This index is less than 6 but more than 5 times the probable error. Professor Monroe would consider it safe to assume a relationship between the measured attributes. In the light of the arbitrary standard voiced by Professor Trow, the relationship is necessarily classified as "present but low". With respect to predictive accuracy the Chemistry section is about 2% efficient in forecasting Orientation first term marks.

The next coefficient is one of the few "interesting" figures contained in Table 8. In the first place, in the 26 indices presented, it is the lowest. Then, again, when the facts -that this section (1) derived the highest criterion (average mark) coefficient, (2) purports to be a fair measure of ability "to do" chemistry, and (3) was prepared by one of the chemistry instructors -are recalled, a "distressing discrepancy" is at once apparent. The writer hold no brief for a commendation or a condemnation of either the test section or the chemistry term marks. It is wholly probable, however, that the test items do not encompass the attributes necessary

TABLE 8

Correlations among individual scores and first term marks
Massachusetts State College Scholastic Aptitude Test
Individual Scores

Term Marks	Botany		Chemistry		Artificial Language		Mathematics		Memory Reading		Power Reading		Number of Cases
	Coef. of Cor.	P.E. Cor.	Coef. of Cor.	P.E. Cor.	Coef. of Cor.	P.E. Cor.	Coef. of Cor.	P.E. Cor.	Coef. of Cor.	P.E. Cor.	Coef. of Cor.	P.E. Cor.	
Orientation	.31	±.036	.21	±.038	.268	±.037	.195	±.038	.327	±.036	.29	±.036	282
Chemistry	.11	±.04	.09	±.04	.29	±.036	.24	±.038	.16	±.039	.31	±.036	283
Mathematics	.22	±.04	.275	±.04	.129	±.04	.305	±.039	.168	±.04	.15	±.04	2259
English	.217	±.038	.28	±.036	.359	±.039	.106	±.04	.29	±.036	.31	±.036	285
German					.37	±.05							124
French					.409	±.057							100

for successful work in the chemical laboratory -and the writer sets forth this reason as a possible solution. Then, too, the index, $.09 \pm .04$, is unreliable, since it is hardly better than 2 times the probable error. In regard to predicting Chemistry first term marks, the Chemistry section is worthless.

The coefficient obtained by comparing the first term Mathematics marks with the individual scores of the Chemistry section is $.275 \pm .04$. According to Hooper, correlation is practically certain, because the index is over 6 times the probable error. Moreover, the relationship here assumed is present though low. With respect to predicting Mathematics first term marks, the Chemistry section is hardly practical being less than 5% efficient in prognostic accuracy.

The index secured by comparing the individual scores of the Chemistry section with the English first term marks is $.28 \pm .036$. It is greater than 6 times the probable error; hence, a correlation is practically certain. This relationship between the measured variables is present but low. The prognostic accuracy is less than 5%, indicating that the Chemistry section is impractical for predicting first term English marks.

The six coefficients presented in the Artificial Language column represent the various relationships existing between the first term marks and the individual scores of the Artificial Language section. The first, $.268 \pm .037$, represents the

relationship existing between the first term Orientation marks and the individual scores of the Artificial Language section. That the index is dependable is designated by the relative size of the probable error which is less than one-sixth the coefficient. Here, again, the relationship is present but low. In regard to the question of forecasting efficiency, the Artificial Language section is not practical for predicting Orientation first term marks; being less than 5% effective.

The index existing between the individual scores of the Artificial Language section and the Chemistry first term marks is $.29 \pm .036$. Since the index is well over 6 times the probable error, it is reliable, and a relationship may be assumed. Professor Trow would designate the existing relationship "present but low." Moreover, the per cent of forecasting efficiency, .05, signifies that the section is practically useless for predicting first term Chemistry marks.

The third coefficient in this column is that showing the degree to which a change in the Artificial Language individual scores tends to be accompanied by a change in the first term Mathematics marks ($r = .129 \pm .04$). Although Professor Monroe would assume a relationship between the measured attributes, the maxim of Professor Trow suggests that it is "very low". With respect to the forecasting efficiency of the section in predicting Mathematics first term marks, it is less than 1% effective.

The degree of relationship existing between the English first term marks and the Artificial Language section individual scores is denoted by the coefficient of correlation, $.358 \pm .035$. The relative size of the probable error (one-tenth the index) signifies practical certainty for the correlation. The interpretation of this relationship is that it is "present but low" -approaching the border of substantiality. The section is almost useless for predicting success in first term English.

The correlation obtained between the German first term marks and the individual scores of the Artificial Language section is $.37 \pm .05$. This figure is quite significant. It is over 7 times as great as the probable error, insuring a relationship. While the section is hardly efficient for predicting success in German (being about 7% effective), the situation signifies a marked advance over and above the Artificial Language section of the Psychological Examination (Cf. Table 7)

In a similar fashion, and to a more noticeable extent, the index, $.409 \pm .057$, reveals the superiority of the Artificial Language section of the Massachusetts State College Scholastic Aptitude Test over the Psychological Examination Artificial Language section. The same students were used for each test. But the latter-named section has the distinct advantage of practice effect. Nevertheless, when the French first term marks are compared with the two artificial language sections,

the hard fact remains: the Artificial Language section of the Psychological Examination entertains a "low" relationship, while the Artificial Language section of the Massachusetts State College Scholastic Aptitude Test shows a "substantial" relationship -as indicated by the obtained indices.

The next column contains the coefficients showing the relationships existing between the Mathematics individual scores and the various term marks. The first index, $.195 \pm .038$, is a measure of the degree to which a change in the individual scores of the Mathematics section tends to be accompanied by a change in the Orientation first term marks. The coefficient is less than 6 but more than 5 times the probable error. If a correlation here be assumed, it is designated "very low". Since the section is slightly less than 2% effective, it is practically useless for predicting Orientation first term marks.

The relationship existing between the Chemistry first term marks and the individual scores of the Mathematics section is denoted by the coefficient of correlation, $.24 \pm .038$. That a correlation is practically certain is indicated by the fact that the index is 6 times the probable error. Professor Trow would interpret the relationship as "present but low." In regard to the effectiveness of the section for predicting success in first term Chemistry work, the compliment of the coefficient of alienation expressed in per cent is but little

better than .02, indicating non-practicality.

The coefficient of correlation obtained by comparing the individual scores of the Mathematics section with the first term Mathematics marks is $.305 \pm .039$. The ratio existing between the coefficient and the probable error signifies that the section is reliable. The relationship here assumed is "present but low". For predicting academic success in first term Mathematics, the Mathematics section is 5% efficient -5% better than a "guess".

A measure of the degree to which a change in the Mathematics section individual scores tends to be accompanied by a change in the English first term marks is $.106 \pm .04$. As indicated by the relative size of the probable error, the test is not reliable; i.e., a relationship can not be assumed. For prediction, the section is approximately 1/2% effective.

The column entitled Memory Reading contains the indices existing between the individual scores of that section and the various first term marks. The first index is $.327 \pm .036$. This figure is a measure of the extent to which a change in the individual scores of the Memory Reading section tends to be accompanied by a change in the first term Orientation marks. In this instance, it is reasonably safe to assume a correlation -since the coefficient is more than 6 times as great as the probable error. The relationship is "present but low". With respect to prognostic efficiency, the test is slightly

more than 5% accurate.

The next index, $.16 \pm .039$, is found by correlating the Chemistry first term marks with the individual scores of the Memory Reading section. It is less than 6 but more than 4 times the probable error. If a relationship be assumed to exist between the attributes measured, it is very low. The accuracy of the Memory Reading section for predicting academic success in the first term Chemistry course is of impractical efficiency: being only 1% efficient.

The coefficient, $.168 \pm .04$, represents the relationship between the Memory Reading individual scores and the Mathematics first term marks. The index is less than 6 but more than 4 times the probable error. In the opinion of Professor Monroe, it is safe to assume a correlation. If it be assumed, it is very low. The prognostic efficiency of the Memory Reading section with respect to Mathematics first term marks is less than 2%.

When the English first term marks are compared with the individual scores of the Memory Reading section, a correlation of $.29 \pm .036$ is found. That the test is reliable is signified by the relative size of the probable error. For predicting first term English marks, the section is about 5% efficient; i.e., it is impractical.

The last column presenting the coefficients found by correlating the individual scores of the Power Reading section

with the various first term marks is entitled Power Reading. The first index, $.29 \text{ } -.036$, denotes the extent to which a change in the Power Reading section individual scores tends to be accompanied by a change in the Orientation first term marks. An inspection of the probable error shows that the section is dependable. Moreover, the assumed relationship is "present but low". Since, in this instance, the forecasting efficiency of the Power Reading section is slightly below 5%, it is practically useless for predicting academic success in first term Orientation.

The next coefficient is $.31 \text{ } -.036$. This figure is a measure of the degree of relationship existing between the individual scores of the Power Reading section and the first term Chemistry marks. The Power Reading section is dependable -as denoted by the relative size of the probable error. The existing relationship is "present but low." In regard to prediction, however, it is but slightly better than 5% efficient.

The index, $.15 \text{ } -.04$, signifies the degree to which a change in the Mathematics first term marks tends to accompany a change in the individual scores of the Power Reading section. The fact that the index is less than 4 times the probable error indicates that it is unsafe to assert a relationship between the two variables measured. For predicting academic success in first term Mathematics, the section is practically ineffective, being less than 1% efficient.

The last coefficient here presented is $.31 \pm .036$. It denotes the degree of relationship existing between the individual scores of the Power Reading section and the English first term marks. Being over 6 times as great as the probable error, the index is reliable. Hence, a relationship is practically certain. The correlation is present but low. The fact that the Power Reading section is only about 5% accurate, indicates that it is practically useless for predicting academic success in first-term, freshman English.

Summary and interpretation. In view of the method used and the results obtained in the secondary analysis, the following statements seem justifiable.

- (1) A composite of the gross scores is the most valid measure of academic success ($\underline{r} = .53 \pm .03$).
- (2) Of the three mental tests used in the study, the Psychological Examination is lowest in prognostic validity, ($\underline{r} = .347 \pm .035$), the Army Group Examination Alpha is second ($\underline{r} = .387 \pm .034$), and the Massachusetts State College Scholastic Aptitude Test ranks first ($\underline{r} = .488 \pm .033$).
- (3) While many of the indices secured among the gross scores and first term marks are significant, the majority indicate that the tests are of little prognostic value.
- (4) Of the coefficients obtained among the individual

scores and the first term average marks, only the Mathematics section of the Psychological Examination shows a correlation ($\underline{r} = .45 \pm .034$) above the normal ($\underline{r} = .40$), indicating that the Mathematics section, although the most valid, is (along with the other sections) practically useless for accurate prediction.

