A COMPARATIVE STUDY OF TWO EXAMINATION SYSTEMS USED IN GRADE IX

A Thesis

Presented to

the Faculty of Graduate Studies and Research

University of Manitoba

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

---

by

Abram P. Hildebrand

August 1967

ACKNOWLEDGMENT

The writer expresses his sincere gratitude to Professor A. M. McPherson for his invaluable assistance in the writing of this thesis. The writer expresses his appreciation especially for the frequent words of encouragement.

Special words of gratitude go out to Professors J. W. Peach and C. C. Wood. Their helpful advice and constructive criticism made it possible to crystalize the format of this thesis.

A note of recognition is also due to all the friends and colleagues who assisted by showing interest, by offering criticism, and by cross checking procedures.

Finally a word of appreciation goes forth to all the faculty members who have assisted in the development of this research.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

TABLE PAGE

TABLE PAGE

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

Programs of education are undergoing drastic changes in every province of Canada. More up-to-date courses are made available in the academic field as well as in the vocational field. New advances in technology make a better education a pre-requisite to obtaining a worth-while job. The education need not include university training but it should enable the student to do well at the job that he enjoys.

One of the reasons for the rise in the Manitoba high school enrolment is a demand for better educated employees. To ensure that the young people have a higher education upon leaving school, the government of Manitoba has raised the high school leaving age to sixteen. It is the school's responsibility to institute a testing program that will help the student select an appropriate program of study. Since earlier education was almost exclusively academic, a realistic approach must be taken to ascertain whether the pursuit of an academic training is in the best interest of the student.

Different testing systems could be used to test the performance of a student. At present the two major systems in use are the term testing system and the continual

testing system. Research confirms that very few attempts have been made to predict students' grades in individual courses at the secondary school level. [1]More information on what testing systems are able to do for the educator must be made available through research.

Statement of the Problem

It was the purpose of this study to compare two testing systems used in Grade IX; the term testing system and the continual testing system. The two systems were compared (1) as predictors of June marks; (2) as methods of motivation; and (3) on the basis of their acceptability to students, parents, teachers and inspectors as revealed through a questionnaire study.

Significance of the Problem

The significant changes in Manitoba's program of education have made it necessary that students be assisted in choosing a proper course of study. Until recently, students in rural Manitoba did not have the same opportunities as students of large urban centres. Only in a few exceptions were rural students in circumstances in which

---

[1]Marian M. Schusler, "Prediction of Grades by Computer for High School Students: A Cross Validation and Experimental Placement Study" (unpublished Doctoral Dissertation University of Pittsburgh, Pennsylvania, 1964)

they had to make a choice between two or more courses. The formation of school divisions and consequent centralization of student population has made it possible to offer a greater choice of courses to the students in rural Manitoba. Now that these opportunities exist, students must be assisted in making the proper choice. If such assistance is to be made available, research must be carried out to present necessary information.

Formerly, research was carried on primarily in Grade XII in order to predict success in an institute of higher learning. Consequently, little is known about achievement indicators within the high school system. Grade IX is the crucial year. In this grade the students must select a course which they intend to study in Grade X, the first grade to offer alternate courses. A study such as this has immediate value. It provides information which students may use to choose the appropriate Grade X course. The Guidance Director would be in a much better position to make pertinent suggestions if early indications should show that the pursuit of an academic course was not in the best interest of the student. This would assist the student to make the necessary mental adjustments for a rewarding effort in a course suited to his needs.

A student also requires assistance within the academic field. A choice between academic courses would also arise when indications pointed out that a student was courting disaster by continuing the University Entrance Course. A mental adjustment is necessary, particularly due to the university orientated approach that has been so prevalent until recently in rural Manitoba. Such an adjustment should be made while the student is still scoring passing marks. In this way a student will be spared the stigma of failure, if he chooses to examine suggestions carefully and select a course in which indicators point to success.

Students are in school to learn; to grow;--achievement tests should measure the results of such growth. Much time and effort is spent in writing tests and examinations, marking them and reporting to the parents. All this testing procedure seems somewhat absurd if such results show no significant relationship to the June departmental marks. If either of the examination systems were found to be a good predictor of June departmental results, it could be possible to eliminate the June departmental examinations. This would be in keeping with the trend towards making schools responsible for promotions, as is the case in Grade X. Such recommendation would follow research on a much wider scale than is possible in a locale study of the kind attempted in this thesis.

Student motivation is important to teachers. If a teacher had a choice between two methods of motivation, he should know which method is the better before making a selection. A comparison of the two testing systems should provide an answer to this problem of selection. This study has immediate value by providing such information.

Definitions, Assumptions, Limitations

Definitions

1. Term testing is taken to denote a system of examinations where students are tested after a given period of time, usually a matter of three months. Such a period of time is called a term.

2. Continual testing is taken to denote a system of testing where no such fixed periods of time exist, but where tests are given at the discretion of the instructor. A minimum number of tests per month is one; the maximum number is left to the discretion of the instructor. A minimum number per month is required in order to report progress to the students and to the parents. Every assignment that is marked by the teacher or marked in class under the guidance of the teacher is recorded and used to arrive at the mark for the month. Testing may also consist of objective tests, review tests, timed tests, chapter tests or any other manner of checking devised by the instructor.

3. June examinations refer to the departmental examinations which the Grade IX students write in June. These examinations are set by a committee appointed by the Department of Education. They are written under strict departmental regulations and are marked at a central point by a committee of markers appointed by the Department of Education. Some of the subjects are also examined in such a manner that the responses that the students make may be machine scored. This is usually done in subjects such as science and mathematics.

4. First term average refers to the average mark scored at the end of the first term by term marked students. It also refers to the average mark of the first three months scored by a student under continual testing.

5. Average Mark of the Year is the average of the three term marks scored by term tested students. It is also the average mark of the ten monthly marks scored by the students under continual testing.

6. Core subjects refer to the four subjects in Grade IX in which the department of education sets final examinations. They are language, social studies, mathematics and science.

7. Control group refers to the group subjected to term testing.

8. Experimental group refers to the group subjected to continual testing.

Assumptions

1. Since Grade IX departmental finals were the deciding factor for promotion from Grade IX to Grade X in June of 1963 to 1966 inclusive, they were accepted as having validity even though no absolute validity. The case for using departmental final marks in any combination of years rested on this point. The June departmental examinations of the four years, 1963 to 1966, were considered as forms A, B, C and D of the same test for Sample A students. For Sample B students, the June departmental examinations of 1967 were considered on par, or form E, with the June departmental examinations of 1963 to 1966 as a standard on which to base comparisons.

2. The population from which the samples were taken was assumed to be similar to and representative of other small rural centres of student enrolment of between 300 and 450 students.

3. The situation that existed was also assumed to be similar to other rural centres in regard to number of teachers involved over a four year period, teacher difference, approach and results of instruction.

Limitations

Since the population from which the samples were taken was quite small as attested to by the enrolment, 300 to 450 students from Grades IX to XII, this thesis can

be accepted only in situations of equal or similar enrolment. It must also be borne in mind that this study was carried out in a rural school division of Manitoba. No general inferences can be drawn from the processed data without reference to conditions and situations in centres similar to the one under study.

Another limitation is the extent to which two persons can be matched by I.Q. ratings. Since the sampling was small, there was bound to be some difference between the four groups of Grade IX students from 1963 to 1966. Even if the students could be matched for I.Q. there would still be other factors which would impinge upon the individual student to make him different. An example of this would be the ease in which an individual student could be motivated.

Procedure

Study in General

A review of some literature on prediction of success or failure in the field of education is given in Chapter II. Much use was made of other theses covering this field. The review also deals, in part, with the values and the limitations of testing. Considerable attention was paid to what other countries are doing in testing.

Chapter III deals with the materials used and groups studied. It describes the sampling and the criteria used

for pairing. This chapter also describes in detail all statistical methods used.

The actual testing program and the results of Sample A are presented, analyzed and summarized in Chapter IV. The testing program was run in a rural centre in Manitoba. The high school student enrolment for that centre was between 300 and 450 pupils. This study involved the results of the Grade IX students only. The Otis Quick Scoring Intelligence tests were used during the years under consideration. Samples of all the Intelligence tests used may be found in the appendix. Traditional term tests were used from September, 1962, to June, 1964, while a continual testing program was carried on from September, 1964 to June, 1966. During the school year of September, 1966, to June, 1967, the term testing system was used to test a class of Grade IX students while the continual testing system was used to test their counterparts in another Grade IX class.

The average marks for the year in the core subjects were correlated with the June marks. This was done for both testing systems. A similar comparison was made of the first term results and the June results. This also was done for both testing systems. T-scores were then calculated in all cases to establish whether there was a significant difference between the averages scored in term testing and those scored in June. The same procedure was

carried through for continual testing. The correlation and t-score procedure was used on the averages scored by the group, by the boys and by the girls. These relationships were utilized in the development of regression equations. Final conclusions and recommendations were drawn from this phase of the study.

The second phase of the study attempted to compare the two testing systems as methods of student motivation. The mean June departmental mark obtained by the term tested students was compared to the mean scored by the continually tested students. The marks were subjected to correlation coefficients and t-tests to check for a significant difference between the means.

This comparison was again made for the four core subjects in Grade IX in which June departmental examinations were set. Result comparisons were made between the total number in each system as well as between the boys and the girls of one system with their counterparts in the other system. The results were presented in tabular form and conclusions were drawn.

Chapter V used Sample B to duplicate the procedure used with Sample A in Chapter IV, while Chapter VI compares the results of the two samples. This procedure was felt necessary due to the difference of variables. Since all

the data was available for Sample A, it was felt that the results of such data should be cross-examined by a method in which all possible controls were exercised. Such controls should make it possible to claim that probably the results obtained were due to the testing system used.

The content of Chapter VII deals with phase three of the study. Phase three consisted of the questionnaire approach. A questionnaire was constructed to solicit responses to questions pertinent to the two testing systems. These questionnaires were forwarded to all the students, to the parents of the students, to the faculty and to the inspectoral staff. The results were compiled and analyzed and the information acquired was used in the final comparison of the two testing systems.

The results of this research are summarized in the final chapter. Some conclusions have been drawn. The final chapter also includes some suggestions as to applications of the findings as well as suggestions for further research in this field.

CHAPTER II

Review of Literature

Authorities agree that testing, in one form or another, is carried on by all institutions of instruction. Testing is so well established that few instructors would care to terminate a course of instruction without some form of grading. The public in general has been conditioned to such procedure and expects, even demands, some form of grading to indicate the degree of excellence attained.

Some authorities, such as J. W. M. Rothney[1], deplore that frequently evaluation is no more than a "series of exercises and tests, marking them, adding or averaging marks, and entering them on a small card which is to be taken home to be signed by a parent, usually the mother. This process is often quick, simple, and terminal."[2] Rothney goes beyond this point, however, and establishes major points in evaluation: that learning is a progressive process; that evaluation must be carried on at all times with the co-operation of the learner, and that such re-

---

[1] John W. M. Rothney, "Evaluating and Reporting Pupil Progress", Department of Classroom Teachers, American Educational Research Association of the National Educational Association, May 1960

[2] Ibid., p.3

[3] Ibid.

cords must be cumulative in order to reveal progress.[3] J. L. Brereton[4], L. J. Cronbach[5], and R. H. Bauernfeind[6] agree that the frequent writing of tests is beneficial to the student because immediate goals are established, the objectives of the course are emphasized and knowledge is reinforced. P. L. Dressel[7], Cronbach and J. C. Nunnally[8], however, all emphasize that maximum benefits are derived only when results are discussed shortly after writing the test. This approach, they claim, reinforces learning most effectively.

Cronbach[9], particularly, presents a comprehensive approach on what tests should do. His approach is positive. In his opinion, students should be able to glean

---

[4]Joseph L. Brereton, Examinations, Where Next?, Pacific Northwest Humanist Publications, 1965

[5]L. J. Cronbach, Educational Psychology, New York: Harcourt, Brace, and World Inc., 1963

[6]R. H. Bauernfeind, Building a School Testing Program, Boston: Houghton Mifflin Company, 1963

[7]L. P. Dressel, and Associates, Evaluation in Higher Education, Houghton Mifflin Co., Boston, 1961, p. 110

[8]J. C. Nunnally, Educational Measurement and Evaluation, McGraw-Hill Book Co., Toronto, 1964, p. 13

[9]Cronbach, op. cit., p. 539

information from tests. Tests should indicate what needs to be learned. At the same time, tests should be constructive and motivate work. They should assist the learner to realize how he should change or develop his performance. When appropriate change has taken place, tests should furnish satisfaction of achievement.

Most authorities on examinations will agree with Cronbach, that tests should furnish information on which decisions can be made about the learner. Information such as what course he should take, what remedial treatment is required, or what job or college should be recommended are frequently necessary. At the same time, tests should furnish the teacher with material by which he may judge the calibre of his instruction.

Marks have played and are playing an important part in an individual's life. J. W. M. Rothney states that, "Marks are the coin of the school realm. They continue to be the measure of school success--the keys that open doors of educational institutions for entrance and exit."[10] Since this is the case, instructors must be concerned at all times that the mark is a true rating of the student's grasp of the subject matter. Administrators cannot allow a single rating influence a decision. They must be willing

---

[10]Rothney, op. cit., p. 8

to obtain a cumulative record on which to base a decision. Employers, likewise, must be careful in weighing the pros and cons of a single reference sheet.

Tests and marks can be most useful but care must be taken that they do not become ends in themselves. Cronbach[11] ascribes such limitations to tests because teachers and students tend to regard them as the only important evidence of accomplishment. In this regard, marks become the end-all of learning but do not fulfill their true purpose--being indicators of level of achievement. Brereton[12] tends to agree with Cronbach. Brereton argues that if the results of the final examinations carry too much weight, pupils tend to cram. They are more concerned about the mark they receive than how well they understand the work.

Improper evaluation of marks indicates that examinations really control what is taught and how it is taught. If marks are all important, then marks we shall have! Students and teachers are interested in drilling for facts. Both stress memorization of facts with the hope that enough facts will be called for on the examination to permit a student to score a high mark. Both students and teachers become slaves to textbooks because

---

[11] Cronbach, op. cit., p. 539

[12] Brereton, op. cit., p. 434

in too many cases textbooks are prescribed for a course. These texts become the sole authority on which too many examinations are based.

In his book, Examinations, where next?, J. L. Brereton[13] studied examination procedures and methods of marking in nine different countries: England, Sweden, West Germany, Portugal, Senegal, Canada, United States, East Germany and Czechoslovakia. Some of his findings bear out claims of other writers and educators. Examination systems of these countries convinced Brereton of the value of well devised competitions in the promotion of any kind of training. Probably the main value consists in setting goals towards which to strive.

Brereton notes that there are few books on the theory of conventional examinations[14]. This is probably due to teachers seldom having an opportunity to study examinations systematically.

Brereton found examinations in Britain to be among the best in the world. He noted one weakness--that too much emphasis was placed on reliability. Brereton noted that the reliability of most objective examinations was

---

[13]Brereton, op. cit., p.3

[14]Brereton, op. cit., p.4

higher than that of essay type examinations. He recommended, however, that more attention should be given to essay type examinations because such examinations elicit the best teaching and that such results are desirable even at the expense of some accuracy of marking. Brereton recommended a balanced approach of written, oral and practical examinations.

Brereton also noted that where examination papers were marked centrally, results were scaled in order to retain the standard required for passing. Such conditions existed in England and in the provinces of British Columbia and Alberta in Canada. In other countries, Sweden, for example, the examiner simply rated a paper as passing or failing, leaving the degree of passing or failing to the individual instructor. The argument here was that an instructor capable of instructing should also be capable of rating. Outside rating merely assisted on a broad basis.

On the topic of rating, Brereton[15] reported that informed opinion in different parts of the world agreed on a letter scale in preference to a percentage rating. Such letter grades tended to have from four to six gradations and were used in class grading as well as in

---

[15]Brereton, op. cit., pp. 59 - 60

external examinations. An incident was, however, cited where parents and press were dissatisfied with a letter rating and demanded percentage rating instead.[16]

One of Brereton's statements is extremely telling, especially in view of the extensive study of examination systems and methods of evaluation. Brereton states:[17]

"In 1599 the Jesuits published a complete account of the examinations to be used in their schools. A study of this document shows that we have learned very little since then."

Frequently administrators wish to know how well a given rating will predict a future rating. Much research has been undertaken to determine predictive values of certain tests. Two such studies at the University of Manitoba are "High School Averages and Supplementals as Predictors of First Year University Success." by H. Pollock[18] and "The Differential Aptitude Tests as Predictors in Education I at the University of Manitoba" by David Friesen.[19]

---

[16] Ibid. [17] Brereton, op. cit., p. 71

[18] H. Pollock, "High School Averages and Supplementals as Predictors of First Year University Success." (unpublished Master's thesis, University of Manitoba, Winnipeg, 1959)

[19] D. Friesen, "The Differential Aptitude Tests as Predictors in Education I at the University of Manitoba." (unpublished Master's thesis, University of Manitoba, Winnipeg, 1958)

Friesen correlated the various sections of the Differential Aptitude Tests with marks achieved by the same students in Education I at the University of Manitoba. It was found that a positive relationship existed between class marks of the students in Education I and their marks on the Verbal Reasoning, Abstract Reasoning and Sentence tests. Pollock attempted, in a similar work, to show the predictive value of marks obtained in June departmental examinations. His study, however, attempted to establish whether a real difference existed between students obtaining a clear pass and students passing on supplementals.

W. G. Flemming in "Factors Affecting the Predictive Accuracy of Ontario Grade XIII Results"[20] found that some of the significant factors affecting the degree of relationship between the Grade XIII and University averages were the type of school, the economic level of the community in which the school was located, the proportion of academic specialists on the staff and the current expenditure of the school board. It was also found that higher correlation between Grade XIII and University averages permitted a higher degree of precision than lower correlations when

---

[20] W. G. Flemming, "Factors Predicting Accuracy of Ontario Grade XIII Results." Bulletin No. 16, Department of Education Research, University of Toronto, Toronto, 1955, p. 3

limits within which a particular student's average will fall are predicted. It was noted that prediction for particular university subjects was not much affected by the school background.

In similar studies by I. R. Bou and F. L. Stovall[21] at the University of Puerto Rico, regression equations showed that students from large high schools achieved better grades than predicted while students from small high schools achieved lower grades than predicted. This study indicates that high school grades are rather unreliable for prediction of success at University. Francis F. Smith[22], at Fresno State College, however, found that the previous year's work was the best single indicator. This finding by Smith was corroborated by S. B. Schmitz[23] of St. Benedict's College, Atchison, Kansas. Schmitz found that the high school average was the most efficient single instrument for prediction of success at college.

---

[21] I. R. Bou and F. L. Stovall, "Relationship Between High School Grades and First Year Achievement in the University of Puerto Rico" Journal of Educational Psychology, May, 1950, Vol. 41, pp. 309 - 320

[22] F. F. Smith, "The Use of Previous Records in Estimating College Success", Journal of Educational Psychology, March, 1945, Vol. 36

[23] S. B. Schmitz, "Predicting Success in College: A Study of Various Criteria", Journal of Educational Psychology, September, 1937, Vol. 28, pp. 465 - 473

W. H. McIntyre[24] compared marks handed out by teachers with marks achieved in June departmental examinations. The study examined the degree by which teachers tend to over-rate or under-rate a pupil's achievement. McIntyre noted that a great variation of discrepancies existed On the average, however, he found that school marks assigned by teachers of accredited collegiates were a fair indication of achievement as measured by the external June departmental examinations.

In summary, it is accepted that examinations have value to students, instructors, administrators and employers. Examinations must be viewed as a means to an end, not an end in themselves. Various approaches are used in different countries but all agree that rating a student in some manner is necessary. Studies show that schools rate differently but that the school averages are still the most efficient single instrument to predict success at college.

This review is only a brief summary of some of the vast amount of literature available on these topics. Studies on prediction were of particular interest to this

---

[24]W. H. McIntyre, "A Comparative Study of Collegiate and Department of Education Marks in Manitoba" (unpublished Master's thesis, University of Manitoba, Winnipeg, 1952)

study. Former trends acted as guide lines for the format and procedure to be used. The values and limitations of testing were also of significance due to their invaluable assistance to teachers. Cronbach and Brereton were particularly helpful in this area.

CHAPTER III

MATERIAL USED AND GROUPS STUDIED

Traditional term examinations had been used for years in the school in which the experiment was carried out. Because of factors such as the extension of the compulsory school age, greater stress on scholastic application and the tremendous impact made by automation, the old system was no longer achieving its purpose. Before the heavy emphasis to stay in school started, students that stayed in high school were generally those with a desire to achieve. Students were in school to study and the results show that they did reasonably well. With the greater emphasis on higher education in order to secure employment after graduation, the young people were indoctrinated by teachers, employers and parents to stay in school; to graduate.

This emphasis brought more students into the class rooms. These students would like to benefit from the results of their attendance but they either had no interest in the courses offered or were unwilling or unable to apply themselves continuously to a task, the results of which lay far in the future.

The immediate task lay in stimulating learning throughout the school year instead of relying on the old system of "coasting and cramming". The faculty thought

that most of the students would respond favorably to immediate reward. Therefore, the faculty decided to make each assignment as meaningful as possible. To be meaningful, each assignment should carry some credit towards the student's monthly grade. These monthly grades should, in turn, constitute the final criteria for failure or promotion.

The faculty next discussed the problem of running a continual testing program. It was decided that chapter tests and unit tests were desirable testing units and that such tests should be used. It was also decided that for factual information, objective tests would be quite suitable. In order to help the student organize and present material, it was decided that a student should also have ample opportunity to write essays. These would be marked by the teachers, using uniform symbols to indicate errors, and using as uniform a marking system as possible.

It was reasoned that if the day-by-day class room assignments were recognized as having worth, the student might take more care and pride in executing these tasks. After a continued period of meaningful application, it was reasoned, the student would develop a habit and a liking for continual study and application and consequently discard the method of "coasting and cramming".

Sample A

To compare the results of a change from the old traditional method to the new continual testing system required two groups of students. Fifteen boys and 20 girls were selected from the Grade IX students in 1962 to 1963 and 15 boys and 20 girls were selected from the Grade IX students in 1963 - 1964. These 70 students were selected as the control group and were subjected to the familiar term examinations.

These 70 students were matched as nearly as possible in terms of age, sex and intelligence to 15 boys and 20 girls in Grade IX during 1964 to 1965 and 15 boys and 20 girls in Grade IX during 1965 to 1966. This latter group of 70 students was subjected to the new approach--continual testing. It was the experimental group.

All students were chosen from the same school in order to minimize changes and methods of approach. In this manner it was possible to offer the same course, the same methods and the same standards of testing for any particular year. Since the students and teaching staff changed over the four years, changes did come. These changes were kept to a minimum by introducing procedures and approaches to the incoming staff and students. Selection from the same school, however, limited the number of students that could participate in the study due to enrolment, suitable

age and intelligence quotient.

Since the purpose of the study was to compare methods of testing, variables such as different groups of students taught by different teachers had to be considered as common factors which would be found province-wide in school situations such as used for this study. Inferences could be made as indicative of trends in situations under consideration.

Sample B

Since the argument for the previous sampling was questioned, a second, more comprehensive approach was suggested. This second approach ran parallel to the previous sampling. It attempted to aid in finding solutions to the same problems. At the same time the two approaches made it possible to compare the two sampling methods used.

The second method of sampling allowed much greater control over the variables. It permitted isolation of the objective--the testing system itself.

Upon opening of school in the fall term of 1966, I.Q. tests were administered to all the incoming Grade IX students. There were 160 in number. Out of the total Grade IX population of this school 27 students were paired with a second group of 27 students. Each group consisted of 14 boys and 13 girls. The pairing was identical to the method used in the previous sampling; age, sex and I.Q.

In all pairing, attention was also paid to socio-economic factors and ethnic background. Because such factors are intangible, it is hard to estimate what bearing this safeguard had on the final results.

The second sampling gave more control over teaching methods, testing methods, general efficiency of teacher, general atmosphere and approach. Teacher A taught mathematics to both groups. Group I, the control group, was tested by term testing while group II, the experimental group, was tested by continual testing. A similar approach was used in all of the core subjects in Grade IX.

Criteria for pairing

Intelligence tests were administered to assist in the matching of the pairs of students. The results are contained in the appendix. The Otis Gamma test was used for the matching of pairs in 1966 to 1967. Combinations of Otis beta tests Forms A and DM were used from 1962 to 1965. Students were considered to be a pair if their I.Q. rating did not differ by more than three points. Members of the same pair had to be of the same sex. Their age had to differ by not more than three months. These statistical details of each pair may be found in Tables XLIX - Ll, pp. 158 - 160, of Appendix A.

Detailed outline of statistical approaches used.

As stated in the problem, the purpose of this study

was to compare the two testing systems in three specific ways. The first comparison was to be made in predictive values. This part of the study, therefore, concerned itself with the different sections to be compared as predictors of June results.

First, the results of the entire control group were compared with the results of the entire experimental group. The total groups were than separated into male students and female students. The results achieved by the male students and the female students of the control group were then separately compared with their counterparts in the experimental group. The multiple comparisons of the groups were made for the first term results as well as for the average mark for the year. The complete procedure of comparison was carried out for all four core subjects.

The raw scores of a particular subject were tabulated as soon as they were available. They were then statistically analyzed in terms of standard deviation, product-moment correlation coefficient, and t-score according to the procedure described by L. E. Tyler[1]. The product-moment correlation coefficient (r) is the mathematical statement of the relationship between two sets of

---

[1] E. L. Tyler, Tests and Measurements, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1965, pp. 17 - 24

scores. Its formula is:

$$r = \frac{\Sigma (d_x)(d_y)}{N(SD_x)(SD_y)} \qquad 2$$

(In this formula d stands for deviation from the mean, N stands for the number of cases, SD_x stands for the standard deviation of the first set of scores, and SD_y stands for the standard deviation of the second set of scores.) The relationship may be any decimal fraction. If it happens to be 1.00, the relationship is perfect. Merely arriving at "r" is not sufficient. It is still necessary to prove whether this relationship is statistically significant for that set of scores. The t-statistic supplies this information. The t-score can be computed by the method outlined by L. E. Tyler.[3] The level at which the relationship is of statistical significance can be read from Tables of "t" as prepared by Fisher and Yates.[4] For statistical purposes, relationships that are significant only at a level higher than 5% are usually discarded. It is an indication that such a situation would occur more than five times out of a hundred by random chance.

---

[2] Ibid., p. 19

[3] Ibid., p. 23

[4] Fisher and Yates, Statistical Tables for Biological, Agricultural, and Medical Research, Oliver and Boyd, Ltd., Edinburgh

Another way of comparing the results is the theory of the "null hypothesis".[5] The null hypothesis for this study assumes that there is no significant difference between the average mark scored by the students in the first term and the average mark scored by the same students in the June examinations. Any variation that exists is assumed to be due to chance fluctuation. If the level of statistical significance is low (5% or lower) the null hypothesis can be rejected and it can be stated that there is a real difference between the first term and the June marks. It also states that this relationship would not happen more than five times out of a hundred by random chance for this group of students in this subject.

Once the null hypothesis can be rejected for a given case, the correlation coefficient can be used to formulate a prediction formula or regression equation. This can be done by using the following equation:[6]

Prediction for X from Y.

$$X' = r. \frac{S_x(Y - M_x)}{S_y} + M_y$$

Predicting Y from X. $$Y' = r. \frac{S_y(X - M_x)}{S_x} + M_y$$

After the regression equation has been developed, it

---

[5] J. W. Best, Research in Education, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1965, pp. 226 - 227

[6] Ibid., p. 233

becomes important to know the index of forecasting efficiency. "This index indicates the percentage improvement in predictive ability of a coefficient of correlation over a pure chance guess."[7] The predictive index was the goal for this phase of the study.

June marks could be predicted after the regression equation had been calculated. Since the predictive value is rarely accurate, it was necessary to estimate the amount of error. The standard error of deviation is found by the formula:

$$\text{Standard error of deviation} = S_y \sqrt{1-r^2}$$

$$= S_x \sqrt{1-r^2}$$

where S is the standard deviation of Y scores and S is the standard deviation of X scores and r is the correlation coefficient.[8]

Tests can also be used as methods of student motivation. The second comparison was run in an attempt to establish which testing system is the better method of assisting students to score higher June marks. This phase of the study was concerned with the June results only. The

---

[7] Ibid., p. 241

[8] E. F. Linquist, A First Course in Statistics, Houghton, Mifflin Co., Boston, 1942

June results scored by the term tested groups of 1963 and 1964 were compared with the June results scored by the continually tested groups of 1965 and 1966.

The correlation coefficients and t-statistics were again derived as in the study of the predictors. These comparisons gave the relationships of the results. They also provided the level of significance as before. This time the null hypothesis, however, stated that there was no real difference between the average marks scored by the term tested groups and the average marks scored by the continually tested groups. It assumed that any difference that was noted was due to random chance and not due to the use of different testing systems. The null hypothesis could be rejected if the t-score were significant at the 5% level. It could then be stated that there possibly was a difference between the two June average marks and that this difference could possibly be due to the two testing systems used.

The comparison was considered complete after the possibility of a difference had been made. If such a difference did exist, the system producing the higher June results was considered the better method of student motivation.

The questionnaire was a direct result of student and parent interest in the testing program. The main factors

under discussion usually included time, attitude towards assignments, attitude towards examinations and final results. Parents, particularly, were concerned about the psychological effects that examinations had on students. This questionnaire was designed in an effort to find some consensus. It was felt that it would constitute an excellent sub-division in a study which dealt primarily with statistical facts.

CHAPTER IV

EXPERIMENTAL STUDY--SAMPLE A

Procedure

The control group wrote traditional term tests from September, 1962 to June, 1964 to establish the grade achievement of the student after covering a certain amount of material programmed for the first term. Special time was set aside for this testing procedure. No classes were in operation and no new work was being covered. Generally, review sessions preceeded the testing sessions in order to prepare students for the tests. Normally classes were slow in picking up after the testing period because the teaching staff was busy marking papers and tabulating results.

These tabulated results were computed, analyzed and statistically compared with the final marks obtained by the same students in the June departmental examinations. Tables LII - LXXV, pp. 161-184, Appendix A contains the raw scores scored by the students at the end of the first term, their average mark for the year as well as the actual mark scored in June. The table shows a breakdown for boys and girls as well as for the four subjects studied.

A similar procedure was also undertaken with the experimental group from September, 1964 to June, 1966. The continual testing system was used during this time. This group kept up normal, routine work but was tested when and

how the instructor decided. The daily assignments and exercises were subject to scrutiny and rating towards a month-end grade. The average results of the first three months of the fall term were considered a fair comparison with the first term mark of the term testing system. This average mark, as well as the average mark for the year, was statistically compared to the actual mark scored in the June departmental examinations. This procedure was followed for the four core subjects. The comparison was also carried through for the female student group and the male student group. The raw scores are presented in Tables LII - LXXV, pp. 161 - 184, of Appendix A.

Statistical Analysis

The statistical analysis was carried out as carefully as possible. When peculiarities did arise, the procedure, data and calculations were re-examined in an attempt to spot the error. When an error was spotted, all similar possibilities were rechecked for similar errors. The Olivetti-Underwood calculator was used to assist in the calculation of the means, the standard deviations and the coefficient correlations as well as the t-scores.

After all the raw scores were in, the correlation coefficients were calculated. The t-score was then applied in order to check whether a significant difference existed between the various averages under consideration. The null

hypothesis, that there was no real difference in the average marks, was rejected at the 5% level and accepted at any percentage higher than that. The regression equations were worked out in order to predict June marks on the first term score as well as on the average score for the year. This was repeated for each of the core subjects and for the male and female groups. Indices of forecasting efficiency and estimates of error were computed. Finally, a comparison was made between the two testing systems based on the statistical findings.

The Results As Predictor

Analysis of Tables I and II

In language, the control group shows a significant difference between the first term results and the June results. Therefore this relationship has predictive value and may be used for the male group and the female group. The percentage of forecasting efficiency is low. It will predict the June mark 29% better than a straight guess. The actual mark could vary 8 marks on either side of the predicted mark. Again this also holds true for the male group and female group.

Continual testing also manifests a significant difference between the first term averages and the averages scored on the June departmental examinations. Here the relationship is negative but is significant at the 1% level.