- (5) None of the indices existing among the individual scores of (1) the Psychological Examination and the Massachusetts State College Scholastic Aptitude Test and (2) the first term marks indicates a practical degree of forecasting efficiency. Of the two tests, the Massachusetts State College Scholastic Aptitude Test is superior.

Moreover, it is a highly significant, theoretically, to note that of the three tests here used, the rough model of the "learning" test is the most valid prognostic indicator. When compared with the normal, the Massachusetts State College Scholastic Aptitude Test alone transcends that value -and by nearly 9 points. Though far removed from perfection ($\underline{r}=1.00$), and hardly practical for prediction, the "learning" test evidences a substantial increase in forecasting efficiency over and above the tests based primarily on recognition and recall. With respect to the gross scores and term marks (though here, too, the results indicate impractical efficiency), the obtain-

ed indices point definitely to the greater validity of the Massachusetts State College Scholastic Aptitude Test. Is it, then, venturing beyond due bounds to assert that these meagre data suggest the probable truth of the a priori hypothesis set forth in Chapter I; namely, that "any indirect measurement must be as nearly as possible in harmony with the attribute to be predicted" if valid and reliable results are desirable? As implied in the foregoing question, the obtained results can do little more than suggest the answer -let alone clinch it.

The examination of the frequency distributions revealed that the range of the composite scores is greater than the range of the gross and individual scores. Therefore, since an increase in range, other factors being equal, increases the chances of a better correlation, and since the range of the composite scores is exactly four times as great as the range of the Massachusetts State College Scholastic Aptitude Test, it is possible that the highest obtained index falls to the composite because of these influences. Yet, the Psychological Examination gross scores have the next highest range and the lowest index when compared with the average marks. The wide range in the latter case is probably occasioned by (1) skillful weighting -thus increasing the range mechanically -and (2) practice effect. In regard to the effects of practice, certain test workers¹ have shown conclusively that practice generally

I. Glick, H.N., Effects of Practice, Bulletin No. 27, University of Illinois, 1925.

raises the scores¹ and because of variation in individual differences, e.g., in speed and accuracy -also the range. The Army Group Examination Alpha and the "learning" test have the lowest and the same range, 100. These facts raise the following question: Will an increase in the range of the Massachusetts State College Scholastic Aptitude Test (possibly secured by "weighting" and practice periods) raise its prognostic efficiency? Nothing less than further research can produce a satisfactory answer.

1. This contains the germ of an idea which may prove fruitful. If two forms of the same test be given to the entering freshmen, would the correlation index be raised accordingly? The testimony of controlled experimentation is awaited.

CHAPTER V

THE HIERARCHY OF INTELLIGENCE.

The problem and the method. By the hierarchy of intelligence¹ is meant that series of responses obtained in a novel situation, in which certain patterns are relatively superior, as measured by the extent of correlation existing among all specific response patterns. According to Professor Hull² of Yale University, in the final selection of the parts which go to make up a good prognostic test, the hierarchical measure, two principal considerations must be observed:

"(I) The tests should each correlate as highly with the aptitude criterion as possible.

"(II) They should correlate as low with each other as possible."

These pragmatic dicta are based on the following deduction: if two test correlate highly with the criterion and highly with each other, they are measuring essentially the same function; two tests producing a low correlation are measuring different traits. For prediction, then, it is a waste of time and labor to use two sections in the same test which correlate highly with each other. Thus, in estimating the validity and practicality of a test, it is necessary to compare each section of a test with the remaining sections.

1. C.F. Freeman, F.N., op. cit., p. 223.

2. Hull, C.L., Aptitude Testing, World Book Co., 1928, p. 254.

The technic employed in this procedure is again the product-moment method of correlation. In regard to results thus far obtained, Hull¹ states that human nature is so constituted that when high criterion correlations are secured, "reliable correlations extending below zero are rarely encountered." Hence, the standard, or goal, is a zero or negative correlation between sections correlating highly with the criterion, average marks. As regards a normal, the intercorrelations existing among the sections of the Psychological Examination² are used.

Results: the Psychological Examination. Table 9 shows the obtained coefficients (and the probable errors) existing among all of the five sections of the Psychological Examination. Two hundred and eighty-five cases are studied. The correlations range from $.14 \pm .039$ ("lowest") to $.68 \pm .022$ ("highest"). The average of the coefficients is .352. Although this figure is quite some distance from zero, the standard, it represents the relative achievement of the most widely used test toward the goal.

Of the 5 sections in the Psychological Examination, the highest correlation with the criterion (c) is achieved by the

1. Hull, C.L., op. cit., p. 450-1.

2. As stated elsewhere, the Psychological Examination has been subjected to rigorous statistical investigation, and revised in accordance with the results. The intercorrelations for the 1932 issue are probably available. Yet, in order to hold the factors involved (methods, number of cases, differences in administration, time of administration, etc.) as constant as may be, the writer deems it necessary to work out the actual indices from these data.

Arithmetic section (Ar.); i.e., $r(\text{Ar-G}) = .45 \pm .03$. The Completion section is next in validity, $r(\text{Cp-G}) = .327 \pm .036$. Here, then, are the two sections which correlate most highly with the criterion, the average marks for the first term. The coefficient existing between the Completion and the Arithmetic sections is $.41 \pm .034$ (Table 9). The interpretation of these facts is that the sections are measuring more the same thing than is desirable. Also, since the Completion and Arithmetic sections each correlate $.81 \pm .015$ with the gross scores, and since the gross scores correlate but $.347$ with the criterion, it may be stated that the Completion section is practically as valid as the Psychological Examination in predicting success for the first term. The same statement holds for the Artificial Language index, $r(\text{A.L.-G}) = .31 \pm .036$. And, as was indicated in Chapter IV, the Mathematics section is more valid than the gross test.

The coefficient obtained by comparing the individual scores of the Completion section with the Artificial Language section individual scores is $.32 \pm .035$. Table 6 shows the indices existing between the Completion section individual scores to be $.31 \pm .036$. These facts signify that the two sections here examined are measuring more of the same thing (what ever it maybe) than is desirable -since they measure each other to approximately the same degree which they measure the criterion.

TABLE 9

Intercorrelations among individual scores.

Psychological Examination

(N = 285)

	Completion	Artificial Language	Arithmetic	Analogies	Opposites	Gross
	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.	Coef. of P.E. Cor.
Completion	-----	.32 ±.035	.41 ±.034	.21 ±.038	.68 ±.022	.81 ±.015
Artificial Language	.32 ±.035	-----	.33 ±.035	.14 ±.039	.40 ±.034	.70 ±.021
Arithmetic	.41 ±.034	.33 ±.035	-----	.385 ±.034	.36 ±.035	.81 ±.015
Analogies	.21 ±.038	.14 ±.039	.385 ±.034	-----	.28 ±.036	.58 ±.026
Opposites	.68 ±.022	.40 ±.034	.36 ±.035	.28 ±.036	-----	.93 ±.005

The index, $.41 \pm .034$, indicates the extent to which a change in the Completion section scores tends to be accompanied by a change in the scores of the Arithmetic section. Since these sections correlate $.327 \pm .036$ and $.45 \pm .03$ respectively with the criterion, average marks, the intercorrelation coefficient shows that the two sections are measuring too much the same function.

The intercorrelation coefficient existing between the Arithmetic section individual scores and the individual scores of the Artificial Language section is $.33 \pm .035$. While this overlapping is rather large, the criterion correlations are $.45 \pm .03$ and $.31 \pm .036$ respectively -signifying that the situation is not altogether unsatisfactory.

The coefficient secured between the Analogies section individual scores and the Completion test individual scores is but $.21 \pm .038$. With respect to the criterion, the obtained index is $.259 \pm .037$ for the Analogies scores; for the Completion, $.327 \pm .036$. Though the present situation is admittedly far from the standard, it is relatively satisfactory. An examination of all three indices shows that the two sections under consideration tend to satisfy the first general maxim of Professor Hull -that the correlations between the criterion and the test sections should be relatively high. So they are, in this instance, with reference to the intercorrelation index. In regard to the second condition set forth by Dr. Hull, that

the correlation coefficient between two sections be as close to zero as may be -the coefficient $.21 \pm .039$ signifies a marked advance in this direction.

The correlation index obtained by comparing the individual scores of the Analogies section with the Arithmetic test individual scores is $.385 \pm .034$. Table 6 shows that the coefficients secured between the Analogies section scores and the criterion, average marks, is $.259 \pm .037$; that between the average marks and the Arithmetic individual scores, $.45 \pm .03$. Here, again, the total situation is not altogether satisfactory. The intercorrelation index is decidedly high, especially when one of the criterion coefficients is more than 12 points lower. Too much overlapping occurs between the two sections.

The index, $.68 \pm .022$, signifies the extent to which a change in the individual scores of the Completion test tends to accompany a change in the individual scores of the Opposites section. These sections also correlate $.327 \pm .036$ and $.14 \pm .036$ respectively with criterion. Obviously, the relationships are quite unsatisfactory. Although the correlation between the Opposites section scores and the average marks is but $.14$, the decisive factor is that the intercorrelation is $.68 \pm .022$. The most conservative of the statistical "stand-patters" would undoubtedly concede a substantial relationship between the two functions here measured. Moreover, since the criterion corre-

lation involved is markedly inferior, the great overlapping would seem to warrant the rejection of the Opposites section of the Psychological Examination -as far as the results of this investigation are concerned.

The intercorrelation coefficient secured between the Opposites section individual scores and the Artificial Language individual scores is $.40 \pm .034$. This means that there is a deal of overlapping between the two sections -with respect to the function measured. The criterion coefficients, $.14 \pm .039$ for the Opposites test and $.31 \pm .036$ for the Artificial Language section, only serve to emphasize the significance of the non-essentiality of the Opposites section -since it correlates $.14 \pm .039$ with the criterion and $.40 \pm .034$ with the Artificial Language test which, in turn, is a much more valid section when contrasted with college marks ($r = .31 \pm .036$).

In a similar manner, the secured index between the Opposites test individual scores and the Arithmetic test individual scores ($r = .36 \pm .035$) points to the necessary exclusion¹ of the section. The Opposites section scores correlation but $.14 \pm .039$ with the average marks, while the Arithmetic individual scores correlate $.45 \pm .03$. Not only are these sections

1. The computation of the correlation coefficient existing between the criterion and the Psychological Examination with the effect of the Opposites section partialled out is further needed to justify fully this recommendation.

measuring too much the same function, but also they deviate too markedly from the criterion; i.e., the Arithmetic test in the correct, and the Opposites section in the wrong, direction-as regard validity.