TABLE I

PREDICTION OF JUNE RESULTS FROM FIRST TERM LANGUAGE RESULTS IN TERM TESTING

Case	First Term Average Mark 1962 and 1963	June Averages for 1963 and 1964	r	t	sig. level	null hypothesis
Group	68	63	.70	4.8	1%	rejected
Boys	64	60	.71	2.35	5%	rejected
Girls	71	65	.70	4.0	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(.70)\ (1.03)\ (Y-68)+63$	29%	8
Boys	$X'=(.71)\ (1.10)\ (Y-64)+60$	29%	8
Girls	$X'=(.70)\ (1.0\ )\ (Y-71)+65$	29%	8

Where X' is the predicted June mark and Y is the first term mark.

TABLE II

PREDICTION OF JUNE RESULTS FROM FIRST TERM LANGUAGE RESULTS IN CONTINUAL TESTING

Case	Average First Term	Average June	r	t	sig. level	null hypothesis
Group	72	58	-.67	13	1%	rejected
Boys	69	54	-.57	8.9	1%	rejected
Girls	75	61	-.62	9.6	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(-.67)\ (1.14)\ (Y-72)+58$	26%	7
Boys	$X'=(-.57)\ (1.24)\ (Y-69)+54$	18%	7
Girls	$X'=(-.62)\ (1.06)\ (Y-75)+61$	22%	8

Where X' is the predicted June mark and Y is the first term agerage.

Consequently the correlations have a predictive value. The index of forecasting efficiency is even lower than that of term testing.

From the observations based on the Tables I and II, the first term results of term testing predict the June results slightly better than the first term results of continual testing. This is the case for the entire group, for the boys and for the girls. The predictive value for term testing is, however, low and the estimate of error high.

Analysis of Tables III and IV

Except for the female group, the findings of Tables I and II pages 37 and 38 were repeated when the average of the year was compared with the average mark scored in the June Departmental Language examinations. The continual testing proved a better predictor than term testing in the case of the female group. The other marked difference was the generally higher efficiency percentage of prediction and a lower estimate of error. Table I, page 37, shows a prediction of 29% for all three groups in the first term results. In comparison Table III, page 40, shows predictions of 51%, 46% and 37% for the entire male and female groups respectively. The estimate of error is two points lower for the entire group, and one point lower for each of the male and female groups.

TABLE III

PREDICTION OF JUNE RESULTS FROM THE YEAR'S LANGUAGE RESULTS IN TERM TESTING

Case	Average Year 1962 - 1964	Marks June 1963 - 1964	r	t	sig. level	null hypothesis
Group	66	63	.87	3.26	1%	rejected
Boys	63	60	.84	2.24	5%	rejected
Girls	68	65	.78	2.38	5%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	X'=(.87) (.95) (Y-66)+63	51%	6
Boys	X'=(.84) (.98) (Y-63)+60	46%	7
Girls	X'=(.78) (.93) (Y-68)+65	37%	7

Where X' is the predicted June mark and Y the average mark for the year.

TABLE IV

PREDICTION OF JUNE RESULTS FROM THE YEAR'S LANGUAGE RESULTS IN CONTINUAL TESTING

Case	Average Year 1964 1966	Marks June 1965 1966	r	t	sig. level	null hypothesis
Group	70	58	.80	14.29	1%	rejected
Boys	67	54	.73	9.09	1%	rejected
Girls	73	61	.81	11.02	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(.80)\ (1.15)\ (Y-70)+58$	40%	6
Boys	$X'=(.73)\ (1.30)\ (Y-70)+54$	32%	6
Girls	$X'=(.81)\ (1.08)\ (Y-73)+61$	41%	4

Where X' is the predicted June mark and Y is the average mark for the year.

TABLE V

PREDICTION OF JUNE RESULTS FROM FIRST TERM SOCIAL STUDIES RESULTS IN TERM TESTING

Case	Average First Term 1962 1963	Mark June 1963 1964	r	t	sig. level	null hypothesis
Group	61	59	.78	1.43	20%	accepted
Boys	60	59	.83	0.0	90%	accepted
Girls	62	59	.75	1.05	30%	accepted

None of the cases have a significant difference at the 5% level. Therefore they have insignificant predictive value for the purpose of this study.

TABLE VI

PREDICTION OF JUNE RESULTS FROM FIRST TERM SOCIAL STUDIES RESULTS IN CONTINUAL TESTING

Case	Average First Term 1964 1965	Mark June 1965 1966	r	t	sig. level	null hypothesis
Group	55	53	.77	1.60	10%	accepted
Boys	51	52	.68	0	90%	accepted
Girls	57	54	.80	1.18	30%	accepted

Again the relationships are not significant at the 5% level. Consequently the data does not have significant predictive value for the purpose of this study.

Analysis of Tables V and VI

In Social Studies the first term average mark in term testing does not show a significant difference to the average mark scored in June. This is the case for the entire group as well as for the male and for the female groups. Therefore the results of the first term marks in Social Studies are of insignificant predictive value for June results. This is also true for the first term averages of continual testing.

Analysis of Tables VII and VIII

The conclusions reached on the first term results of term testing, Table V page 42, were also true of the results for the year. Social Studies showed no significant difference between the average mark for the year and the average mark scored in June departmental examinations. Consequently, the correlation coefficient between these marks could not be used to arrive at regression equations.

The only difference found in continual testing is in the entire group and in the female group. The level of significance is at the 1% for both of these groups. Regression equations were worked out, and the forecasting efficiencies are 31% and 34% respectively. The estimate of error is high; 10 marks either way.

Since this study was a comparison study, continual testing is the better predictor for the entire groups and

TABLE VII

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SOCIAL STUDIES RESULTS IN TERM TESTING

Case	Average Year 1962 1964	Marks June 1963 1964	r	t	sig. level	null hypothesis
Group	61	59	.92	1.04	40%	accepted
Boys	60	59	.93	0	90%	accepted
Girls	61	59	.91	1.60	20%	accepted

These findings are not statistically significant, consequently they are of insignificant value as predictors for this study.

TABLE VIII

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SOCIAL STUDIES RESULTS IN CONTINUAL TESTING

Case	Average Year 1964 1965	Marks June 1965 1966	r	t	sig. level	null hypothesis
Group	57	53	.72	3.08	1%	rejected
Boys	55	52	.65	1.36	20%	accepted
Girls	59	54	.75	3.18	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(.72)\ (1.17)\ (Y-57)+53$	31%	10
Boys	of insignificant predictive value		
Girls	$X'=(.75)\ (.88)\ (Y-59)+54$	34%	10

Where X' is the predicted June mark and Y is the average mark for the year.

for the female groups in Social Studies. No comparison can be made between the male groups.

Analysis of Tables IX and X

In Science, both term testing and continual testing show significant differences between the first term and June marks for the entire group and for the female group. Both testing systems show no significant difference between first term and June marks for the male group. The percentage of forecasting efficiency is higher for term testing than for continual testing. On the basis of these statistics, term testing is superior to continual testing for purposes of prediction for the entire group and for the female group. Neither system is significant as a predictor for the male group.

Analysis of Tables XI and XII

In science, both term testing and continual testing are good predictors of June marks if such prediction is based on the average mark for the year. The difference between the average mark for the year and the June mark is not significant for the male group in term testing. Consequently, the continual testing system is the better predictor for the male groups in this section. The percentage of prediction efficiency is 39% the estimated error is 6 marks either way. The term testing system

TABLE IX

PREDICTION OF JUNE RESULTS FROM FIRST TERM SCIENCE RESULTS IN TERM TESTING

Case	Average First Term 1962 1963	Mark June 1963 1964	r	t	sig. level	null hypothesis
Group	70	62	.74	5.97	1%	rejected
Boys	66	62	.75	1.86	10%	accepted
Girls	73	62	.77	6.96	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(.74)\ (1.08)\ (Y-70)+62$	33%	10
Boys	Insignificant relationship between scores		
Girls	$X'=(.77)\ (1.12)\ (Y-73)+62$	37%	9

TABLE X

PREDICTION OF JUNE RESULTS FROM FIRST TERM SCIENCE RESULTS IN CONTINUAL TESTING

Case	Average First Term 1964 1965	Mark June 1965 1966	r	t	sig. level	null hypothesis
Group	64	59	.60	2.61	2%	rejected
Boys	59	57	.56	1.37	20%	accepted
Girls	67	61	.61	2.80	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(.60)\ (1.23)\ (Y-64)+59$	20%	10
Boys	Insignificant relationship between scores		
Girls	$X'=(.61)\ (1.24)\ (Y-64)+61$	21%	10

TABLE XI

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SCIENCE RESULTS IN TERM TESTING

Case	Average Year 1962 1964	Mark June 1963 1964	r	t	sig. level	null hypothesis
Group	67	62	.83	4.95	1%	rejected
Boys	65	62	.84	1.72	10%	accepted
Girls	68	62	.84	5.31	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(.83)\ (1.17)\ (Y-67)+62$	45%	8
Boys	of insignificant predictive value		
Girls	$X'=(.84)\ (1.15)\ (Y-68)+62$	46%	7

Where X' is the predicted June mark and Y is the average mark for the year.

TABLE XII

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SCIENCE RESULTS IN CONTINUAL TESTING

Case	Average Year 1964 1966	Mark June 1965 1966	r	t	sig. level	null hypothesis
Group	65	59	.80	5.31	1%	rejected
Boys	63	57	.79	3.68	1%	rejected
Girls	67	61	.79	4.73	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	X'=(.80) (1.45) (Y-65)+59	40%	7
Boys	X'=(.79) (1.55) (Y-63)+57	39%	6
Girls	X'=(.79) (1.43) (Y-67)+61	39%	6

Where X' is the predicted June Mark and Y is the average mark for the year.

again proves to be the superior predictor for the entire groups and for the female groups. The percentages of efficiency of 45% and 46% respectively must be rated superior to the 40% and 39% respectively for the continual testing.

Analysis of Tables XIII and XIV

All the differences between the first term results and the June average scores are significant for both testing systems in Mathematics. In this situation, continual testing proves to be the superior predictor for the entire group and the female group. Term testing proves the superior predictor for the male group. All have high estimates of error.

Analysis of Tables XV and XVI

In Mathematics the average mark for the year and the June mark have no significant difference. This holds true in every case in Tables XVI and XVII. In this case both systems are unacceptable as predictors for the purpose of this study. Neither can be considered the superior predictor of the June mark.

TABLE XIII

PREDICTION OF JUNE RESULTS FROM FIRST TERM MATHEMATICS RESULTS IN TERM TESTING

Case	Average First Term 1962 1963	Mark June 1963 1964	r	t	sig. level	null hypothesis
Group	77	70	.71	3.82	1%	rejected
Boys	79	74	.83	2.28	5%	rejected
Girls	76	67	.62	3.34	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	X'=(.71) (1.14) (Y-77)+70	29%	13
Boys	X'=(.83) (1.09) (Y-79)+74	45%	10
Girls	X'=(.62) (1.16) (Y-76)+67	22%	14

Where X' is the predicted June Mark and Y is the first term average.

TABLE XIV

PREDICTION OF JUNE RESULTS FROM FIRST TERM MATHEMATICS RESULTS IN CONTINUAL TESTING

Case	Average First Term 1964 1965	Mark June 1965 1966	r	t	sig. level	null hypothesis
Group	67	62	.81	3.81	1%	rejected
Boys	67	62	.79	3.26	1%	rejected
Girls	67	62	.83	2.72	1%	rejected

	Regression Equation	Forecasting Efficiency	Estimate of Error
Group	$X'=(.81)\ (1.22)\ (Y-67)+62$	41%	9
Boys	$X'=(.79)\ (1.16)\ (Y-67)+62$	39%	8
Girls	$X'=(.83)\ (1.26)\ (Y-67)+62$	45%	9

Where X' represents the predicted June mark and Y represents the first term averages.

TABLE XV

PREDICTION OF JUNE RESULTS FROM THE YEAR'S MATHEMATICS RESULTS IN TERM TESTING

Case	Average Year 1962 1964	Mark June 1963 1964	r	t	sig. level	null hypothesis
Group	72	70	.85	1.64	10%	accepted
Boys	75	74	.82	1.43	20%	accepted
Girls	70	67	.85	1.73	10%	accepted

None of the above differences are significant.

TABLE XVI

PREDICTION OF JUNE RESULTS FROM THE YEAR'S MATHEMATICS RESULTS IN CONTINUAL TESTING

Case	Average Year 1964 1966	Mark June 1965 1966	r	t	sig. level	null hypothesis
Group	63	62	.87	.9	40%	accepted
Boys	63	62	.90	.61	60%	accepted
Girls	63	62	.86	.67	50%	accepted

None of the above relationships are significant.

TABLE XVII

FIRST TERM PREDICTION FACTORS IN TERM TESTING

Case	Average First Term 1962 -1963	Mark June 1963 1964	r	t	sig. level	null hypothesis
Language						
Group	68	63	.70	4.80	1%	rejected
Boys	64	60	.71	2.35	5%	rejected
Girls	71	65	.70	4.00	1%	rejected
Social Studies						
Group	61	59	.78	1.43	20%	accepted
Boys	60	59	.83	0.00	90%	accepted
Girls	62	59	.75	1.05	30%	accepted
Science						
Group	70	62	.74	5.97	1%	rejected
Boys	66	62	.75	1.86	10%	accepted
Girls	73	62	.77	6.96	1%	rejected
Mathematics						
Group	77	70	.71	3.82	1%	rejected
Boys	79	74	.83	2.28	5%	rejected
Girls	76	67	.62	3.34	1%	rejected

TABLE XVIII

FIRST TERM PREDICTION FACTORS IN CONTINUAL TESTING

Case	Average First Term 1964 1965	Mark June 1965 1966	r	t	sig. level	null hypothesis
Language						
Group	72	58	-.67	13.00	1%	rejected
Boys	69	54	-.57	8.90	1%	rejected
Girls	75	61	-.62	9.60	1%	rejected
Social Studies						
Group	55	53	.77	1.60	10%	accepted
Boys	51	52	.68	0.00	90%	accepted
Girls	57	54	.80	1.18	30%	accepted
Science						
Group	64	59	.60	2.61	2%	rejected
Boys	59	57	.56	1.37	20%	accepted
Girls	67	61	.61	2.80	1%	rejected
Mathematics						
Group	67	62	.81	3.81	1%	rejected
Boys	67	62	.79	3.26	1%	rejected
Girls	67	62	.83	2.72	1%	rejected

TABLE XIX

THE YEAR'S PREDICTION FACTORS IN TERM TESTING

Case	Average Year 1962 1964	Mark June 1963 1964	r	t	sig. level	null hypothesis
Language						
Group	66	63	.87	3.26	1%	rejected
Boys	63	60	.84	2.24	5%	rejected
Girls	68	65	.78	2.38	5%	rejected
Social Studies						
Group	61	59	.92	1.04	40%	accepted
Boys	60	59	.93	0.00	90%	accepted
Girls	61	59	.91	1.6	20%	accepted
Science						
Group	67	62	.83	4.95	1%	rejected
Boys	65	62	.84	1.72	10%	accepted
Girls	68	62	.84	5.31	1%	rejected
Mathematics						
Group	72	70	.85	1.64	10%	accepted
Boys	75	74	.82	1.43	20%	accepted
Girls	70	67	.85	1.73	10%	accepted

TABLE XX

THE YEAR'S PREDICTION FACTORS IN CONTINUAL TESTING

Case	Average Year 1964 1966	Mark June 1965 1966	r	t	sig. level	null hypothesis
Language						
Group	70	58	.80	14.29	1%	rejected
Boys	67	54	.73	9.09	1%	rejected
Girls	73	61	.81	11.02	1%	rejected
Social Studies						
Group	57	53	.72	3.08	1%	rejected
Boys	55	52	.65	1.36	20%	accepted
Girls	59	54	.75	3.18	1%	rejected
Science						
Group	65	59	.80	5.31	1%	rejected
Boys	63	57	.79	3.68	1%	rejected
Girls	67	61	.79	4.73	1%	rejected
Mathematics						
Group	63	62	.87	.90	40%	accepted
Boys	63	62	.90	.61	60%	accepted
Girls	63	62	.86	.67	50%	accepted

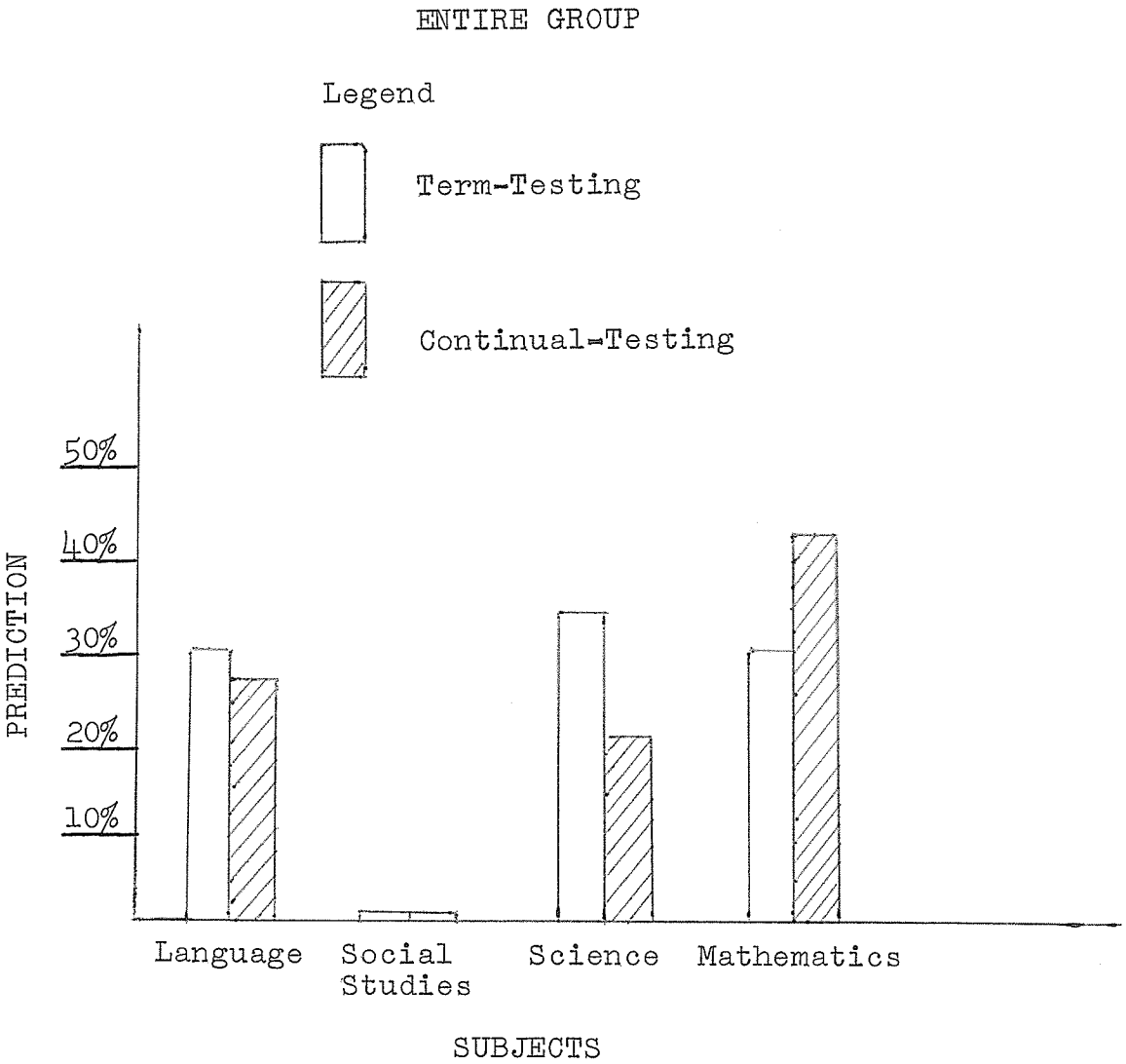


Fig. 1. First Term Results as Predictors of June Results.

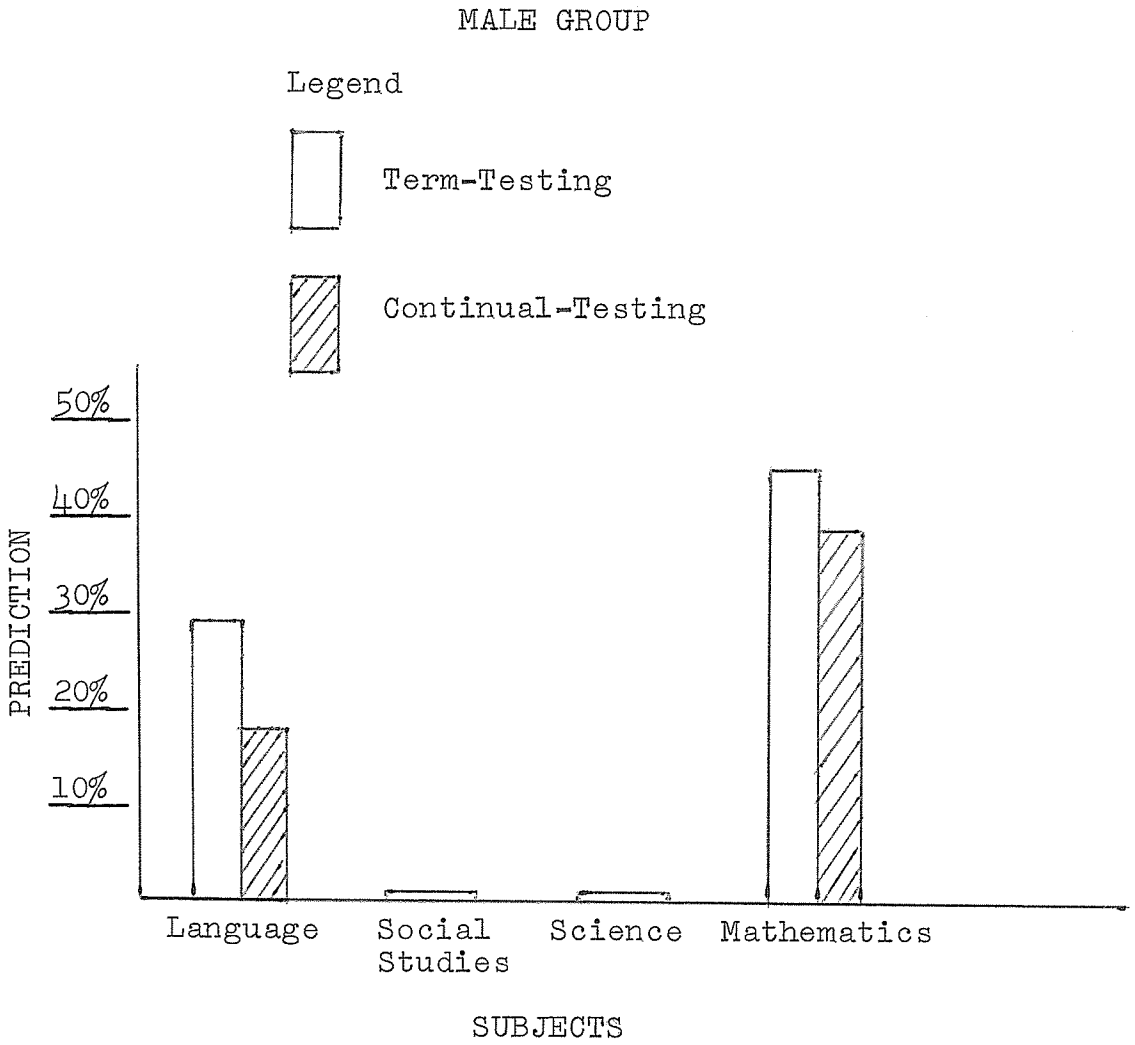


Fig. 2. First Term Results as Predictors of June Results.

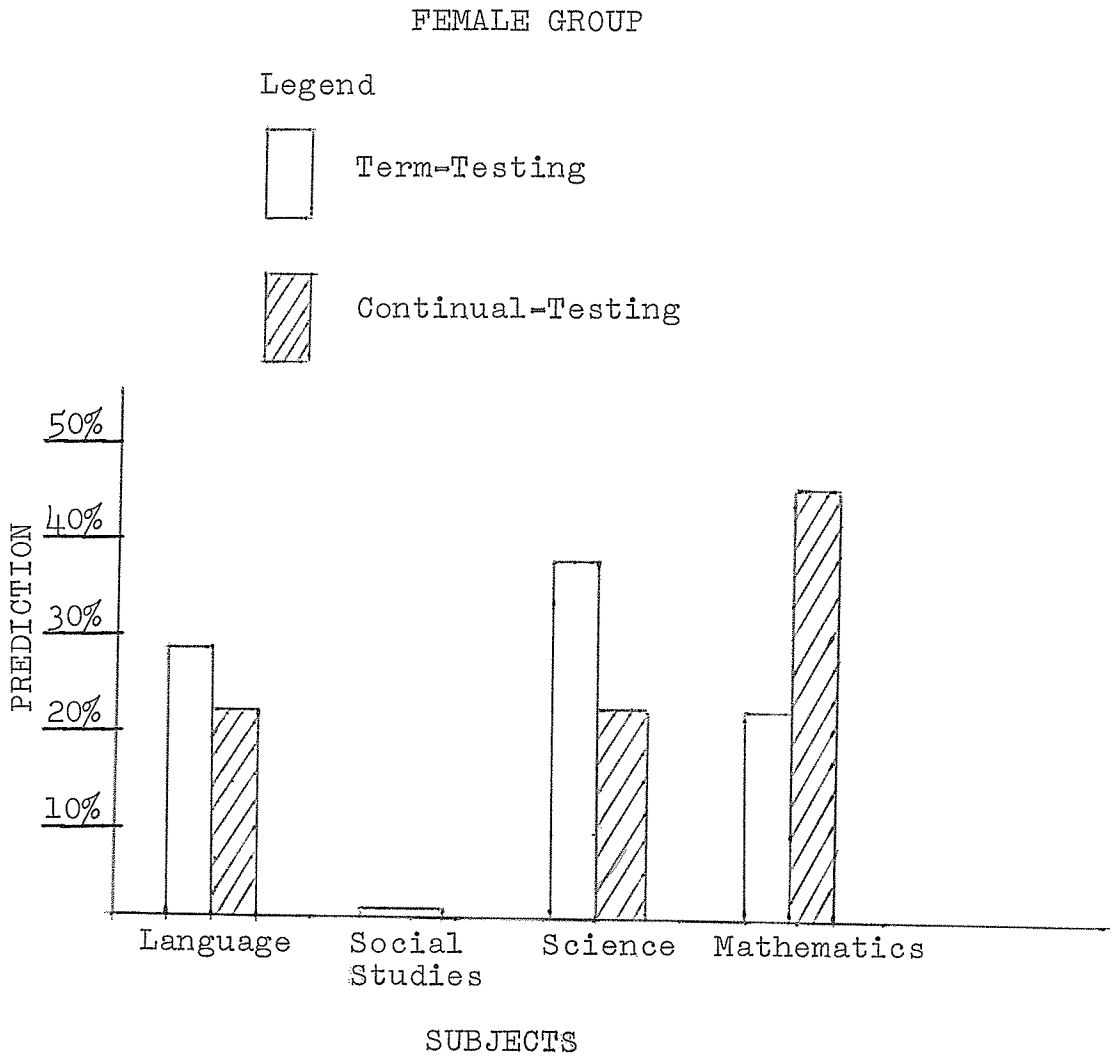


Fig. 3. First Term Results as Predictors of June Results.