The coefficient of correlation obtained by comparing the individual scores of the Opposites section with the Analogies test individual scores is $.28 \pm .036$. Table 6 shows that the indices secured between the Analogies section individual score, and the criterion, average marks, is $.259 \pm .037$; that between the average marks and Opposites section, $.14 \pm .039$. In no way is this situation satisfactory. The intercorrelation coefficient is greater than either of the criterion indices. According to the dicta set forth in the beginning of the chapter, this is wholly undesirable. Instead of high criterion coefficients, relatively low ones exist. Instead of low correspondence between the two sections, a relatively high (high when compared with the criterion coefficients) index is in evidence.

Then, too, in regard to what has been stated concerning the necessary disqualification of the Opposites section -it may be added that the coefficient of correlation between the section individual scores and the gross scores of the Psychological Examination is $.93 \pm .005$ -signifying that they are measuring practically the same thing. The remaining indices (Table 9) range from $.58$ to $.81$, and measure the degree of relationship between the attributes measured by each section and the

gross scores.

Results: The Massachusetts State College Scholastic Aptitude Test. Table 10 shows the secured coefficients (and the probable errors) existing among the six sections of the Massachusetts State College Scholastic Aptitude Test. As in the case of the Psychological Examination, the responses of two hundred and eighty-five subjects are studied. The correlation indices range from $.13 \pm .039$ to $.34 \pm .035$. The average of the intercorrelation coefficients existing among the various sections of the Psychological Examination. The latter named index is henceforth designated the normal -since it is the average of the intercorrelation coefficients existing among the most widely used mental test, and is computed from indices obtained from the data of this research. However, since all the coefficients in Table 10 are below this value, with the possibility of one or two exceptions the same relative standard employed in the immediately previous comparison will obtain.

The index secured by comparing the individual scores of the Botany section with Artificial Language individual scores is $.33 \pm .035$. Table 6 reveals that the index existing between the individual scores of the Botany test and the criterion, average marks, is $.267 \pm .037$; that between the average marks and the Artificial Language test individual scores, $.315 \pm .036$. Though the intercorrelation value is below the normal ($.352$), the situation is quite unsatisfactory. Each of the criterion

TABLE 10
Intercorrelations among individual scores
Massachusetts State College Test

(N = 285)

	Botany		Artificial Language		Chemistry		Mathematics		Memory Reading		Power Reading		Gross	
	Coef.	P.E.	Coef.	P.E.	Coef.	P.E.	Coef.	P.E.	Coef.	P.E.	Coef.	P.E.	Coef.	P.E.
Botany	-----		.33	±.035	.199	±.038	.24	±.037	.34	±.035	.13	±.039	.646	±.025
Artificial Language	.33	±.035	-----		.225	±.037	.27	±.037	.207	±.038	.13	±.039	.608	±.026
Chemistry	.199	±.038	.225	±.037	-----		.23	±.037	.24	±.037	.205	±.038	.59	±.026
Mathematics	.24	±.037	.27	±.037	.23	±.037	-----		.24	±.037	.147	±.039	.565	±.028
Memory Reading	.34	±.035	.207	±.038	.24	±.037	.24	±.037	-----		.289	±.036	.53	±.029
Power Reading	.13	±.039	.13	±.039	.205	±.038	.147	±.039	.289	±.036	-----		.43	±.03

coefficients is inferior to the intercorrelation index, which is in direct opposition to the tenets advanced by Professor Hull. Not only is the overlapping too great, but also the criterion coefficients are relatively much too small.

The coefficient obtained between the Chemistry section individual scores and the Botany test individual scores is only $.199 \pm .038$. With respect to the criterion, the secured index is $.339 \pm .035$ for the Chemistry section individual scores; for the Botany test scores, $.267 \pm .037$. Although the relative positions of the three coefficients evidences a marked departure from the standard, they are fairly satisfactory. A critical examination shows that the two sections under consideration tend decidedly to comply with the general rule-of-thumb set forth by Dr. Hull; the criterion correlations are relatively high and the intercorrelation coefficient is relatively low. As stated previously no definite rule can be followed absolutely. For practical purposes, the best available must often serve as poor substitutes of the theoretical "best". In this instance, the relationship is practically satisfactory -since it is one of the "best available", and since it marks an advance toward the goal.

The index, $.225 \pm .037$, signifies the extent to which a change in the individual scores of the Botany section tends to accompany a change in the individual scores of the Artificial Language test. These sections also correlate $.267 \pm .037$ and

.315 \pm .036 respectively with the criterion average marks (Table 6). Here, too, the relative significance of the indices is rather satisfactory. The criterion coefficients are fairly high and the intercorrelation coefficient is substantially lower than either.

The coefficient obtained between the individual scores of the Mathematics section and the Botany section individual scores is .24 \pm .037. Since the relationship between the criterion and the Botany test scores is .267 \pm .037 and that between the Mathematics section scores is .257, the situation is hardly ideal. Though the criterion indices are greater than the intercorrelation, the margin is too narrow. That is to say, either the intercorrelation index is too large -indicating too much overlapping -or the criterion indices are too small according to Professor Hull. The judgment of the writer is that both adverse conditions here obtain.

The intercorrelation coefficient secured between the Mathematics section individual scores and the Artificial Language individual scores is .27 \pm .037. The criterion coefficients obtained between the two sections are .257 \pm .037 and .315 \pm .036 respectively. This means that there is entirely too much overlapping between the Mathematics and the Artificial Language sections of the test. The index between the criterion and the Artificial Language individual scores is but 4.5 points higher than the intercorrelation coefficient, and

the remaining index is below it.

In regard to the Mathematics section and the Chemistry test, the condition is more favorable. The intercorrelation index is $.23 \pm .037$. With respect to the criterion, the index is $.257 \pm .037$ for the Mathematics section individual scores and $.339 \pm .035$ for the Chemistry test scores. Thus, since the criterion indices are both greater than the intercorrelation coefficient, the condition is permissible.

The coefficient of correlation obtained by comparing the individual scores of the Memory Reading section and the Botany section individual scores is $.34 \pm .035$. Table 6 shows that the index existing between the criterion, average marks, and Memory Reading individual scores is $.307 \pm .036$; that between the Botany individual scores and average marks, $.267 \pm .037$. These proportions are obviously unsatisfactory. Though the correlation coefficient between the Memory Reading test scores and the criterion is rather high, here the decisive factor is the still higher intercorrelation index, $.34 \pm .035$ -signifying that the two sections considered are measuring the same thing to a relatively large, and therefore undesirable, degree. Such overlapping is subversive to a good, prognostic test.

A similar scrutiny of the index secured by comparing the Memory Reading individual scores and the individual scores of the Artificial Language section ($r = .207 \pm .038$) signifies that

here is an instance of relative promise. The criterion coefficients (Table 6) are $.307 \pm .036$ and $.315 \pm .036$ respectively. Each is well over 6 times the probable error, and therefore reliable. Each is relatively high; thus meeting the initial pandect of Dr. Hull. The intercorrelation coefficient is 10 points below the lesser of the two criterion indices. Here, indeed, is a relatively satisfactory situation.

The index secured by comparing the individual scores of the Memory Reading section with the Chemistry section individual scores is $.24 \pm .037$. The coefficient is at least 6 times the probable error, denoting dependability. The index existing between the Memory Reading test individual scores and the criterion, average marks, is $.307 \pm .036$; that between the average marks and the individual scores of the Chemistry section, $.339 \pm .035$. The intercorrelation coefficient is below the normal by ten points, which marks a substantial advance toward the standard, or goal. Each of the criterion coefficients is superior to the intercorrelation index by a good margin. These facts indicate the sections under consideration tend to conform to the general dictum (#1) set forth as a goal. The overlapping is not too great between the sections, nor are the criterion indices low enough relatively to occasion much derogatory criticism.

The coefficient obtained between the Memory Reading section individual scores and the Mathematics section individual

scores is $.24 \pm .037$ -the same figure secured by comparing the Chemistry section and the Memory Reading test individual scores. Since the probable error is less than one sixth the coefficient the result is reliable. With respect to the average marks, the secured index is $.257 \pm .037$ for the Mathematics section individual scores, and for the Memory Reading test individual scores, $r = .307 \pm .036$. As indicated by the probable error, both indices are dependable. No claim can here be made for a marked advance as regards the goal. Yet, the relative position of the coefficients indicate that the condition is permissible. Nevertheless, the coefficient existing between the criterion and the Mathematics test individual scores is but little better than the intercorrelation. The fact that it is superior, and the fact that the other criterion index is 5 points higher signify that for the initial attempt at a valid and practical prognostic test, the condition is not wholly unfavorable.

The intercorrelation coefficient secured between the Power Reading section individual scores and the Botany test individual scores is $.13 \pm .039$. Although this index is not satisfactorily reliable, it is the lowest obtained in the hierarchical investigation -and therefore, according to C. L. Hull, it is the best. The criterion coefficients previously calculated for the two sections are $.19 \pm .038$ (Power Reading) and $.267 \pm .037$ (Botany). Again, according to Professor Hull the

remarkable progress here made toward the intercorrelation goal is partly damped by the low criterion coefficient of the Power Reading section individual scores. However, since Botany-criterion index is more than 2 times as large as the intercorrelation index, and the criterion-Power Reading coefficient is larger by 6 points than the intercorrelation coefficient, the situation is quite satisfactory -at least for the present.

In regard to the Artificial Language section and the Power Reading test, the condition is somewhat more favorable. The intercorrelation index is the same, $.13 \pm .039$, and consequently receives the same interpretation. In a like manner, the criterion coefficient for the individual scores of the Power Reading section ($r = .19 \pm .039$) is identical -thus receiving the same interpretation. The progress is made in connection with the criterion index of the Artificial Language section individual scores. Here, the coefficient is $.315 \pm .036$. Since it is well over 6 times the probable error, it is reliable. Relatively speaking, the relationship here represented is the best attained in the present investigation. While the proportion is admittedly quite some distance from the goal, a definite advance is achieved. Therefore, in spite of the very low criterion coefficient of the Power Reading section individual scores, and although from these meagre data it is seemingly not significant for predicting scholastic success the writer contends that since something else is probably being

measured, the section should be retained, studied more in detail, and if possible improved for further practical application.

The coefficient of correlation obtained by comparing the individual scores of the Power Reading section with the Chemistry section individual scores is $.205 \pm .039$. Table 6 shows the index existing between the criterion, average marks, and the individual scores of the Power Reading section to be $.19 \pm .038$; that between the Chemistry section individual scores and average marks, $.339 \pm .035$. These proportions are quite one-sided. The relation between the Chemistry section and the criterion is relatively satisfactory when compared with the intercorrelation index. A different condition exists in regard to the remaining proportion. The intercorrelation index is, of course, rather large. Nevertheless, it is the low criterion coefficient of the Power Reading section individual scores, which is of crucial importance, since it throws the proportion askew.