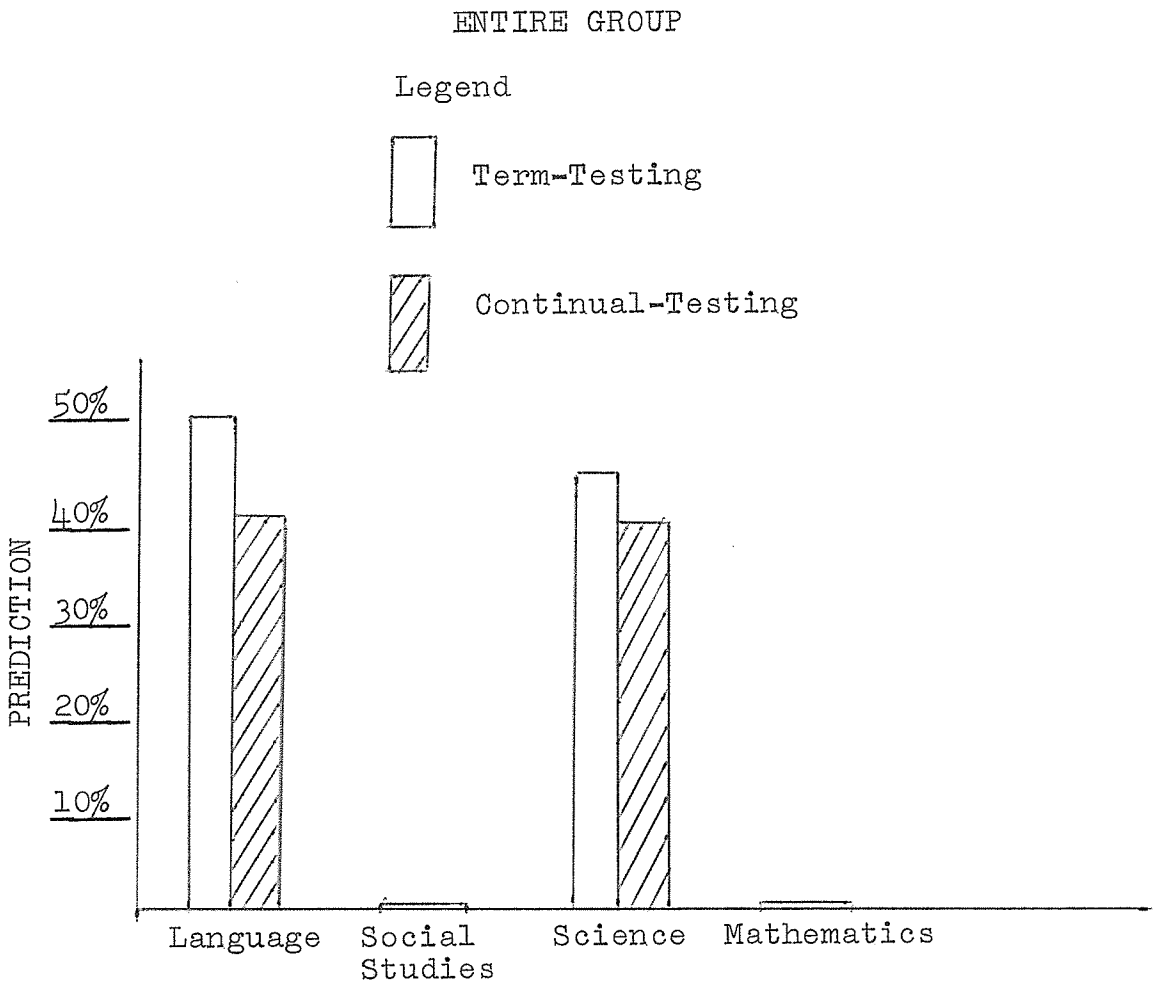


Fig. 4. Average Results for the Year as Predictors of June Results

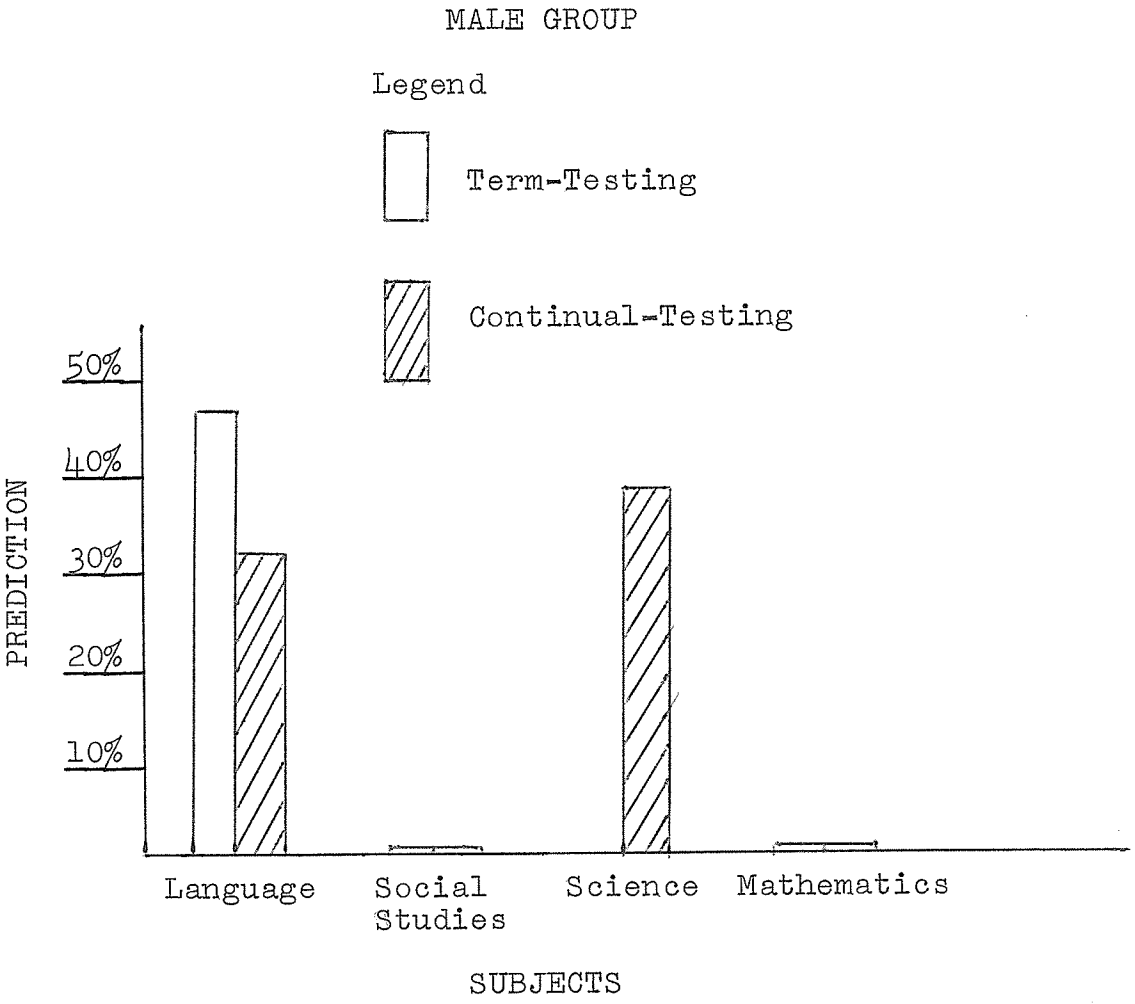


Fig. 5. Average Results for the Year as Predictors of June Results.

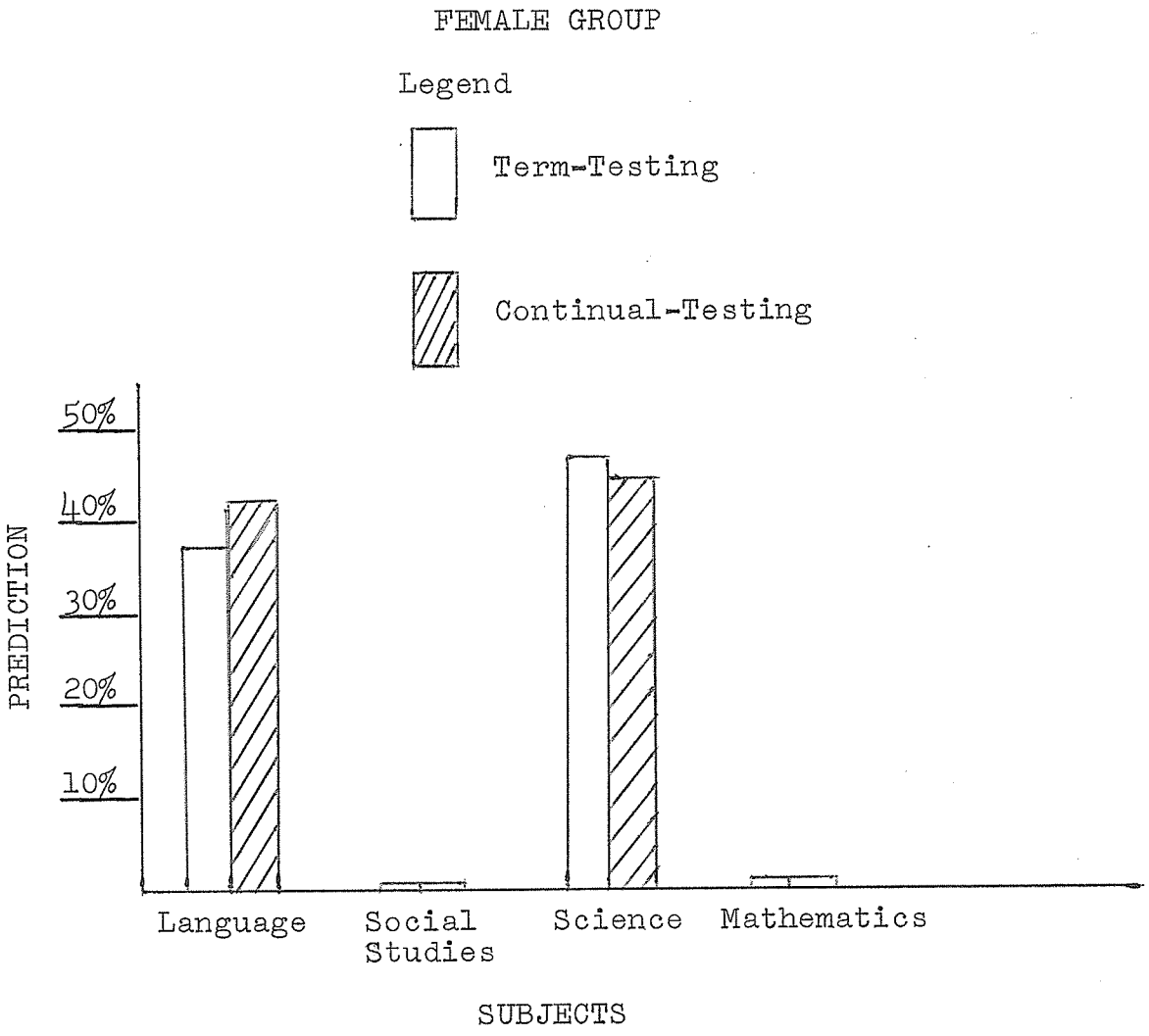


Fig. 6. Average Results for the Year as Predictors of June Results.

Summary

Tables I to XVI present the statistics on which to compare the two testing systems as predictors of June results. Tables XVII to XX present the same material in a repetitive but in concentrated form. Figures 1 to 6 present the percentage of prediction as graphs. These percentage comparisons were the purpose of this phase of the study.

The results in language have a significant difference to the June results in all cases. This holds true for both testing systems. The first term results of term testing are the better predictors of June results for all cases. The average of the year of term testing are the better predictors of June results for the entire group and for the male group. The continual testing system is the better predictor for the female group in the latter comparison.

There are no significant relationships between first term results and June results in Social Studies. Neither are there any significant relationships between the average mark for the year and the June results in this subject. Neither testing system can be considered the better predictor on the basis of these statistics.

Science has only one significant relationship on the basis of first term results. This is the entire group case. Term testing proves to be superior in this case. When the average mark for the year is considered, all cases, except

one, have significant relationships to June results. The one exception is the male group in term testing. Term testing is the better predictor of June results for the entire group and for the female group in Science. Continual testing is the better predictor for the male group. Mathematics is a strong predictor on the basis of the first term results. On the evidence of these results, continual testing is the superior predictor of June results for the entire group and for the female group. Term testing is the better predictor for the male group. Mathematics has no significant relationships when the average mark of the year is the basis of comparisons.

As Motivator

A COMPARISON OF THE TWO TESTING SYSTEMS TO ESTABLISH WHICH SYSTEM PROVIDES BETTER MOTIVATION

The second method of comparison examined the two testing systems as methods to motivate the students. In this comparison the June examination results were compared. An attempt was made to locate the testing system which produced higher June marks.

The same matched pairs were used as for the comparison for predictors. This resulted in comparing the June results of the 1963 students with the June results of the 1965 students. Similarly the June results of the 1964 students were compared with those of the 1966 students.

TABLE XXI

JUNE MEAN RESULTS PRODUCED BY TERM TESTING AND CONTINUAL TESTING FROM SEPTEMBER, 1962, TO JUNE, 1966

Subject	Term	Continual	r	t	level	hypothesis
Language	63	58	.30	3.15	1%	rejected
Social Stud.	59	53	.27	2.32	5%	rejected
Science	61	59	.22	1.31	20%	accepted
Mathematics	70	59	.12	3.72	1%	rejected
MALE GROUPS						
Language	60	54	.36	2.51	2%	rejected
Social Stud.	59	52	.33	2.1	5%	rejected
Science	62	57	.25	1.91	10%	accepted
Mathematics	74	57	.19	4.34	1%	rejected
FEMALE GROUPS						
Language	65	61	.27	1.45	20%	accepted
Social Stud.	59	54	.26	1.20	30%	accepted
Science	62	61	.20	.32	80%	accepted
Mathematics	67	62	.06	1.48	20%	accepted

Again the comparisons were made in the four core subjects. They were further made for the entire groups, the male groups and the female groups.

The raw scores were analyzed and from these r and t tests were computed. This information was used to calculate whether a significant difference existed between the two averages. The null hypothesis was rejected at the 5% level, but accepted at any higher percentage.

The purpose of this section of the study was to establish the higher June averages. The null hypothesis asserted that there was no real difference in the averages produced by term testing and by continual testing. It held that the difference that appeared to be present was due to chance fluctuation of marks. The t-score was applied to the results to establish the level of confidence. The results are tabulated in Table XXI.

Analysis of the Entire Groups

Language showed a difference in the June averages for the entire groups. The null hypothesis was rejected. Therefore, the difference in the average marks was possibly due to the different testing systems used. The difference was a matter of 5 marks. This evidence would claim term testing to provide better motivation for students in language.

Term testing was also the better method of motivation in social studies and in mathematics. The former showed an average difference of 6 marks while the latter showed an average difference of 11 marks. The social study average had a significance level of 5% while the mathematics average was significant at the 1% level.

Science was the only subject of the four core subjects compared that showed no real difference in the June average marks. The null hypothesis was accepted and the numerical difference was possibly due to chance fluctuations. According to these statistics, both testing systems appear to be equally good in providing motivation for the students.

Analysis of the Male Groups

Language showed a difference of 6 marks on the average for the male groups. The t-score showed a significance level at 2%. Consequently the null hypothesis was rejected. Term testing was accepted as the superior system in providing motivation for male students in language.

Social studies and mathematics were also found to have a high level of confidence. The level of significance for these two subjects was 5% and 1% respectively. Here, too, the null hypothesis could be rejected. Therefore, the average difference of 7 marks for social studies and 17 marks for mathematics were possibly due to the different

testing systems used. Term testing was again considered to be the superior motivator for these two subjects.

The 5 mark difference in science had too low a level of confidence to warrant the rejection of the null hypothesis. The significance level was 10%. It was, therefore, assumed that there was no real difference in the average marks resulting from term testing and continual testing. The apparent difference was assumed to be due to chance fluctuation. The level of confidence indicated that this difference in averages could conceivably happen as frequently as 20 times out of 100 by pure chance. Consequently both testing systems were taken to be equally good motivators for students in Grade IX science.

Analysis of the Female Groups

The female groups show a marked difference to the entire groups and the male groups. No subject shows a real difference in June averages. In each subject the June averages show a level of confidence too low to permit the rejection of the null hypothesis. The difference in averages of 4%, 5%, 1% and 5% for language, social studies, science and mathematics respectively were considered to be due to chance fluctuations. These fluctuations could have taken place by pure chance as many as 20, 30, 80 and 20 times out of a 100 for the respective subjects. It was,

therefore, assumed that both testing systems are equally good as motivators for girls in all of the four core subjects.

Summary

Term testing appeared to provide the better motivation in language, social studies and mathematics for the entire groups. This was also the case for the male groups. Both testing systems appeared to be equally good motivators for all three groups in science. The female groups appeared to be an exception. Apparently the testing systems made no real difference to the female groups. They are, therefore, equally proficient as motivators for the female groups in all of the four core subjects.

CHAPTER V

EXPERIMENTAL STUDY--SAMPLE B

Procedure

The testing program of Sample B was carried out identically to that described for Sample A in Chapter IV. Sample B was run from September, 1966 to June, 1967. During this time both the control group and experimental group were run simultaneously. The same teachers taught the same subject to both the groups. The control group was tested by terms while the experimental group was tested continually. Both groups were taken off other incentive programs that were run in school.

Statistical Analysis

The statistical analysis for Sample B was identical to that of Sample A. It should, however, be noted that one of the students, student number five, of the control group discontinued attendance before the end of the school year. His partner was dropped from the experimental group for calculative purposes. Each group, therefore, consisted of 26 students, 13 girls and 13 boys.

The Results As Predictor

Analysis of Tables XXII and XXIII

In language, the control group showed a significant difference between the first term results and the June marks.

TABLE XXII

PREDICTION OF JUNE RESULTS FROM FIRST TERM LANGUAGE RESULTS IN TERM TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	69	59	.87	8.7	1%	rejected
Boys	67	58	.88	5.65	1%	rejected
Girls	71	60	.82	.666	40-50%	accepted

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.87)\ (.92)\ (Y-69)+59$	51%	5
Boys	$X'=(.88)\ (.85)\ (Y-67)+58$	53%	6
Girls	$X'=(.82)(1.09)\ (Y-71)+60$	43%	5

Where Y is the First Term Mark and X' is the Predicted June Mark.

TABLE XXIII

PREDICTION OF JUNE RESULTS FROM FIRST TERM LANGUAGE RESULTS IN CONTINUAL TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	69	60	.89	5.2	1%	rejected
Boys	67	56	.87	4.85	1%	rejected
Girls	70	64	.63	2.76	1-2%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.89)\ (.92)\ (Y-69)+60$	54%	6
Boys	$X'=(.86)(1.13)\ (Y-67)+56$	49%	6
Girls	$X'=(.62)\ (.82)\ (Y-70)+64$	22%	9

Where Y is the First Term Mark and X' is the Predicted June Mark.

This difference was significant at the 1% level. The relationship had predictive value and could be used for the entire group. The male group showed similar value.

The difference in the marks occurred too frequently in the female group to be of predictive value. The 51% and 53% index of forecasting efficiency for the entire group and the male group, respectively, would be decided advantages over a pure guess. The estimates of error were also low, 5 and 6 respectively, enough to make the predicted mark reasonably accurate.

Continual testing showed significant differences between the first term results and the June results for all three groups. These results could, therefore, be used for predictive purposes.

Tables XXII and XXIII indicated that continual testing was slightly more accurate in predicting June marks for the entire group. The difference in forecasting efficiency was 3%. Continual testing, however, showed a slightly higher estimate of error. Term testing held a slight edge over continual testing in the male group. Here the estimate of error was identical. According to the tables, term testing was unacceptable as a predictor because the observed differences happened too frequently by random chance. Consequently, continual testing was the better predictor in

this case. Its forecasting efficiency, however, was low, 22%, and its error of estimate was high, ±9 marks.

Analysis of Tables XXIV and XXV

Term testing and continual testing both showed significant differences between the year's average mark and the average June mark for all groups. The level of significance was 1% for each group in both testing systems. Continual testing, however, had much higher correlation coefficients than term testing for both the entire group and the male group. These were .92 and .95 in comparison to .78 and .72 for the entire group and the male group respectively. Consequently the continual testing had much higher predictive efficiency. Continual testing also had considerably lower estimates of error for these groups.

According to the tables, term testing was a superior predictor of June marks for the female group. Its forecasting efficiency was 63% in comparison to 51% for continual testing. The estimate of error was also lower for term testing than for continual testing, 3 marks in comparison to 5 marks.

Analysis of Tables XXVI and XXVII

Table XXVI showed term testing to have acceptable predictive value for all three groups. Predicting efficiencies were 36%, 38% and 40% respectively for the entire,

TABLE XXIV

PREDICTION OF JUNE RESULTS FROM THE YEAR'S LANGUAGE RESULTS IN TERM TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	69	58	.78	6.84	1%	rejected
Boys	68	57	.72	3.61	1%	rejected
Girls	69	59	.93	10.00	1%	rejected

	Regression Equation	Index of Fore casting Efficiency	Estimate of Error
Group	X'=(.78) (.90) (Y-69)+58	37%	7
Boys	X'=(.72) (.79) (Y-68)+57	31%	10
Girls	X'=(.93)(1.20) (Y-69)+59	63%	3

Where Y is the Year's Average Mark and X' is the predicted June Mark.