A study of the index secured by comparing the Power Reading section and the Mathematics test individual scores ($r = .147 \pm .039$) in relation to the criterion coefficients reveals another relatively promising situation. It is of course true that the criterion coefficients fall far from the goal. Yet, since the proportion is favorable, the relationship is permissible. This means that the approval accorded here is

dependent upon the fact that criterion indices are relatively higher than the intercorrelation coefficient -thus showing that the two tests considered tend to approximate the second part of the accepted theory.

The index showing the relationship existing between the individual scores of the Power Reading section and the Memory Reading individual scores is $.289 \pm .036$. Though this figure is more than 5 points below the normal established for comparison, it is altogether too high when placed in relation to the criterion indices. The criterion indices are $.19 \pm .038$ and $.307 \pm .036$ respectively. That these proportions are undesirable is obvious. The criterion coefficient of the Power Reading section individual scores is almost 10 points less than the intercorrelation index. Then, too, the criterion index of the Memory Reading section individual scores is hardly enough above the intercorrelation coefficient to warrant any discrimination in favor of the situation. The condition here described is farther from the goal than are all the other existing conditions in the Massachusetts State College Scholastic Aptitude Test here described.

There is one point more to be considered before an attempt is made to evaluate the results of the present study. The individual sections must be studied relative to the test itself. In Table 10, under the column heading Gross, may be found the correlation indices (and the respective probable errors) se-

cured by comparing each group of individual scores with the gross scores of the learning test. The first index ($r = .646 \pm .025$) is a measure of the extent to which a change in the individual scores of the Botany section tend to be accompanied by a change in the gross scores. The interpretation of this correspondence is that they are measuring much the same thing. A severe limitation of this interpretation is the fact that the influence of the Botany section has not been partialled out of the gross test (this restriction holds for the remainder of the coefficients subsequently described).

The Artificial Language section coefficient is $.608 \pm .026$. Although the relationship between this section and the gross test is not so marked as was the gross-Botany index, the contribution of the Artificial Language section is more important for prediction -since the criterion-Artificial Language coefficient, $r = .315 \pm .036$, is greater than the other, $r (B-C) = .267 \pm .037$. The extent of this influence is beyond the scope of the present study.

In a similar way, the coefficient ($.59 \pm .028$) found by comparing the individual scores of the Chemistry section with the gross scores is less than the preceding indices. Here, the influence on prediction is probably better than either of the other two tests; $r (Chem-C) = .339 \pm .035$.

The index existing between the Mathematics section individual scores and the gross scores of the Massachusetts State

College Scholastic Aptitude Test is $.565 \pm .028$. The coefficient, like all others in this column, is well over 6 times the probable error, and hence dependable. The Mathematics test is also measuring much the same thing as is the gross test.

The Memory Reading section individual scores correlate $.53 \pm .029$ with the gross scores. The index is lower than the others, yet, the relationship indicated is definitely significant.

In the case of the Power Reading section, the relationship is less marked. The gross test index is $.43 \pm .03$. It is obvious that this section is measuring more the attribute which the test portends to measure than it is measuring academic success. It is just possible that the time element is a significant factor here. A glance at the gross column shows that the coefficients decrease directly with the passage of time. The writer is acquainted with no research which has to do with the influence of time per se on the relative significance of obtained coefficients.

Summary and Interpretation. The intercorrelations among the sections of the Psychological Examinations and the Massachusetts State College Scholastic Aptitude Test indicate -

- (I) That in general the sections of the Psychological Examination correlate more highly with each other than they do with the criterion. The average of the intercorrelations is $.352$; that of the correlations among

Individual scores and the criterion (first term average Marks) .297.

- (2) That in general the sections of the Massachusetts State College Scholastic Aptitude Test correlate substantially lower among themselves (average $r = .229$) than with the criterion (average $r = .280$).

In the light of the dicta of Professor Hull, it seems evident from the results of Chapters IV and V that the "learning" test is the "better" prognostic test since it (1) correlates 14 points higher with the criterion than does the test based primarily on recognition and recall, and since (2) the sections of the test in general have substantially lower intercorrelations (lower by 5.5 points). Thus in resolving the problem- Which is the more extensive Test? -the results point more to the Massachusetts State College Scholastic Aptitude Test than to the Psychological Examination.

Furthermore, the results of this study seem to indicate that the Opposites section of the Psychological Examination should be rejected (contingent upon the results of further research, however), and that possibly all of the Massachusetts State College Scholastic Aptitude Test sections be retained. The problem of "weighting" the accepted sections was not attempted.

The findings of the hierarchical investigation are quite disappointing to the writer. The chief reason for disappoint-

ment is the annoying incompleteness of the results. At the outset, it was hoped that a brief examination of the so-called hierarchy of intelligence would (as happens more often than not) reveal at least certain tendencies. The obtained results indicate definitely that the procedure is quite inadequate for these data. In order to secure useful information in regard to the possibilities of varying combinations of the test sections used in this study, more extensive research is essential. The problem might well be taken up where the present one was left off. The intercorrelation results here set forth form a solid foundation for the multiple correlation study by which the possibilities of all combinations may be determined. The question of maximum correlations once resolved, the natural succedanea are (1) the problem of "weighting" (2) test revision (if necessary), and (3) further experimentation.

CHAPTER VI

PERCENTILE INTERPRETATION

The problem and the method. The results of the foregoing chapters seem to indicate, in general, that the mental tests employed in this study are of little practical use in predicting academic success (for the first term, at least). Consequently, in order the better to evaluate their worth, it is necessary to investigate further the validity of these so-called prognostic tests. This is done by means of percentile comparison. The method used in the percentile comparison and interpretation is that of computing the coefficient of correspondence. Odell¹ defines this index as a measure of the relationship existing between the scores in two series. It is determined by comparing directly the number of items (marks, scores, etc.) in the same percentiles for both functions with the total number of items in each percentile. For example, the marks and scores are ranked in order, from the lowest to the highest. If the percentile be a quartile, or quarter, each series is divided into fourths. The lowest quarter is called the first quartile, the highest quarter, the fourth quartile. The number of corresponding measures which fall in (say) the first quartile of each array is computed, and divided by the total number of cases in that quartile. The obtained quotient is the coefficient of correspondence.

1. Odell, C.W., op. cit., pp. 209-305

Table 11 contains a presentation of the number of cases in which the first term average marks and the mental test composite scores fall in the same quartile or differ by one, two or three quartiles. The column headed Total Misplacement indicates the number of cases which are not in the same average mark and mental test quartiles. The next column contains the point misplacements. The point-misplacement index is found by summing the moments of quartile disagreement. For example, the number of cases differing by one quartile is multiplied by one, the number differing by two quartiles is doubled, etc. The sum of these products represents the total point misplacement. Obviously, then, the lower the total and the point misplacements -the more valid is the test. The method for computing the coefficient of correspondence has already been given.

First quartile comparison between average marks and mental test scores. Table 11 shows the extent of correspondence existing between the mental test (composite and gross) scores and the criterion, average marks. The Army Group Examination Alpha has the lowest coefficient of correspondence, .364; the Psychological Examination, the highest, .400. Here, the Massachusetts State College Scholastic Aptitude Test receives second place. However, with respect to the point misplacement, the "learning" test is first. This means that the gross scores of the Massachusetts State College Scholastic Aptitude Test deviate less (in terms of quartiles) from the criterion array

TABLE 11

Quartile comparison showing degree of correspondence between first term average marks and mental test composite and gross scores.

Mental Tests	First Term Average Marks					Coefficient of Correspondence
	Same Quartile	Differ by One Quartile	Differ by Two Quartiles	Differ by Three Quartiles	Total Misplacement	
Army Group Examination Alpha (gross)	104	109	48	24	1818	253 .364
Massachusetts State College Scholastic Aptitude Test (gross)	107	129	38	11	178	238 .375
Psychological Examination (gross)	114	107	48	16	171	251 .400
Composite	106	117	57	5	179	241 .370

(285 cases)

than does any one of the three other tests here considered. The Psychological Examination has the lowest total misplacement, while the Army Group Examination Alpha shows the highest. In regard to the number of cases which differed by three quartiles, the Composite ranks best; the Massachusetts State College Scholastic Aptitude Test, second; the Psychological Examination, third; and the Army Group Examination Alpha, last.

A further examination of the quartile analysis (Table 11) reveals that with 74% accuracy the Army Group Examination Alpha predicts the relative position of each student within a quartile. The Psychological Examination predicts the same relative position within a quartile with 77% accuracy; the Composite with 79% accuracy. Only 18% of the students differ by 2 or 3 quartiles in regard to the Massachusetts State College Scholastic Aptitude Test average scores and the first term average marks. Hence, it follows that this test is the best of the four prognostic indicators under consideration -since with respect to the relative position of each student within a quartile, the "learning" test forecasts 82% efficiently. Of course, as already noted, it is a fact that the Psychological Examination obtains the highest coefficient of correspondence. Nevertheless, the whole truth must include the fact that it has the second highest point misplacement, and that it has 64 cases, the relative position of which differ by 2 or 3 quartiles: the Massachusetts State College Scholastic Aptitude

has but 49.

Second quartile comparison. Tables 12-15 show the respective quartile comparisons between the first term average marks and (1) the Army Group Examination Alpha, (2) the Massachusetts State College Scholastic Aptitude Test gross scores, (3) the gross scores of the Psychological Examination, (4) Composite scores. In this particular study, the average marks and the gross and composite scores are compared directly, quartile by quartile.

Table 12 presents the results of the quartile comparison between the first term average marks and the gross scores of the Army Group Examination Alpha. The number of students who received fourth-quartile scores, i.e., in highest quarter, on this test and fourth-quartile average marks is 41. This means that of the 69 students in receiving marks in the upper fourth of the first term, 59% were predicted by the test under consideration. It should be noticed that the highest and lowest quartiles receive the highest coefficients of correspondence. Evidently, the first and fourth quartiles are more efficient than the middle quarters.¹ From a practical point of view, the lowest (first) quartile is the most important -since it is with

1. It should be noted, however, that the influence of chance on the scores and marks of second and third quartiles is probably greater than it is on the first and fourth quartile marks and scores.

TABLE 12

Quartile comparison between the first term average marks and the Army Group Examination Alpha gross scores.