TABLE XXV

PREDICTION OF JUNE RESULTS FROM THE YEAR'S LANGUAGE RESULTS IN CONTINUAL TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	67	60	.92	7.76	1%	rejected
Boys	65	56	.95	6.38	1%	rejected
Girls	70	64	.87	4.34	1%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	X'=(.92) (1.08) (Y-67)+60	61%	5
Boys	X'=(.95) (1.12) (Y-65)+56	69%	4
Girls	X'=(.87) (.91) (Y-70)+64	51%	5

Where Y is the year's average mark and X' is the predicted June Mark.

TABLE XXVI

PREDICTION OF JUNE RESULTS FROM FIRST TERM SOCIAL STUDIES RESULTS IN TERM TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	46	55	.78	3.59	1%	rejected
Boys	45	58	.80	3.36	1%	rejected
Girls	46	52	.80	2.74	5%	rejected

	Prediction Formula Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.77)\ (.97)\ (Y-46)+55$	36%	10
Boys	$X'=(.79)\ (.91)\ (Y-45)+58$	38%	13
Girls	$X'=(.80)(1.02)\ (Y-46)+52$	40%	9

Where Y is the First Term Mark and X' is the predicted June Mark.

TABLE XXVII

PREDICTION OF JUNE RESULTS FROM FIRST TERM SOCIAL STUDIES RESULTS IN CONTINUAL TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	56	57	.91	.58	50-60%	accepted
Boys	56	53	.96	1.14	20-30%	accepted
Girls	57	61	.87	2.78	1-2%	rejected

	Prediction Formula Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.92)(1.31)\ (Y-56)+57$	61%	6
Boys	$X'=(.95)(1.31)\ (Y-56)+53$	69%	6
Girls	$X'=(.87)(1.20)\ (Y-57)+61$	51%	5

Where Y is the 1st term mark and X is the predicted June mark.

male and female groups. The estimates of error were high at + or - 10, 13, and 9.

In continual testing, both the entire group and the male group had no value as predictors. These differences happened too frequently by pure chance. The female group, however, showed high predictive value with 51% forecasting efficiency. It also has a low estimate of error, + or - 5 marks.

Term testing appeared to be the better predictor for the entire group and for the male group. Continual testing was the better predictor for the female group. All such predictions of June marks based on the results of the first term testing.

Analysis of Tables XXVIII and XXIX

In social studies only term testing was acceptable as a predictor of June marks when based on the Year's average. It had predictive value for the entire group. The male and female groups showed random occurrence too frequently to be of acceptable predictive value.

This was also the case for all three groups in continual testing. They failed to show a significant difference between the year's average mark and the June mark.

Since the male and female groups showed that this difference occurred too frequently at random, neither testing system could be indicated as the superior predictor.

TABLE XXVIII

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SOCIAL STUDIES RESULTS IN TERM TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	49	55	.92	2.51	2%	rejected
Boys	54	58	.90	1.65	10-20%	accepted
Girls	46	51	.91	1.85	5-10%	accepted

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	X'=(.92)(1.01) (Y-49)+55	61%	7
Boys	X'=(.90)(1.03) (Y-54)+58	56%	8
Girls	X'=(.91)(1.08) (Y-46)+51	59%	6

Where Y is the Year's Average Mark and X' is the Predicted June Mark.

TABLE XXIX

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SOCIAL STUDIES RESULTS IN CONTINUAL TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	57	57	.94	0	100%	accepted
Boys	55	52	.97	1.38	10-20%	accepted
Girls	58	62	.88	1.8	5-10%	accepted

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	X'=(.94) (1.31) (Y-57)+57	66%	5
Boys	X'=(.97) (1.38) (Y-55)+52	76%	4
Girls	X'=(.88) (1.02) (Y-58)+62	53%	6

Where Y is the Year's Average Mark and X' is the Predicted June Mark.

Term testing was the better predictor of June marks from the year's average mark for the entire group. Continual testing showed an almost perfect example of pure random chance. This system failed to show a significant difference between the year's average mark and the June mark.

Analysis of Tables XXX and XXXI

Table XXX indicated that term testing showed significant differences between the first term marks and the June marks for all three groups in science. Consequently all three results could be used for the purpose of predicting June marks. The correlation coefficients, however, were low 24%, 23%, and 31% respectively for the entire male and female groups. The estimates of error were high.

In continual testing, the male group showed no significant difference between the first term average and the June mark. Since this difference appeared too frequently by chance, it was unacceptable as a predictor of June marks. Term testing, therefore, was the better predictor for the male group.

The entire group and the female group showed significant differences at the 1% level. Their forecasting efficiency was also better than those of their counterparts in term testing. These facts, coupled with lower estimates of error for the continual testing groups, showed continual testing a superior predictor of June marks for the entire

TABLE XXX

PREDICTION OF JUNE RESULTS FROM FIRST TERM SCIENCE RESULTS IN TERM TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	76	62	.64	7.57	1%	rejected
Boys	75	62	.63	4.01	1%	rejected
Girls	76	61	.72	5.83	1%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	X'=(.65) (.57) (Y-76)+62	24%	13
Boys	X'=(.64) (.71) (Y-75)+62	23%	11
Girls	X'=(.72) (.52) (Y-76)+61	31%	13

Where Y is the First Term Mark and X' is the Predicted June Mark.

TABLE XXXI

PREDICTION OF JUNE RESULTS FROM FIRST TERM SCIENCE RESULTS IN CONTINUAL TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	69	59	.79	3.3	1%	rejected
Boys	66	59	.79	1.22	20-30%	accepted
Girls	71	59	.80	5.5	1%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	X'=(.78) (1.07) (Y-69)+60	38%	9
Boys	X'=(.79) (1.02) (Y-66)+59	39%	10
Girls	X'=(.80) (1.19) (Y-71)+59	40%	7

Where Y is the First Term Mark and X' is the Predicted June Mark.

group and the female group. Such prediction must be based on the average of the 1st term marks.

Analysis of Tables XXXII and XXXIII

Tables XXXII and XXXIII dealt with statistical results based on a comparison of the year's average mark and the June mark. All three groups showed significant differences between the year's average mark and the June mark. This was the case for both term testing and continual testing. In every case the continual testing correlation coefficient was higher than its term testing counterpart. These tables, therefore, indicated that when based on the year's average, continual testing was a better predictor of June marks for all three groups. The female group in continual testing showed a particularly high correlation coefficient, .98. The consequent forecasting efficiency of 80% and the estimate of error of + or - 2 were quite impressive.

Analysis of Tables XXXIV and XXXV

All three groups showed a significant difference between the first term average and the June average. For the entire group, continual testing had a slight edge as the better predictor of June marks. For the male group, term testing proved to be slightly better than continual testing. These two groups were, however, so close that little preference can be inferred.

TABLE XXXII

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SCIENCE RESULTS IN TERM TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	72	62	.78	7.09	1%	rejected
Boys	74	62	.80	6.06	1%	rejected
Girls	70	61	.93	4.00	1%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.78)\ (.82)\ (Y-72)+62$	37%	8
Boys	$X'=(.80)\ (.93)\ (Y-74)+62$	40%	7
Girls	$X'=(.93)\ (.63)\ (Y-70)+61$	63%	5

Where Y is the Average Mark for the Year and X' is the Predicted June Mark.

TABLE XXXIII

PREDICTION OF JUNE RESULTS FROM THE YEAR'S SCIENCE RESULTS IN CONTINUAL TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	66	59	.88	4.76	1%	rejected
Boys	65	59	.89	2.61	2%	rejected
Girls	67	59	.98	4.35	1%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.88)\ (1.17)\ (Y-66)+59$	53%	6
Boys	$X'=(.89)\ (1.17)\ (Y-65)+59$	54%	7
Girls	$X'=(.98)\ (1.18)\ (Y-67)+59$	80%	2

Where Y is the Year's Average and X' is the Predicted June Mark.

TABLE XXXIV

PREDICTION OF JUNE RESULTS FROM FIRST TERM MATHEMATICS RESULTS IN TERM TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	75	65	.74	4.82	1%	rejected
Boys	73	64	.85	4.03	1%	rejected
Girls	77	65	.40	4.93	1%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	X'=(.73) (1.11) (Y-75)+65	32%	7
Boys	X'=(.84) (1.19) (Y-72)+65	46%	7
Girls	X'=(.39) (.89) (Y-77) +66	7%	7

Where Y is the First Term Mark and X' is the Predicted June Mark.

TABLE XXXV

PREDICTION OF JUNE RESULTS FROM FIRST TERM MATHEMATICS RESULTS IN CONTINUAL TESTING

Case	Average 1st Term 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	80	63	.78	5.59	1%	rejected
Boys	76	53	.77	3.11	1%	rejected
Girls	83	73	.87	5.53	1%	rejected

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.77)\ (2.15)\ (Y-80)+62$	36%	7
Boys	$X'=(.76)\ (2.36)\ (Y-76)+53$	35%	8
Girls	$X'=(.87)\ (1.05)\ (Y-83)+73$	51%	4

Where Y is the First Term Mark and X' is the Predicted June Mark.

Continual testing, however, had a decided edge over term testing for the female group. Continual testing showed a prediction efficiency of 51% while its term testing counterpart had an efficiency of only 7%. Continual testing also showed an estimate of error of 4 marks while that of term testing showed 7.

Analysis of Tables XXXVI and XXXVII

The results of comparing the year's average mark and the June average mark showed that in most cases no real difference existed. In all but the continual testing male group, there was no real difference. Consequently these results could not be used to predict June results from yearly averages. The significance level showed that in all cases, except one, the results indicated in the table could happen too frequently by chance.

The one exception, the male group in continual testing, was significant at the 5% level. The correlation, however, was so low, .24, that the predictive efficiency is only 8% with an estimated error of + or - 16.

The observations indicated that no comparison can be made between the entire and female groups in terms of predicting June results from yearly averages in mathematics. Continual testing was superior, in this regard, to term testing for the male group. The small efficiency percentage and the large estimate of error made the mean of distribution almost as reliable as the prediction formula.

TABLE XXXVI

PREDICTION OF JUNE RESULTS FROM THE YEAR'S MATHEMATICS RESULTS IN TERM TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	65	65	.80	0	100%	accepted
Boys	65	64	.92	0	100%	accepted
Girls	66	66	.31	0	100%	accepted

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	X'=(.80) (1.00) (Y-65)+65	40%	7
Boys	X'=(.92) (1.01) (Y-65)+64	61%	6
Girls	X'=(.31) (.91) (Y-66)+66	05%	7

Where Y is the Year's Average Mark and X' is the Predicted June Mark.

TABLE XXXVII

PREDICTION OF JUNE RESULTS FROM THE YEAR'S MATHEMATICS RESULTS IN CONTINUAL TESTING

Case	Average Year 1966-67	Average June 1967	r	t	sig. level	null hypothesis
Group	68	63	.81	1.64	20-30%	accepted
Boys	65	53	.24	2.18	5%	rejected
Girls	72	73	.86	.75	40-50%	accepted

	Regression Equation	Index of Forecasting Efficiency	Estimate of Error
Group	$X'=(.81)\ (1.75)\ (Y-68)+63$	41%	8
Boys	$X'=(.24)\ (1.73)\ (Y-65)+53$	08%	16
Girls	$X'=(.86)\ (\ .94)\ (Y-72)+73$	49%	5

Where Y is the Year's Average Mark and X' is the Predicted June Mark.

TABLE XXXVIII

FIRST TERM PREDICTION FACTORS IN TERM TESTING

Case	Average First Term 1966	Mark June 1967	r	t	sig. level	null hypothesis
Language						
Group	69	59	.87	8.7	1%	rejected
Boys	67	58	.88	5.65	1%	rejected
Girls	71	60	.82	.666	40-50%	accepted
Social Studies						
Group	46	55	.78	3.59	1%	rejected
Boys	45	58	.80	3.36	1%	rejected
Girls	46	52	.80	2.74	5%	rejected
Science						
Group	76	62	.64	7.57	1%	rejected
Boys	75	62	.63	4.01	1%	rejected
Girls	76	61	.72	5.83	1%	rejected
Mathematics						
Group	75	65	.74	4.82	1%	rejected
Boys	73	64	.85	4.03	1%	rejected
Girls	77	65	.40	4.93	1%	rejected

TABLE XXXIX

FIRST TERM PREDICTION FACTORS IN CONTINUAL TESTING

Case	Average First Term 1966	Mark June 1967	r	t	sig. level	null hypothesis
Language						
Group	69	60	.89	5.2	1%	rejected
Boys	67	56	.87	4.85	1%	rejected
Girls	70	64	.63	2.76	1-2%	rejected
Social Studies						
Group	56	57	.91	.58	50-60%	accepted
Boys	56	53	.96	1.14	20-30%	accepted
Girls	57	61	.87	2.78	1-2%	rejected
Science						
Group	69	59	.79	3.3	1%	rejected
Boys	66	59	.79	1.22	20-30%	accepted
Girls	71	59	.80	5.5	1%	rejected
Mathematics						
Group	80	63	.78	5.59	1%	rejected
Boys	76	53	.77	3.11	1%	rejected
Girls	83	73	.87	5.53	1%	rejected

TABLE XL

THE YEAR'S PREDICTION FACTORS IN TERM TESTING

Case	Average Year 1966-1967	Mark June 1967	r	t	sig. level	null hypothesis
Language						
Group	69	58	.78	6.84	1%	rejected
Boys	68	57	.72	3.61	1%	rejected
Girls	69	59	.93	10.00	1%	rejected
Social Studies						
Group	49	55	.92	2.51	2%	rejected
Boys	54	58	.90	1.65	10-20%	accepted
Girls	46	51	.91	1.85	5-10%	accepted
Science						
Group	72	62	.78	7.09	1%	rejected
Boys	74	62	.80	6.06	1%	rejected
Girls	70	61	.93	4.00	1%	rejected
Mathematics						
Group	65	65	.80	0	100%	accepted
Boys	65	64	.92	0	100%	accepted
Girls	66	66	.31	0	100%	accepted

TABLE XLI

THE YEAR'S PREDICTION FACTORS IN CONTINUAL TESTING

Case	Average Year 1966-1967	Mark June 1967	r	t	sig. level	null hypothesis
Language						
Group	67	60	.92	7.76	1%	rejected
Boys	65	56	.95	6.38	1%	rejected
Girls	70	64	.87	4.34	1%	rejected
Social Studies						
Group	57	57	.94	0	100%	accepted
Boys	55	52	.97	1.38	10-20%	accepted
Girls	58	62	.88	1.8	5-10%	accepted
Science						
Group	66	59	.88	4.76	1%	rejected
Boys	65	59	.89	2.61	2%	rejected
Girls	67	59	.98	4.35	1%	rejected
Mathematics						
Group	68	63	.81	1.64	20-30%	accepted
Boys	65	53	.24	2.18	5%	rejected
Girls	72	73	.86	.75	40-50%	accepted