Army Group Examination Alpha							
Average Marks	Fourth Quartile	Third Quartile	Second Quartile	First Quartile	Total Mis- place- ment	Total Mis- place- ment	Coefficient of Correspondence
Fourth	<u>41</u>	13	5	10	28	53	.597
Third	20	<u>22</u>	22	5	47	54	.305
Second	18	20	<u>12</u>	19	55	74	.174
First	14	19	16	<u>29</u>	49	94	.371
(285 cases)							

this quartile¹ that the Office of the Dean is usually most concerned. Incidentally, since the mental tests are used by many colleges in place of entrance examinations, it is significant to note the validity of each test as regards the prediction of failure along with academic success. The coefficient of correspondence shows that the Army Group Examination Alpha predicted the number of first quartile average marks 37% accurately. This test is much more efficient in the selection of good students than it is in selecting poor ones.

Table 13 shows the results obtained by comparing the first term average marks with the Massachusetts State College Scholastic Aptitude Test by means of quartiles. Thirty-two cases were accurately placed in the fourth quartile.² The coefficient of correspondence signifies that this placement is 46% efficient. The Army Group Examination Alpha is more accurate in predicting students of honor grade than is the "learning" tests. On the other hand, thirty-seven marks below 65 were forecast by the Massachusetts State College Scholastic Aptitude Test. The coefficient of correspondence indicates that the selection is 47% accurate. This means that the Massachusetts State College Scholastic Aptitude Test predicts "good" and "poor" students equally well. Although the Army Group

1. In this study, the lowest quartile includes all average marks below 65.

2. All average marks in the fourth quartile are above 79 -i.e., marks of honor grade.

TABLE 13

Quartile comparison between the first term average marks and the Massachusetts State College Scholastic Aptitude Test gross scores.

Massachusetts State College Scholastic Aptitude Test

Average Marks	Fourth Quartile	Third Quartile	Second Quartile	First Quartile	Total Misplacement	Total Misplacement	Coefficient of Correspondence
Fourth	<u>32</u>	24	5	8	37	58	.463
Third	22	<u>22</u>	16	9	47	56	.318
Second	12	11	<u>16</u>	30	53	65	.232
First	3	12	26	<u>37</u>	41	59	.474

(285 cases)

Examination Alpha forecasts the students to receive honor grades more accurately than does the Massachusetts State College Scholastic Aptitude Test, in regard to predicting "poor" students, the latter test is (in this instance) unquestionably superior. It is also quite significant that only 3 of the students in the lowest average mark quartile were in the first quartile of the Massachusetts State College Scholastic Aptitude Test.

Table 14 contains the results of the quartile comparison made between the first term average marks and the Psychological Examination. Since 44 "good" students were accurately placed, the coefficient of correspondence is .637. This test is most efficient in predicting honor students. It is also rather efficient (46%) in selecting the students whose average mark for the first term is less than 65. That the Psychological Examination is more valid than the Army Group Examination Alpha for predicting the achievement of both "good" and "poor" students is obvious. Just as readily apparent is the fact that the Psychological Examination is more accurate in forecasting the relative standing of honor students than is the Massachusetts State College Scholastic Aptitude Test. The next question is: Of the two latter-named tests, which is the superior in regard to the selection of "poor" students? There is no crux here. In spite of the fact that the Massachusetts State College Scholastic Aptitude Test is but one point higher (in per cent) than

TABLE 14

Quartile comparison between the first term average marks and the Psychological Examination gross scores.

Psychological Examination

Average Marks	Fourth Quartile	Third Quartile	Second Quartile	First Quartile	Total Misplacement	Point Misplacement	Coefficient of Correspondence
Fourth	<u>44</u>	10	12	3	25	43	.637
Third	12	<u>18</u>	23	16	51	67	.261
Second	9	24	<u>16</u>	20	53	62	.231
First	13	11	18	<u>36</u>	42	79	.461

(285 cases)

the Psychological Examination, the writer does not hesitate to assert that the evidence tends to favor the former. The deciding factor is the point misplacement index within the first quartile. For the first quartile of the average marks, the Psychological Examination has a point misplacement of 79; the Massachusetts State College Scholastic Aptitude Test, 59. The chances are that the Massachusetts State College Scholastic Aptitude Test predicts "poor" students more accurately than the Psychological Examination.

Table 15 shows the results of the quartile comparison between the first term average marks and the composite scores. In the fourth -the highest-quartile, 41 accurate placements are made -indicating that the Composite is 58% efficient in regard to the selection of honor students. In predicting poor students, the Composite is less accurate. But 34 students were correctly placed. The coefficient of correspondence is .435.

First term failures. At the close of the first term of the 1931-1932 academic year, 25 students were dismissed because of failure to do work of satisfactory grade. To what extent the mental test composite and gross scores predicted these failures may be seen in Table 16. The lowest quartile of the Army Group Examination Alpha placed 11 or the 25 failures below the level of the 65 average mark. Only 4 are found in the next quartile; i.e., misplaced by one quartile.

The Psychological Examination placed 13, or 52% below the

TABLE 15

Quartile comparison between the first term average marks and the Composite Scores.

Composite Scores.

Average Marks	Fourth Quartile	Third Quartile	Second Quartile	First Quartile	Total Misplacement	Point Misplacement	Coefficient of Correspondence
Fourth	<u>41</u>	11	17	0	28	45	.576
Third	25	<u>12</u>	26	6	57	63	.173
Second	12	20	<u>19</u>	18	50	62	.275
First	5	22	17	<u>34</u>	44	76	.435

(285 cases)

TABLE 16

The mental test quartile position of the twenty-five first term failures.

Mental Test	Fourth Quartile	Third Quartile	Second Quartile	First Quartile	Coefficient of Correspondence
Army Group Examination Alpha	4	6	4	11	.44
Psychological Examination	2	7	3	13	.52
Massachusetts State College Scholastic Aptitude Test	0	3	6	16	.64
Composite	2	5	5	12	.48

65 average mark level. Nevertheless, but 3 cases are found in the second quartile, and 9 are in the upper two quarters.

Sixteen of the failures were selected to receive an average mark less than 65, by the learning test. This means that 64% of the failures for the first term of the 1931-1932 academic year were placed in the lowest quartile of the Massachusetts State College Scholastic Aptitude Test. It is also significant that 6 cases are found in the second quartile. Eighty-eight per cent of all failures are found in the lower quartiles of the Massachusetts State College Scholastic Aptitude Test gross scores. Then, too, no student who was dismissed from the College for poor scholarship at the close of the first term received a score which placed him in the upper fourth of the "learning" test.

A composite of the mental test gross scores placed 12 of the failures in the lowest quartile. The coefficient of correspondence for this placement is .48. Although more accurate than the Army Group Examination Alpha and the Psychological Examination, the Composite scores are less efficient than the gross scores of the Massachusetts State College Scholastic Aptitude Test.

Summary and Interpretation. The results of the quartile analysis indicate:

(1) that the Psychological Examination is the best indicator of honor students (average mark above 79); the Composite

and the Army Group Examination Alpha rank second, and the "learning" test is the least accurate indicator of high scholarship;

(2) that, of the three tests employed, the Massachusetts State College Scholastic Aptitude Test is by far the most efficient in predicting failures, the Psychological Examination is next in accuracy, the Composite is third and the Army Group Examination Alpha is the least efficient.

Other than the fact that substantial progress has been made, there is, of course, little practical significance attached to the results of the percentile analysis. The high coefficient of correspondence (.64) achieved by the first quartile placement of the Massachusetts State College Scholastic Aptitude Test gross score means that 16 of the 25 failures for this term are found somewhere among the lowest quarter of the entire Freshman Class. At present, there is no satisfactory answer to the question: What proportion of good students should be excluded in order to eliminate a given number of poor ones?

CHAPTER VII

SUMMARY AND CONCLUSIONS

The major purpose of this research is to investigate the relative validity of different types of subject matter in certain scholastic aptitude tests for college freshmen. The tests based on recall and recognition are the Army Group Examination Alpha and the Psychological Examination. No tests based primarily on subject matter which called forth learning responses were available. To meet this contingency, Professor H. N. Glick and the writer prepared an empirical, "learning" test, the Massachusetts State College Scholastic Aptitude Test. Subject to certain limitations, the comparative validity of these tests with respect to predicting scholastic success for first term of the 1931-1932 academic year is here determined.

SUMMARY

The preliminary analysis. A qualitative examination in regard to the general nature of the sole criterion, academic success, reveals that the first and second term average marks do not conform to the law of normal distribution, and, with few exceptions, the distributions of the first and second term marks are also "irregular". A similar inspection of the composite, gross, and individual scores shows that here, too

equally noticeable discrepancies obtain.

The secondary analysis. The results of the secondary analysis show that a composite of the gross scores is the most valid measure of academic success. Of the three mental tests used in this study, the Psychological Examination is the least valid; the Army Group Examination Alpha is somewhat better; and the Massachusetts State College Scholastic Aptitude Test ranks first. As regards predicting success in a term course, the tests are practically useless. Nor are the test sections (excepting the sections of the Army Group Examination Alpha, which were not considered) of practical use for predicting general success in the first term and success in a single first course of study. However, the Massachusetts State College Scholastic Aptitude Test evidences a substantial increase in forecasting efficiency over and above the tests based essentially on recognition and recall; i.e., the Army Group Examination Alpha and the Psychological Examination.

The hierarchical investigation. The results of this special study are subject to many limitations. Nevertheless, the evidence at hand indicates in general that the Massachusetts State College Scholastic Aptitude Test contains more "extensive" items than do the other two tests employed in the investigation.

The percentile analysis. An examination of the validity of the tests by means of quartile comparisons shows that the

Psychological Examination is the best measure of upper-fourth average mark division. In regard to the selection of students whose first term average mark is less than 65, the Massachusetts State College Scholastic Aptitude Test is the most valid indicator. Again, the Massachusetts State College Scholastic Aptitude Test predicts failures more efficiently than do the Army Group Examination Alpha and the Psychological Examination.

CONCLUSIONS

That a test essentially measuring ability to learn is more valid for predicting success in the first term of the freshman year than a test based essentially on recall and recognition is the major conclusion to be drawn from this investigation. Not only is the "learning" test superior to specific tests (e.g., the Psychological Examination and the Army Group Examination Alpha), but also it is superior to the majority of the so-called scholastic aptitude tests now widely used to predict college success. That the Massachusetts State College Scholastic Aptitude Test, a specific aptitude measure, selects the first term failures more efficiently than the most popular prognostic indicators of college level is also a fact.

In regard to the question of practicality, it may be said that both kinds of tests are of little use for predicting the accomplishment of particular individuals. That is to say, the mental tests are of little use to the school administrator who

desires to know in advance the "standing" of a certain student (say) at the end of a term of study. On the other hand, the mental tests provide the most valid measure of future academic success now available. Secondary school average marks correlate only $.35^1$ with first year marks in college. ✓ The recommendations of school principals cannot be used because many recommend all their students for admission while others maintain such high standards that only one-fourth of their students are eligible. The rating of individual schools according to the success of their graduates is unsatisfactory, since a college environment may stimulate the student who was "indifferent" in high school work. "College board examination" grades correlate less than $.40^2$ with college marks, and are ✓ therefore practically worthless for predicting college success. What, then, is the college administrator to use as a criterion³ when he is called upon to decide whether or not a "border-line" student is to remain in college?

1. Odell, C.W., "Predicting the Scholastic Success of College Freshmen", Bureau of Educational Research, University of Illinois, Bulletin Number 37, 1927.

2. Crane, E., "An Investigation of Three Plans for Selecting the Students to be Admitted to College", Journal of Educational Psychology, 1926 (17), pp. 322-330.

3. As referred previously, another use of the mental test is to employ it in lieu of the entrance examination. Of late, many institutions of higher learning have had more applicants for admission than they could accept. Here, again, the practical question arises: How can the academic sheep be separated from the goats? Only the results of controlled experimentation can answer.

Since the mental tests are more valid than any other available indicator, and since, of the available tests, the Massachusetts State College Scholastic Aptitude Test is the most efficient for predicting first term scholastic success, it is recommended that this test, or tests composed primarily of subject matter which calls forth learning responses, be used by college administrators -until experimental research produces more fruitful results.

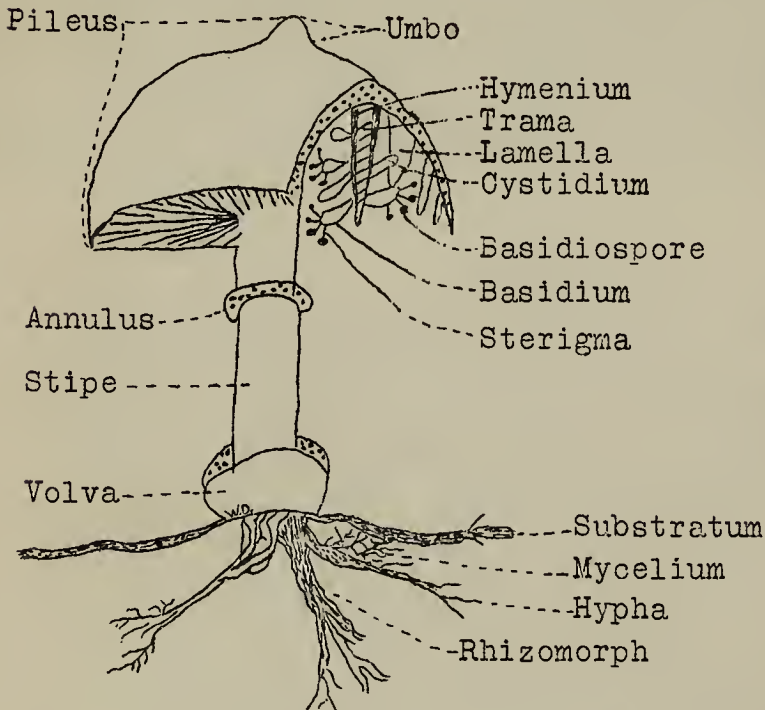
APPENDIX

Section I

Directions:

Study thoroughly the drawing and the printed material on this sheet. You will be asked questions about it later. You will not be asked to reproduce the drawing.

Part A
An Agaric (Mushroom)



Part B

1. The neurone is the structural and functional unit of the nervous system.

2. The neuroglia is the supporting tissue for the neurones.

3. The falx cerebri is the sickle-shaped median partition situated in the fissure between the hemispheres of the brain.

4. Dolichocephalic animals have long narrow skulls.

5. A fungus is a plant which has neither root, stem nor leaf, is void of green chlorophyll and reproduces by spores.

6. A seed is a reproductive body which contains a miniature plant.
7. An ascomycete is a fungus which has its spores in a sack-like organ, an ascus.
8. A pleomorphic bacterium has two or more forms: one may be spherical, the other rod-shaped.
9. A conidium is a reproductive body borne on the end of a stalk and does not contain a miniature plant.
10. Symbiosis is the living in contact of plants and animals for mutual benefit of each.
11. The periorbital is a conical fibrous membrane which encloses the eyeball with its muscles, vessels, nerves, etc.
12. The corpora quadrigemina are the four rounded eminences which lie under the posterior part of the cerebral hemispheres.
13. The collosal sulcus separates the corpus callosum from the gyrus fornicatus.
14. A clone is a mutation within a monsporous isolation.

SECTION II.

Directions: Study carefully the selections given below. You will be asked questions about them later. You will not be expected to commit the selections to memory.

Part A

Morte d'Arthur, by Sir Thomas Malory (c. 1400-1471), was published by Caxton in 1485. Malory's sources were of French origin and were about ten times as long as his own romance. His account of King Arthur approximates an epic view. The work is courtly, simple, fresh in diction, picturesque in expression, and free from affectation. There are frequent blunders in syntax, however, and the style is often marred by lack of paragraph and sentence structure.

Part B

O world, thou choosest not the better part!
It is not wisdom to be only wise,
And on the inward vision close the eyes,
But it is wisdom to believe the heart.
Columbus found a world, and had no chart,
Save one that faith deciphered in the skies;
To trust the soul's invincible surmise
Was all his science and his only art.
Our knowledge is a torch of smoky pine
That lights the pathway but one step ahead
Across a void of mystery and dread.
Bid, then, the tender light of faith to shine
By which alone the mortal heart is led
Unto the thinking of the thought divine.

Santayana

Massachusetts State College Scholastic Aptitude Test

FOR COLLEGE FRESHMEN

Prepared by H. N. Glick and Alfred H. Holway

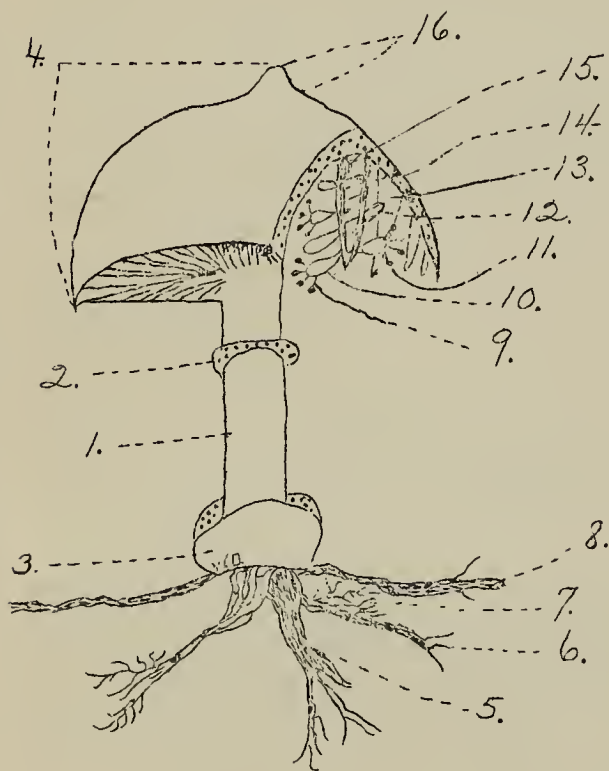
Name.....
 (Last name) (Given names or initials)
Age..... Last school attended.....

	Score
Section I.....
Section II.....
Section III.....
Section IV.....
Section V.....
Section VI.....
Gross Score.....

SECTION I

Directions (Part A):

Here is the drawing you studied; the parts are numbered and the names of the parts are at the right of the drawing. You are to copy the number of each part in the parentheses after the name of that part. Examples: Number 1 is the stipe, so 1 is placed in the parentheses after stipe. Do nothing with the names of parts which do not belong to this drawing.



Stipe(1)
Annulus.....()
Sterigma.....()
Pileus.....()
Substratum()
Mycelium()
Basidiospore...()
Mesodermis.....()
Lamella.....()
Rhizomorph()
Umbo.....()
Hypha.....()
Basidium.....()
Hymenium()
Stipule.....()
Cystidium()
Volva.....()
Trama.....()

Do not stop. Go to next sheet without further instructions.

SECTION I (continued)

Directions: (Part B)

Below are the definitions which you studied with the terms defined left out. The terms defined are in the list of terms below. What you are to do is to copy the number of the definition in the parentheses after the term which the definition best defines. Example: Definition number 1 defines neurone, so 1 is placed in the parentheses after neurone. Do nothing with the terms in the list which are not defined by any of the definitions.

1. The ---- is the structural and functional unit of the nervous system.
2. The ---- is the supporting tissue for the neurones.
3. The ---- is the sickle-shaped median partition situated in the fissure between the hemispheres of the brain.
4. ---- animals have long narrow skulls.
5. A ---- is a plant which has neither root, stem nor leaf, is void of green chlorophyll and reproduces by spores.
6. A ---- is a reproductive body which contains a miniature plant.
7. An ---- is a fungus which has its spores in a sack-like organ, an ascus.
8. A ---- bacterium has two or more forms; one may be spherical, the other rod-shaped.
9. A ---- is a reproductive body borne on the end of a stalk and does not contain a miniature plant.
10. ---- is the living in contact of plants and animals for mutual benefit of each.
11. The ---- is a conical fibrous membrane which encloses the eyeball with its muscles, vessels, nerves, etc.
12. The ---- are the four rounded eminences which lie under the posterior part of the cerebral hemispheres.
13. The ---- separates the corpus callosum from the gyrus fornicatus.
14. A ---- is a mutation within a monsporous isolation.

neurone.....()	clone()
ascomycete.....()	pleomorphic.....()
dolichocephalic.....()	corpora quadrigemina.....()
conidium.....()	brachycephalic.....()
falx cerebri.....()	symbiosis()
zygote.....()	seed()
periorbita.....()	neuroglia.....()
collosal sulcus.....()	fungus.....()

SECTION II

Directions: On this sheet we have a vocabulary, some rules and some sample sentences of an artificial language. On the opposite page are some English sentences and just beneath each English sentence is its translation into the artificial language. Some of these translations are correct and some are incorrect. You are to study the language on this sheet and draw a line through every word which is incorrectly translated on the opposite sheet. Do not try to memorize the vocabulary and forms on this sheet but consult them freely while checking the translation. If correctly translated words are marked, they will count against you.

VOCABULARY

RULES.

I--y
you--ye
he--ul
to--ig
at--ik
that--tuc
the--tuk
remember--noi
live--par
is--uk
give--ko
affirm--nu
go--di
girl--sen
boy--tec
book--lik
statement--tat
medicine--dem
home--moh
long--su
good--lo
bitter--to
difficult--psi
large--kno

1. Plurals are formed by adding "z".
Only pronouns have plurals.
Examples: we--yz
 they--ulz
2. Past time is indicated by placing
"ar" before the verb.
Examples: gave--arko
 went--ardi
3. Opposites are formed by adding "ng"
to the word.
Examples: affirm--nu; deny--nung
 difficult--psi; easy--psing
4. The objective case is indicated by
placing "ob" before the noun or
pronoun.
Examples: him--obul
 them--obulz

Samples: (The incorrect translations
 are marked).