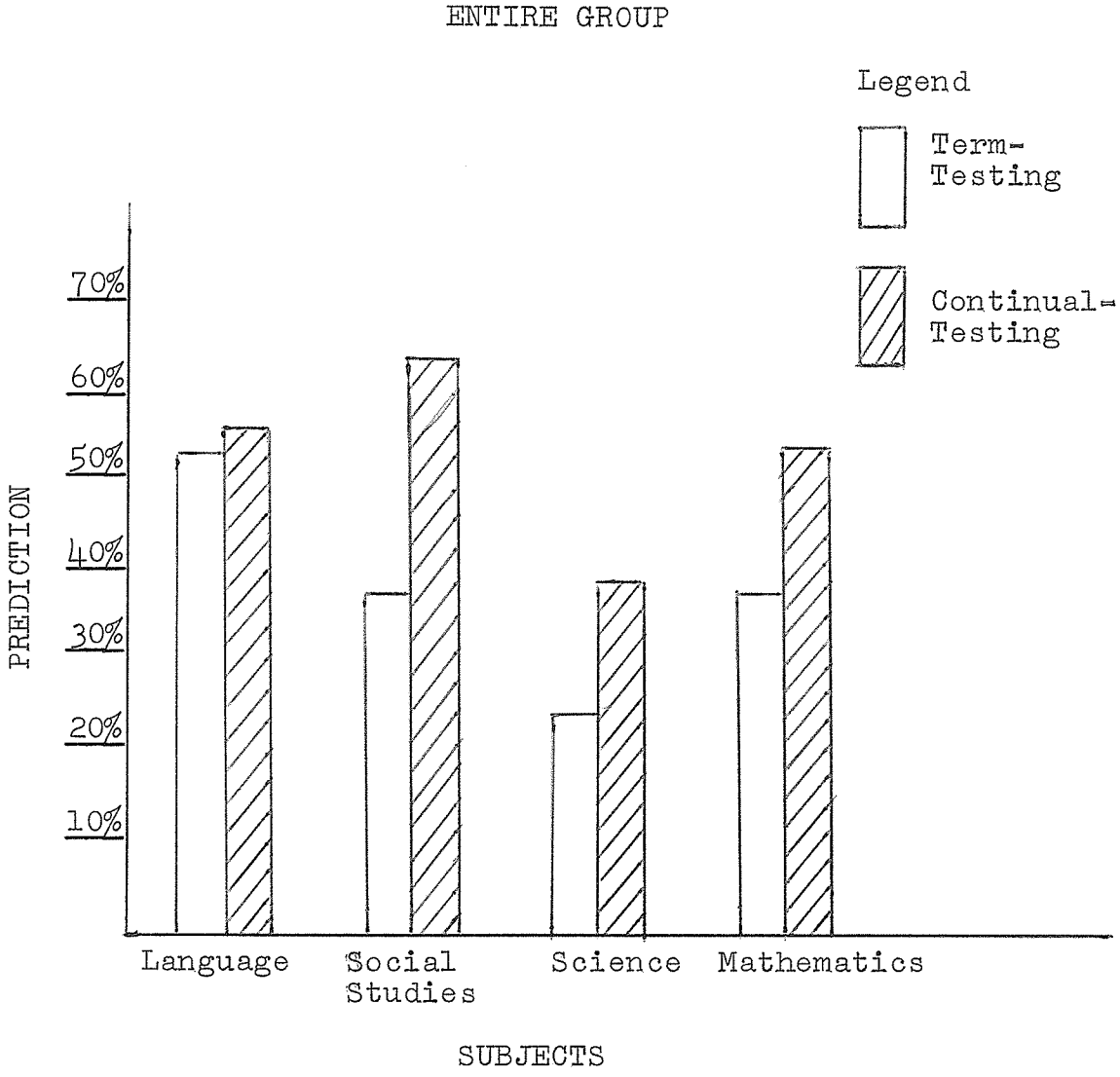


Fig. 7. First Term Results as Predictors of June Results.

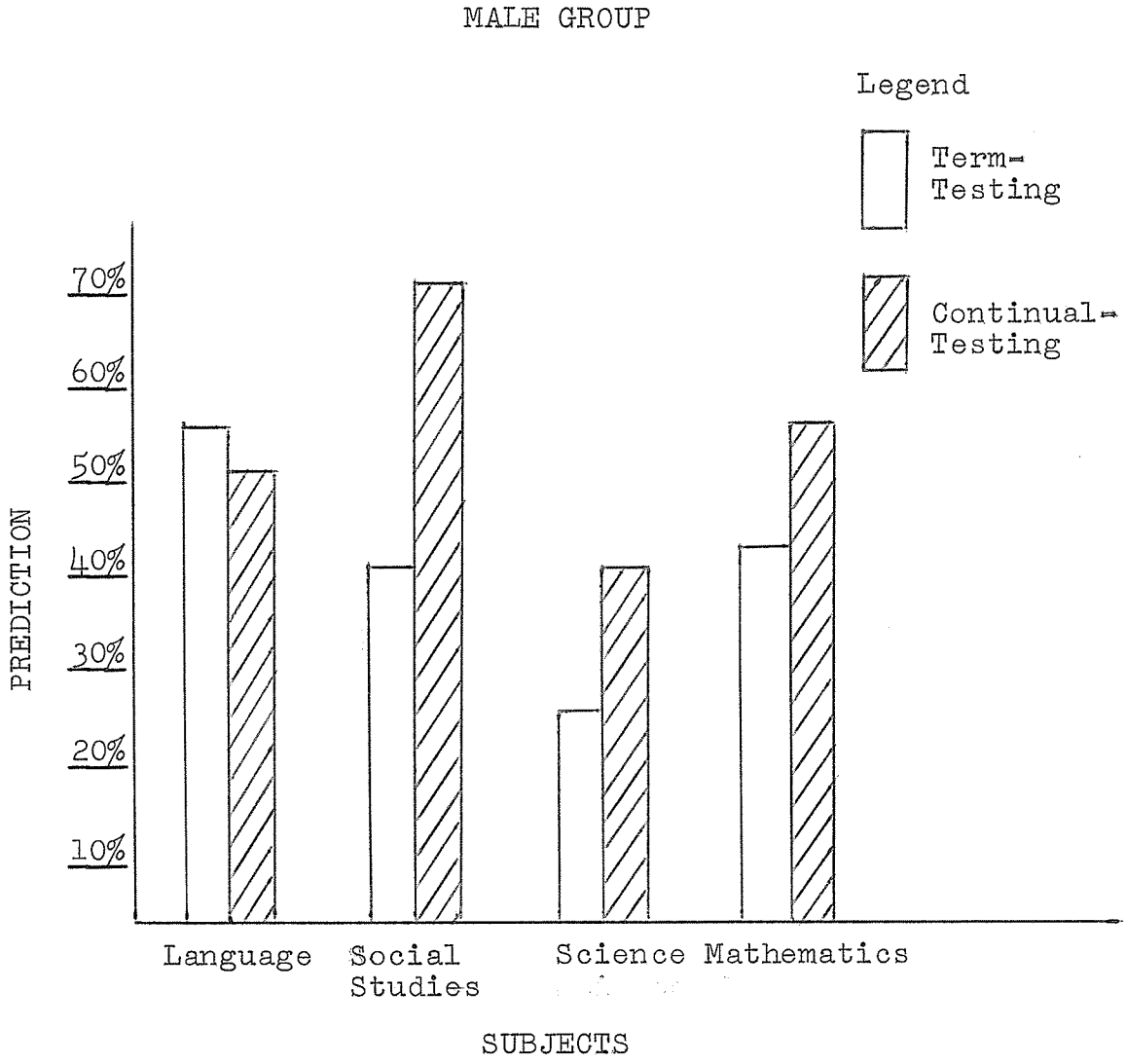


Fig. 8. First Term Results as Predictors of June Results.

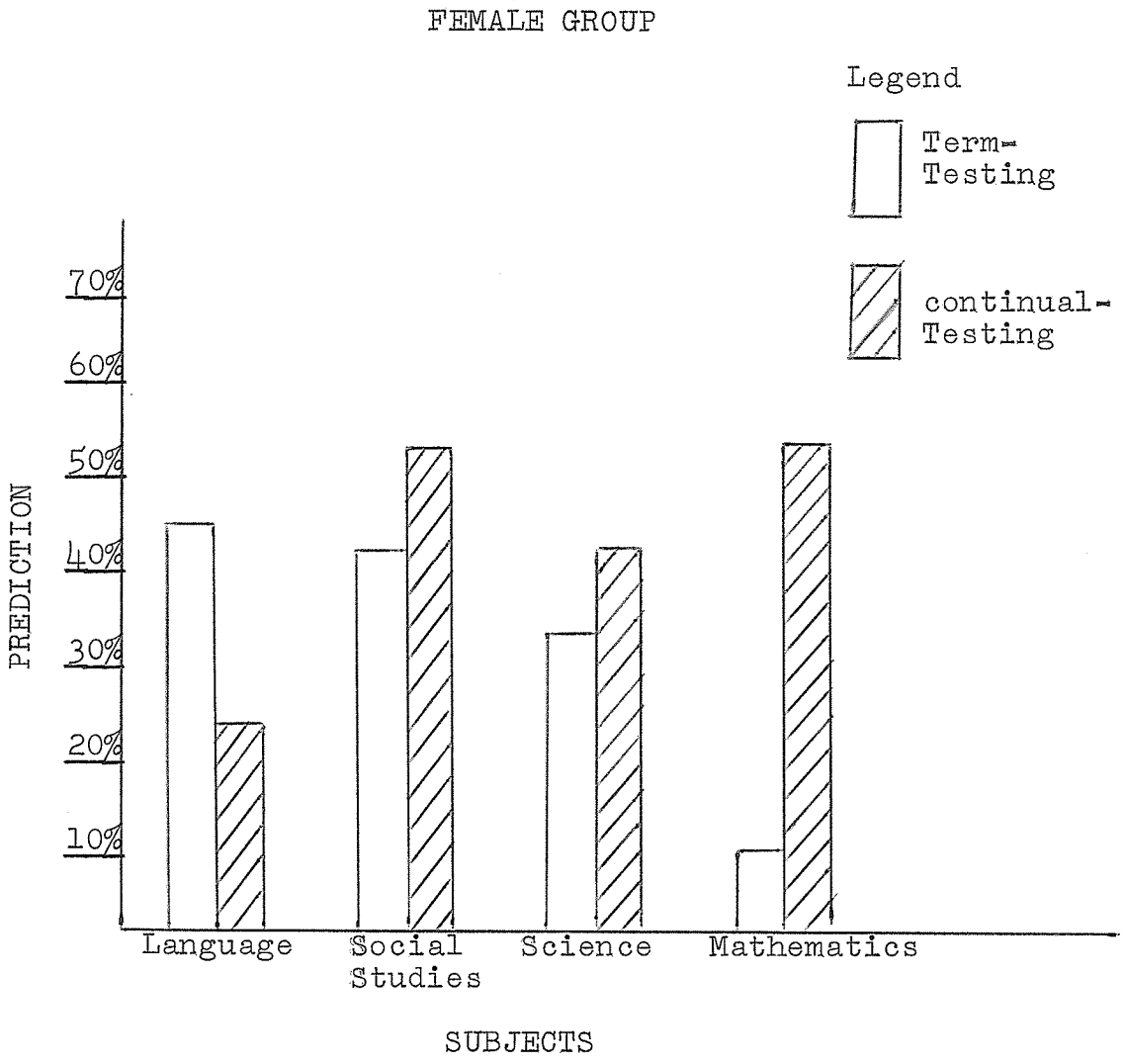


Fig. 9. First Term Results as Predictors of June Results.

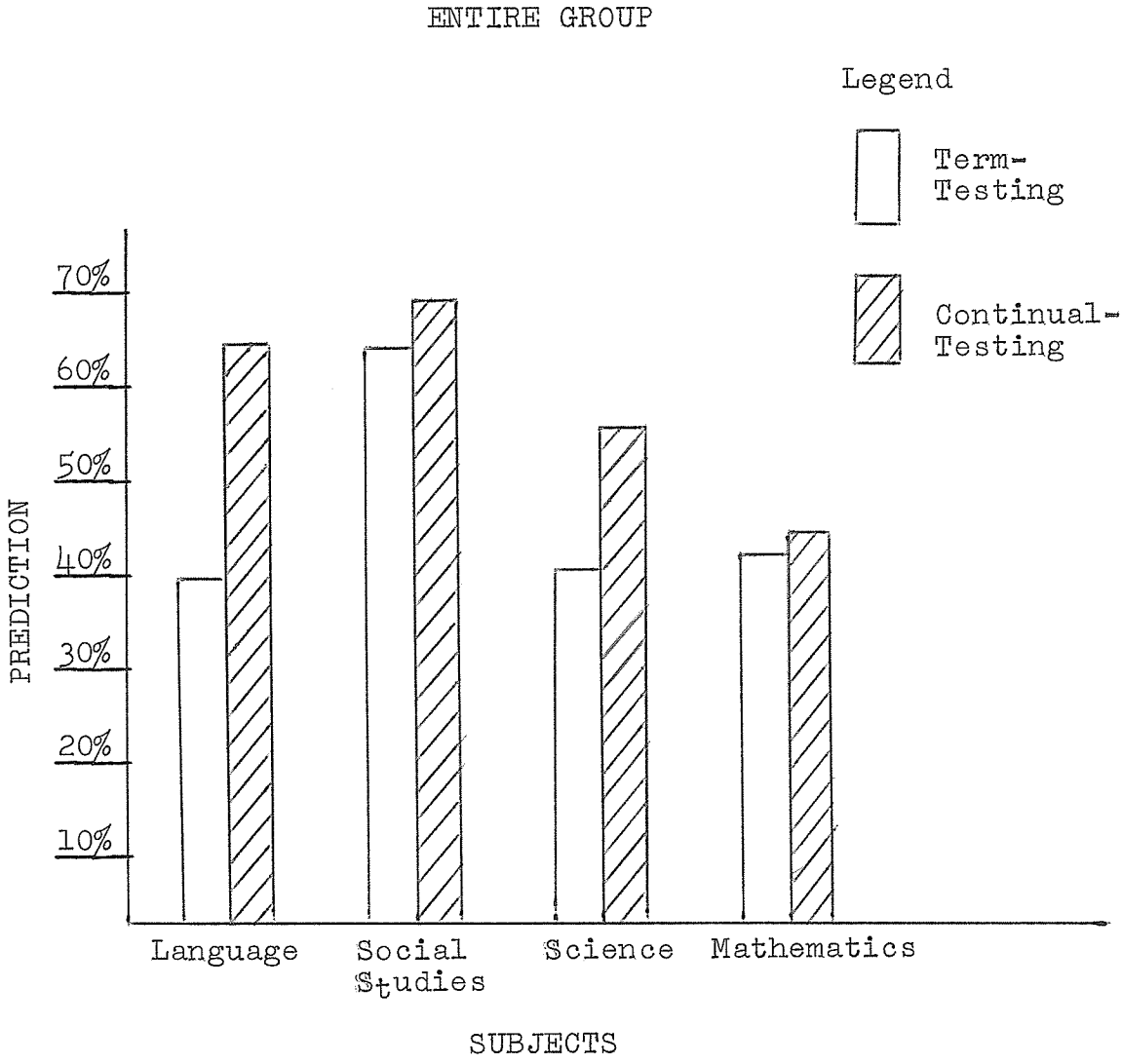


Fig. 10. Average Results for the Year as Predictors of June Results

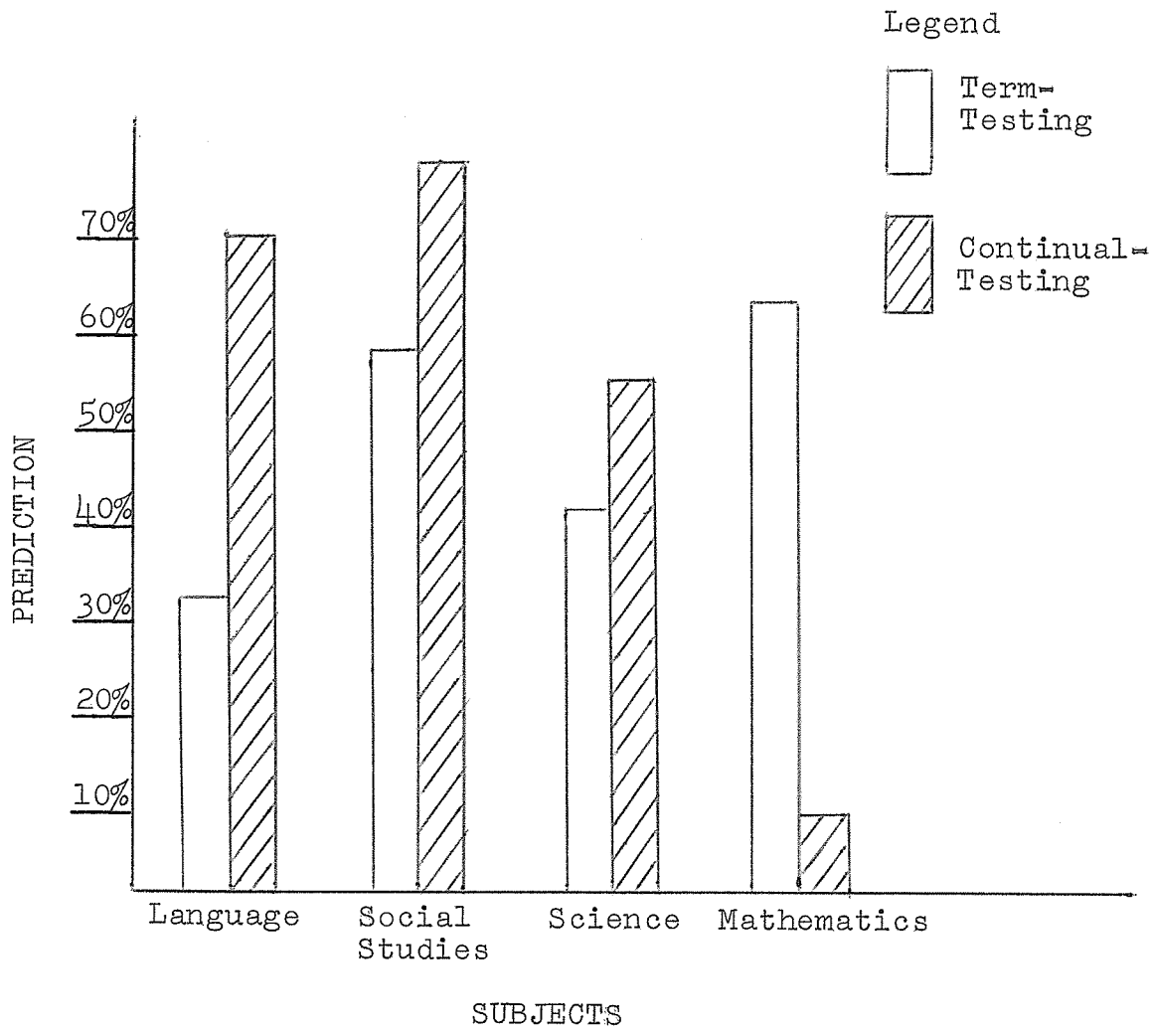


Fig. 11. Average Results for the Year as Predictors of June Results.