- A. I gave the book to you.
y ke tuk ~~lik~~ ig obye.
- B. I gave the book to you.
Y arko ~~tue~~ oblik ig ye.

Do not stop. Go to next sheet without further instructions.

Section II (Continued)

1. I gave the book to him.
Ul arko tuk oblik ig obul.
2. He gives the book to me.
Ul ko tuk oblik ik oby.
3. He lives at home.
Ul par ik obmoh.
4. The medicine is bitter.
Tuk obdem uk to.
5. The book is at home.
Ul lik uk ik obmoh.
6. The statement is difficult.
Tec obtat uk psi.
7. The boy went home.
Tuk tec arding obmoh.
8. He lived long.
Ul arpar su.
9. The book is small.
Tuk lik ik knong.
10. He denied the statement.
Ye arnung tuk tat.
11. The bad girl took the large book.
Tuk lo sen arkong tuk kno oblik.
12. The little girl died at home.
Tuk knong sen arparng ik moh.
13. He is giving the book to the girl.
Ul ko tuk oblik ig tuk obsen.
14. He remembered the long statement.
Ul armoing tuk su obtat.
15. The girl forgot the bitter medicine.
Tuk obsen arnoing tuk to obdem.
16. The small book is easy.
Tuk knong oblik uk psing.
17. The medicine was sweet.
Tuk obdem aruk tong.
18. I remember that the book was difficult.
Y arnoi tuc tuk lik aruk psing.
19. He denied that the statement was difficult.
Ul arnung tuc tuk obtat aruk psi.
20. He took the medicine; he died.
Ye arkong tuk obdem; ul arparng.
21. He forgot that the medicine was bitter.
Ul arnoing tuc tuk obdem aruk to.
22. We denied that he gave the book to us.
Y arnung tuc ul arko tuk oblik ig oby.
23. They gave the medicine to the girl.
Ulz arko tuk dem ig tuk obsen.
24. I remembered that I lived at home.
Y noi tuc Y par ik obmoh.
25. I gave the books to the boys at home.
Y arko tuk oblikz ig tuk obtec ik obmoh. Stop here.

SECTION III

Directions: Part A. Below is the substance of the first selection you read with some numbered blanks indicating that some of the words are left out. The words left out are listed below. You are to copy the number of each blank in the parentheses after the word which belongs in the blank.

(1) was written by (2). The book was published in (3) by (4). The author's sources were of (5) origin. His account of (6) approximates an (7) view. The work is (8), simple, (9) in diction, (10) in expression, and free from (11). It contains frequent blunders in (12), however, and the (13) is often (14) by lack of paragraph and sentence (15).

Morte d'Arthur.....()	French.....()
courtly()	picturesque.....()
structure.....()	syntax.....()
Thomas Malory.....()	1485.....()
fresh()	King Arthur.....()
marred.....()	affectation.....()
Caxton.....()	style.....()
knightly.....()	epic.....()
1548.....()	crystal.....()
chivalry.....()	

Part B

Directions: Answer the following questions according to the poem which you read. If a statement is true, check (✓) true; if it is false check false; if the poem doesn't say whether a statement is true or false, check didn't say. The samples are checked correctly.

Samples:	true	false	didn't say.
1. Columbus used a chart.		✓	
2. Columbus had no chart.	✓		
3. Columbus used a compass.			✓
1. The poem implies the existence of God			
2. Believe nothing which science cannot prove.			
3. The world is full of mystery & dread.			
4. Life is not worth living.			
5. It may be better to follow inward visions than to follow knowledge.			
6. Immortality is certain.			
7. Columbus acted more upon faith than upon knowledge.			
8. Faith will never lead to knowledge.			
9. The highest wisdom may come from the heart.			
10. We become wise only through the acquisition of knowledge.			
11. Divine thoughts can be reached only through faith.			
12. The past is a void of mystery and dread.			

Stop
here

SECTION IV

Directions: This section of the test is divided into four parts. In each part some facts and principles are stated, and these are followed by some problems. You are to study the facts and principles and then solve the problems. In every part the solution of the problems is based upon the facts and principles in the same part.

PART A

Facts and Principles. Elements which differ in atomic weight but not in atomic number are called isotopes. Elements which differ in atomic number but not in atomic weight are called isobars. Following are the names of some elements with their atomic weights and numbers:

Elements	Atomic weight	Atomic number
Uranium X ₁	234	90
Uranium 2 ¹	234	92
Ionium	230	90
Radium	226	88

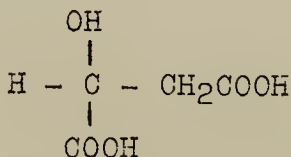
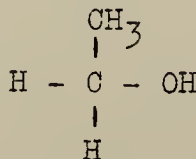
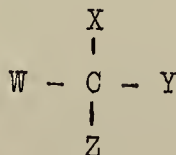
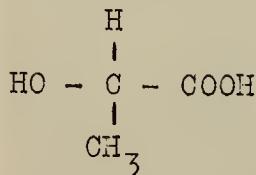
Problems Indicate by check (✓) whether the following pairs of elements are isotopes, isobars or neither.

Elements	Isotopes	Isobars	Neither
Uranium X ₁ + Ionium	_____	_____	_____
Uranium X ₁ + Uranium 2	_____	_____	_____
Uranium 2 ¹ + Ionium	_____	_____	_____
Uranium X ₁ + Radium	_____	_____	_____
Uranium 2 ¹ + Radium	_____	_____	_____

PART B

Facts and Principles. If a carbon atom (C stands for carbon atom) is attached to four different elements or groups of elements, the carbon atom is said to be asymmetric.

Problems. Place a cross (X) on each carbon atom which is asymmetric.



Do Not Stop. Go to next sheet without further instructions.

Section IV (continued)

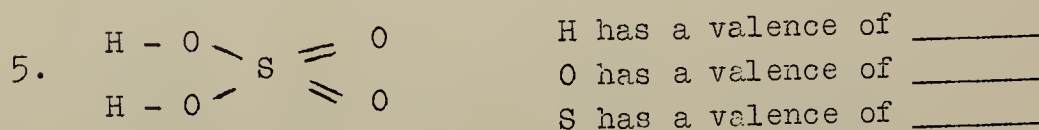
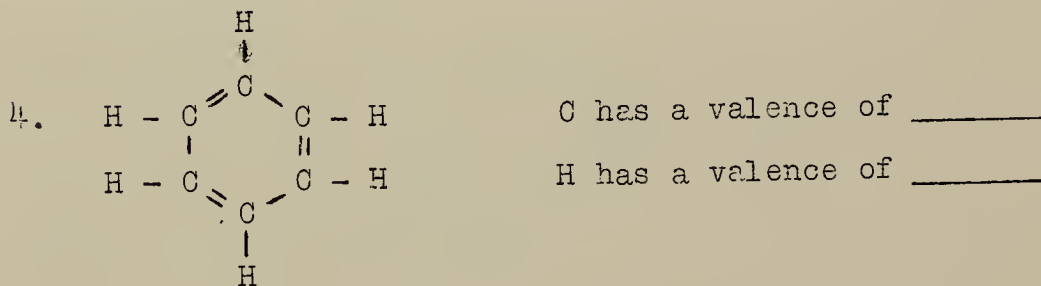
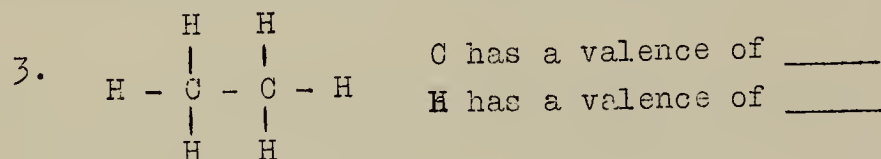
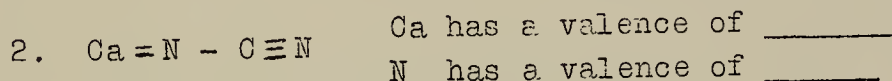
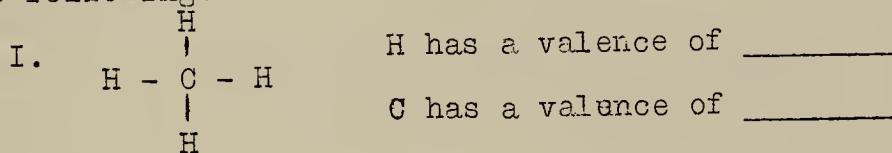
Part C.

Facts and Principles: The valence of an element is the number of atoms of hydrogen or chlorine which one atom of the given element combines with or displaces. Valence is frequently marked by lines pointing toward the symbol for an element. The number of lines in each case indicates the valence of the atom which the symbol represents. (Note: The symbols for elements are either large letters standing alone or else a large letter and a small letter together). For example:

H - Cl. Here H & Cl each have a valence of 1.

$\begin{array}{c} \text{O} \\ \parallel \\ \text{S} \\ \parallel \\ \text{O} \end{array}$ Here S has a valence 4 & O a valence of 2.

Problems. Determine the valence of the atoms indicated in the following:

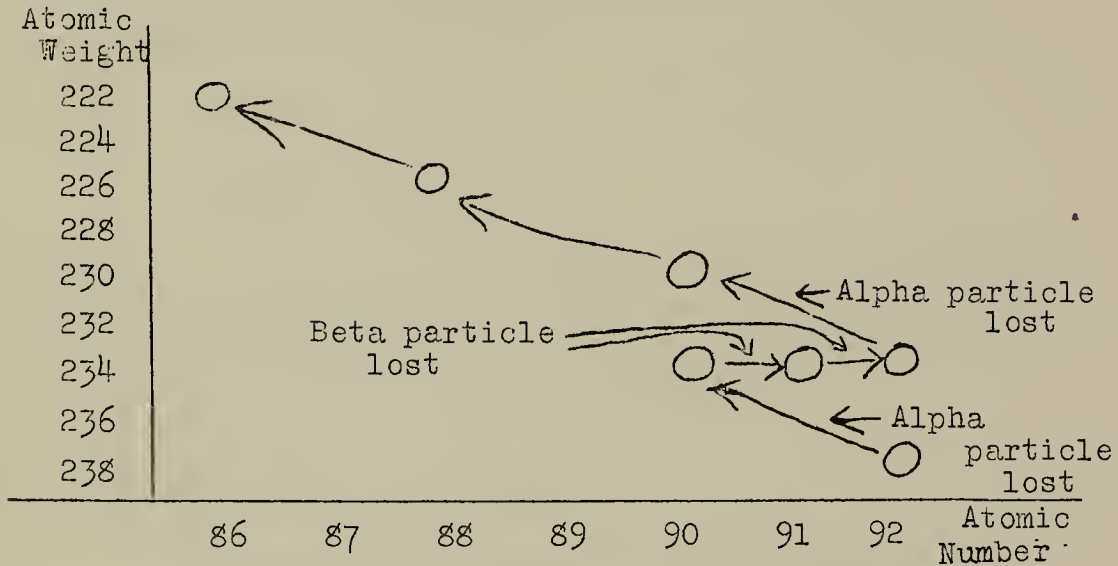


Do Not Stop. Go to next sheet without further instructions.

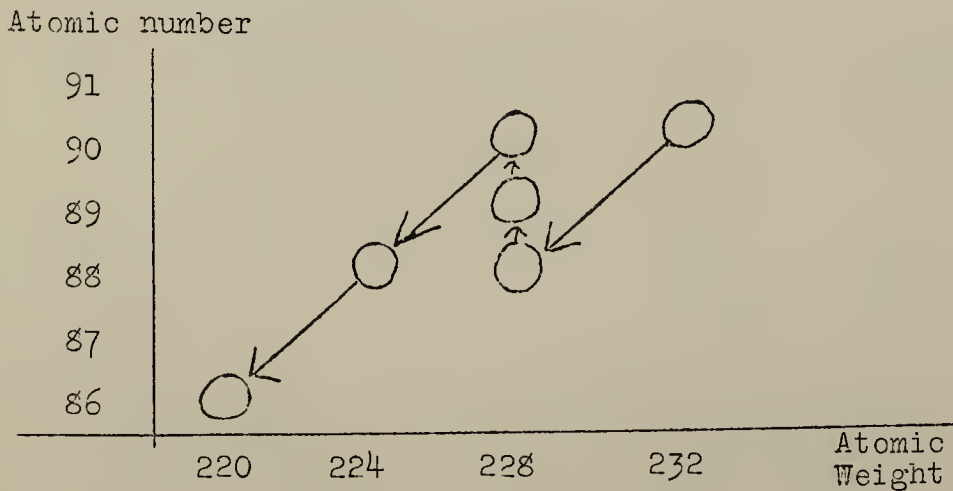
SECTION IV(Continued)

Part D

Facts and Principles. The following figure shows the disintegration series of radium. This figure shows that when an element loses alpha particles there is a decrease in the atomic weight of 4 and a decrease in the atomic number of 2. The figure also shows that when beta particles are lost there is no change in the atomic weight but an increase of 1 in atomic number.



Problem. The circles in the following figure represent certain elements. You are to place the letter A in the circles which have lost alpha particles, and the letter B in the circles which have lost beta particles.



Stop here. Wait for further instructions

SECTION V

Directions: Below are some facts and principles followed by some problems. Study the facts, principles and examples and solve as many problems as you can in the time allowed. You may figure on the edge of this page.

Facts and Principles. 1. The process of finding a function, given its derivative, is called integration.

2. Just as the sign \div in the ration $-a \div b$ signifies division, the sign \int indicates integration.

3. The following equation is a general integration formula:

$$\int u^n du = \frac{u^{n+1}}{n+1} + C.$$

That is to say, the integral (\int) of u to the n th power (u^n) times the derivative of u (du) equals ($=$) u raised to the $n+1$ power (u^{n+1}) divided by $n+1$ ($\frac{u^{n+1}}{n+1}$). The constant of integration (C) is added to the function.

Sample problems.

Evaluate the following integrals:

a. $\int u^2 du =$

1. Applying the general integration formula (stated above), we find that

2. $\int u^2 du = \frac{u^{2+1}}{2+1} + C = \frac{u^3}{3} + C = \text{Ans.}$

b. $\int (u^2 + 3u^4) du =$

1. Simplifying,
 $\int (u^2 + 3u^4) du = \int u^2 du + 3 \int u^4 du =$

2. Applying the general integration formula, we find that

$$\int u^2 du + 3 \int u^4 du = \frac{u^3}{3} + \frac{3u^5}{5} + C = \text{Ans.}$$

Problems: Evaluate.

a. $\int x^2 dx =$

b. $\int x^3 dx =$

c. $\int y^7 dy =$

d. $\int (u^4 + u^7) du =$

e. $\int u^2 du + \int x^3 dx =$

f. $\int dy =$
 Hint: $1 = y^0$

g. $\int 3 dx =$

h. $\int (3x^6 + 2x^3) dx =$

i. $\int (3 + 2y^5) dy =$

j. $\int (x^2 + 2)^2 dx =$

Stop here. Wait for further instructions.

SECTION VI

Directions: Answer the questions at the bottom of the page according to the ideas directly expressed or implied in the following selection. If a statement is true according to the selection, check (✓) true; if it is false, check false; but if the selection gives no idea whether it is true or false, check didn't say.

"The monad is the element of all being and partakes of its nature, and being is activity. If we ask further concerning the nature of the monad or activity, we can only be told that it is most like perception. Activity and consciousness are thus two words for the same thing and lie at the bottom of nature.

The monad is indestructible, uncreatable, and immutable, but it is not static. It undergoes a continuous process of development in accordance with its own laws, but loses in developmental change neither its identity nor its unity.

Immutable, uncreatable, indestructible monads can have no effect upon each other, for what could they do but change, create, or destroy each other? Thus it appears that the world is an infinite pluralism of independent monads. Thus, too, there are no causes. A cause would either imply mutual effects between monads (and there are none) or else an analysis of the monad so that its internal development is a causal chain (but the monad is unitary and has no parts). Cause as anything more than coincidence is sheer illusion. A monad is like a watch, perfectly constructed, wound up, and set going forever. It will continue without an external agent according to the laws of its own nature. Two such watches will be found always to agree, and yet neither is the cause of the other. Thus harmony in nature comes about without effective causation because harmony preëxists in the laws of the monads."

True False Didn't
Say

1. The monad develops and acts according to the laws of the universe.			
2. The monad changes from one stage to another until it loses its identity.			
3. The world developed from one original monad.			
4. There are at least two atoms in one monad.			
5. Both mind and matter develop from the monad.			
6. The monad was created by God.			
7. Each monad is held in position by other monads.			
8. The energy which causes the monad to move comes from the sun.			
9. Knowing and doing are the same thing.			
10. The principles of cause and effect in reality do not exist.			
11. Harmony in nature is due to the inter-relationships of the monads.			
12. The monad does not change but it moves.			
13. The monad is composed of an infinite number of parts.			

Stop here

REFERENCES

1. Bolton, T.L., "The Growth Of Memory In School Children"; American Journal of Psychology, (14), 1891.
2. Boring, E.G., A History of Experimental Psychology, The Century Co., 1929.
3. Boring, E.G., "Intelligence As The Tests Test It"; New Republic, June 6, 1932.
4. Boring, E. G., "The Logic of The Normal Law of Error In Mental Measurement"; American Journal of Psychology, (31), 1920.
5. Boring, E.G., "Scientific Induction and Statistics"; American Journal of Psychology, (37), 1926.
6. Constance, C.L., "Instructors' Use Of Mental Test Scores," Journal Of Educational Research, January, 1932.
7. Crane, E., "An Investigation of Three Plans for Selecting The Students to be Admitted to College," Journal of Educational Psychology, (17), 1926.
8. Cleeton, G.U., "The Predictive Value of Certain Measures of Ability in College Freshmen," Journal of Educational Research. Volume 15: pp 357-360. 1927.
9. Dearborn, W.F., Intelligence Tests, Houghton Mifflin Co., 1828.
10. Ellis, R. S., The Psychology Of Individual Differences, J.Appleton and Co., 1930.
11. Edgerton, H. A., Academic Prognosis in the University. Warwick and York, 1930.
12. Edgerton, H. A., and Toops, H., "Academic Progress," Ohio State University Studies. Contribution in Administration. No. 1. 1929.
13. Freeman, F.N., Mental Tests, Houghton Mifflin Co., 1926.
14. Glick, H.N., "Effects of Practice", University of Illinois Bulletin, (31), 1925.
15. Glick, H.N., "The 1929 Psychological Examination", Educational Record, April, 1930.

16. Harper, F.H., Elements of Practical Statistics, MacMillan Co., 1930.
17. Held, O.C., and Mason, George, Critical Levels on the Thorndike Psychological Examination. Unpublished Study. University of Pittsburgh College Personnel Office. 1928.
18. Hull, C.L., "The Correlation Coefficient And Its Prognostic Significance", Journal of Educational Research, (15), May, 1927. ✓
19. Kohler, W., Gestalt Psychology, Horace Liveright, 1929.
20. Love, C.E., The Calculus, MacMillan Co., 1926.
21. MacPhail, A.H., The Intelligence of College Students, Warwick and York, Baltimore, 1927. ✓
22. Mills, F.C., Statistical Methods, Holt and Co., 1924.
23. Monroe, W.S., The Theory of Educational Measurement, Houghton Mifflin Co., 1923.
24. Odell, C.W., Educational Statistics, The Century Co., 1925.
25. Odell, C.W., "Predicting the Scholastic Success of College Freshmen", University of Illinois Bulletin, (37), 1927. ✓
26. Otis, A.S., Statistical Method In Educational Movement, World Book Co., 1926. ✓
27. Pintner, R., Intelligence Testing, Holt and Co., 1931.
28. Rugg, H.D., Statistical Methods Applied to Education, Houghton Mifflin Co., 1917.
29. Terman, L.M., The Measurement of Intelligence, Houghton Mifflin Co., 1916.
30. Toops, H.A., "The Prediction of Scholastic Success in College". ✓
School and Society. Volume 25; pp 265-268.
31. Trow, W.C., Educational Psychology, Houghton Mifflin Co., 1931.
32. Wheeler, R.H., The Science of Psychology, Crowell Co., New York, 1929.
33. Woodworth, R.S., Psychology, Holt and Co., 1931.

ACKNOWLEDGMENTS

The writer is indebted to Dr. W. Davis for permission to reproduce his drawing of the agaric, to Dr. P. Serex for permission to use the problems contained in Section IV of the appended test, and to The Century Company for permission to use certain published material.

Especial thanks are due Professor H. N. Glick, co-author of the test, under whose direction the work developed: whatever the present study merits in care, thoroughness, and perspective derives originally from him.

Approved by:

H. N. Glick

Walter E. Prince

F. C. Moore

Graduate Committee

Date June 13, 1932