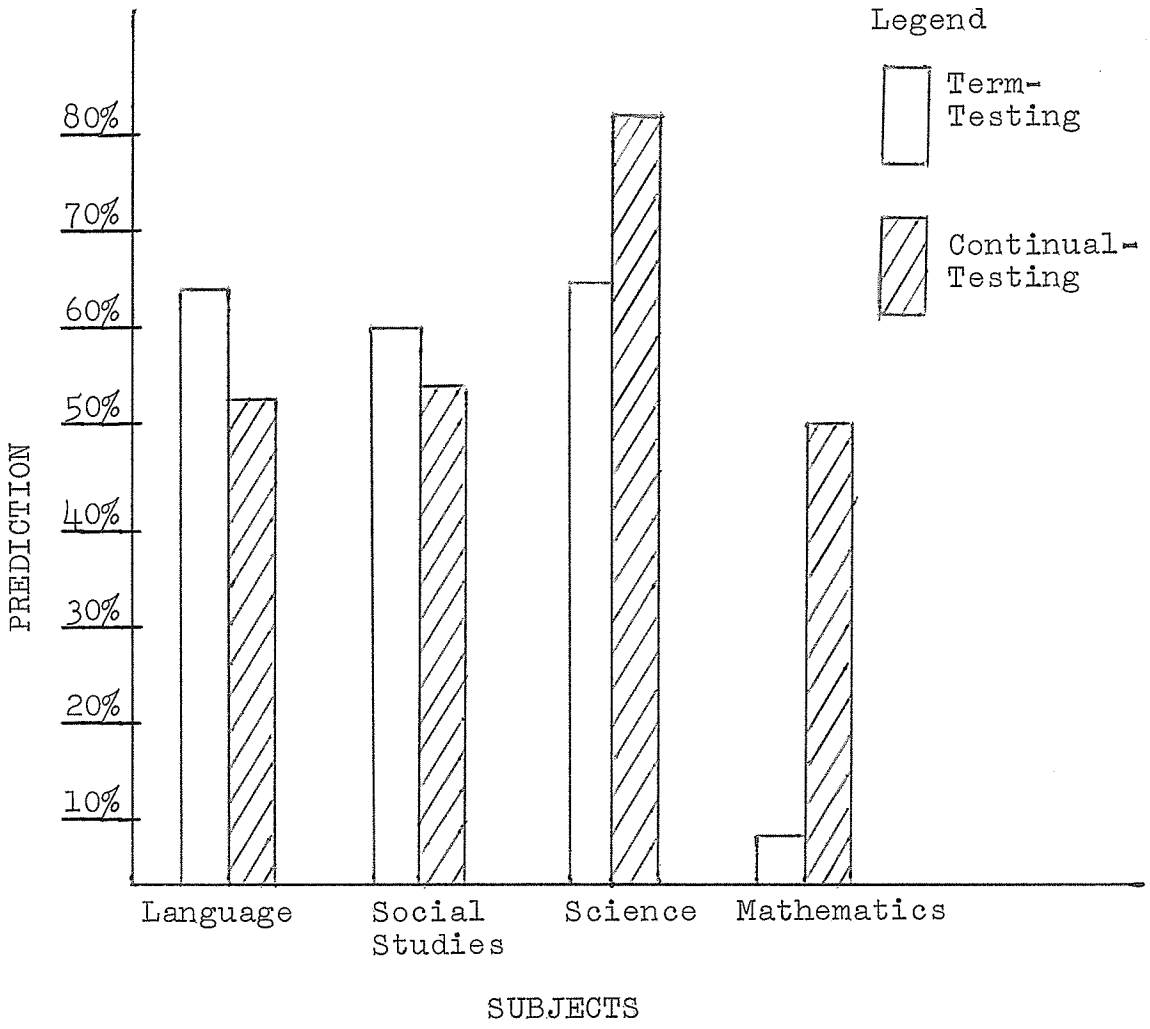


Fig. 12. Average Results for the Year as Predictors of June Results.

Summary

Tables XXII to XXXVII presented the statistics on which to compare the two testing systems as predictors of June results. Tables XXXVIII to XLI presented the same material in a repetitive but in concentrated form. Figures 7 to 12 presented the percentage of prediction as graphs. These percentage comparisons were the purpose of this phase of the study.

A significant difference was found in every language comparison, except one. The female group in the first term comparison in term testing did not produce a significant difference. The first term results for term testing proved to be almost identical to first term results for continual testing. Consequently term testing and continual testing were equally good predictors of June marks, if such prediction were made on the basis of the first term results. The one exception, the female group, indicated that continual testing was the better predictor for this group.

In comparison based on yearly averages, all three groups showed significant differences both in term testing and in continual testing. Continual testing, however, proved to be the better predictor of June marks for the entire group as well as for the male group. Term testing proved to be a more efficient predictor of June marks for the female group.

The entire group and the male group did not show significant differences in social studies when the first term and the June results were compared in continual testing. They did, however, show such significance for term testing. The study, therefore, indicated that for the entire group and the male group term testing was a better predictor of June results when based on first term averages.

Social studies showed a significant difference in first term averages and June averages for the female group. Continual testing showed a slightly higher prediction efficiency. Consequently, continual testing was a better predictor of June results for the female group, providing that such prediction were based on first term averages.

When social study results are compared on a yearly averages versus June average, only one significant difference was noted. This difference was seen for the entire group under term testing. Consequently term testing was the better predictor of June results for the entire group, if such prediction were based on the yearly average. No comparisons could be made for the female and male groups because no significant differences were established. The results that were obtained could happen so frequently by chance that they were unacceptable as predictors for this study.

All except one, first term and June comparisons for

science showed significant differences. The only exception was continual testing for the male group. For the male group term testing proved the better predictor of June marks. For the entire group and the female group, continual testing proved to show a higher predictive efficiency. Such prediction, however, must be based on the first term average mark. Even for these two groups there was only slight advantage on the side of continual testing.

When the yearly average marks and the June marks were compared in science, all comparisons showed significant differences. The predictive differences were, however, small in each case. Slight though the difference was, continual testing proved to be the better predictor of June results for all of the three groups, if such prediction were based on the yearly average marks.

All six comparisons in first term averages and June averages for the two testing systems showed significant differences. Continual testing was only slightly better than term testing for the entire group. Term testing had a slight advantage over continual testing for the male group, while continual testing had a decided advantage over term testing for the female group. In all cases the comparisons were based on the first term averages and the point of comparison were as predictors of June results.

When mathematic results were compared on the yearly average mark basis, the study showed a complete reversal. Only one set of comparison showed a significant difference. Consequently, when based on yearly average marks, term testing was a better predictor of June marks for the male groups. No comparison could be made for the entire groups and the female groups in this study. The results that were obtained could have occurred too frequently by chance to be of predictive value.

TABLE XLII

JUNE MEAN RESULTS PRODUCED BY TERM TESTING AND CONTINUAL TESTING FROM SEPTEMBER, 1966, TO JUNE, 1967

Subject	Term	Continual	r	t	sig. level	null hypothesis
ENTIRE GROUP						
Language	59%	60%	.30	.33	70-80%	accepted
Social Stud.	55	57	-.05	.37	70-80%	accepted
Science	62	59	-.07	.005	100%	accepted
Maths.	65	63	.60	.51	60-70%	accepted
MALE GROUP						
Language	58	56	.24	.42	60-70%	accepted
Social Stud.	58	53	-.004	1.27	20-30%	accepted
Science	62	59	-.001	.005	100%	accepted
Maths.	64	53	.73	1.62	10-20%	accepted
FEMALE GROUP						
Language	60	64	-.10	1.02	30-40%	accepted
Social Stud.	52	61	-.20	1.62	10-20%	accepted
Science	61	59	-.18	.004	100%	accepted
Maths.	65	73	.46	2.99	1-2%	rejected

A COMPARISON OF THE TWO TESTING SYSTEMS TO ESTABLISH WHICH SYSTEM PROVIDES BETTER MOTIVATION

As in Sample A, the second method of comparison examined the two testing systems as methods to motivate the students. Here again, the two testing systems were compared in terms of June examination results. The purpose was an attempt to locate the testing system which produced higher June marks.

The same matched pairs were used as for the comparison for predictors. This compared the June results of 1967 of the term tested students with the June results of 1967 of the continual tested students. The same subjects and the same groups were investigated as in Sample A.

No change was attempted in the statistical approach. Consequently, correlation coefficients, t-tests and the null hypotheses were treated similarly to the previous sample study. The results are tabulated in Table XLII, page 109.

Analysis of the Entire Group

The 1967 June results failed to show any significant difference for the term tested group and the continual tested group. The entire group tested by term scored equally well on their June papers as did the entire group tested in the continual manner. This held true for all the four subjects under observation in this study. The correla-

tion coefficients were very low, in fact some were low and negative. On the basis of these findings, both systems are equally good as motivators for the entire group.

Analysis of the Male Group

The calculations on the male group showed no significant differences in the 1967 June results. The term tested students did just as well on their June final papers as their continual tested counterparts. Again, this held true for all the four subjects compared in this study. These findings indicate that the two testing systems are equally good motivators for the male group.

Analysis of the Female Group

Mathematics was the only subject to show a significant difference in the June results for the female groups. Even here the correlation coefficient was quite low, .46. This was the only case where statistical proof existed to substantiate a claim that one testing system was superior to the other as a motivator. In this case, continual testing held a slight edge for the female group. All the other subjects showed no significant difference in the June results. On this basis, both systems were equally good motivators in language, social studies and science for the female groups.

Summary

The statistics accepted the null hypothesis consistently. According to these findings, term testing and continual testing are equally good at motivating students.

CHAPTER VI

A COMPARISON OF SAMPLES A AND B

Samples A and B were subjected to the same statistical treatment. Sample A consisted of students picked over a four year period while Sample B consisted of students taking the same course during the same year. Sample A presented an immediate problem by posing numerous variables over which no control could be exercised. These included different teachers, different term tests as well as continual tests, different June examinations and different school atmosphere. All of these factors were felt to be very important in the student's performance.

During the school year, 1966-67, Sample B was tested in order to remove some of these variables. The term testing and continual testing programs were run simultaneously. This removed some of the previous variables. During this year the same teacher taught mathematics to both of the mathematics classes. The only difference lay in the manner of testing. This was also the case for language, social studies and science. Considering all these controls, it was felt that any difference in achievement either during the year or in June could possibly be contributed to the difference in testing procedure.

Comparison of Sample A and Sample B first term results Term Testing

According to Tables XVII and XXXVIII the first term results for language showed a significant difference between first term results and June results for the entire group and the male group for both samples. Sample B, however, showed a higher correlation coefficient for both the entire group and the male group. Consequently, Sample B would have a higher predictive efficiency. No significant difference was shown for the female group in language in Sample B. Sample A, however, showed a significant difference between the first term marks and the June results. Sample A, therefore, claims to be the better predictor of June results.

Sample A indicated no significant difference between the first term results and the June results in social studies. This was the case for all three groups in this sample. Sample B, on the other hand, showed a significant difference for all three groups. The latter sample also showed correlation coefficients of .78, .80 and .80 for the entire, male and female groups respectively.

Science showed significant differences for similar comparisons for the entire groups and for female groups. The correlation coefficients, and consequently the predictive efficiencies, were higher for Sample A than for

Sample B. The male group in Sample A showed no significant difference, and was therefore unacceptable as a predictor in the study.

Both samples showed significant difference for all groups in mathematics. Sample B showed higher predictive efficiency for the entire group and the male group, while Sample A showed higher predictive efficiency for the female group.

Continual Testing

Tables XVIII and XXXIX showed language to have a significant difference between first term averages and June results for all three groups in continual testing. Sample A showed a negative correlation for all three groups while Sample B showed a positive correlation. Sample B was observed to be a better predictor of June marks for the entire group as well as for the male group. There was really very little difference as far as the female group was concerned, except that Sample A showed a negative correlation while Sample B showed a positive correlation.

Sample A had no significant difference for any of the three groups in social studies. The only group with a significant difference in this comparison was the male group in Sample B. This group had a correlation coefficient of .87 and a predictive efficiency of 68%.

In science, both samples showed no significant difference for the male groups. Neither sample had acceptable predictive value for this study. The entire groups and the female groups, however, showed significant differences. Sample B showed the better correlation coefficient.

All differences were significant in the mathematics comparisons. The correlation coefficients were almost identical for the comparable groups. Sample B held a slight advantage over Sample A in the female groups, while Sample A held a slight edge over Sample B in the entire groups and the male groups.

Comparison of Sample A and Sample B Yearly Results

Term Testing

All groups show significant differences in language when the year's average was compared with the average June mark. Sample A showed .87 and .84 correlation coefficients for the entire group and the male group respectively while Sample B showed correlation coefficients of .78 and .72 for these respective groups. Sample B, however, showed a decided advantage over Sample A in the female groups. Here Sample B produced a correlation coefficient of .93 in comparison to Sample A's .78.

The only group to show a significant difference between the year's average mark and the average June mark in social studies was the male group in Sample B. It, however,

had a correlation coefficient of .92 and consequently a high efficiency of prediction. All the other comparisons had no significant differences. They were, therefore, unacceptable as predictors for this study.

The entire groups and the female groups indicated predictive value in science. The male group of Sample B showed similar value. All groups, except one, showing predictive value had correlation coefficients of .80 or better. The one exception, the entire group of Sample B, had a correlation coefficient of .78.

There were no significant differences in the mathematics comparisons. All of the results could have happened too frequently at random to have predictive value.

Continual Testing

When comparing the results of Tables XX and XLI, language comparisons were again significant for all groups. The entire group and the male group in Sample B, particularly, showed a high correlation coefficient. They were considerably better than those of Sample A. The female group of Sample B also showed a slight improvement over Sample A. All the correlation coefficients were high. The Sample A, male group, was the only one which had a correlation coefficient of under .80.

Sample B had no significant differences in the social studies comparison. All three groups in this sample

had no predictive value for the purpose of this study. The male group in Sample A was in a similar situation. The entire group of Sample A had a correlation coefficient of .72.

All groups in science showed predictive value. Sample B correlation coefficients were higher for all respective groups. The female group of Sample B had a particularly high correlation coefficient, .98. All other coefficients were .79 or better. The statistical evidence claimed Sample B to have the greater predictive value.

Mathematics had only one comparison that showed a significant difference. Even here, in the Sample B male group, the correlation was very low, .24. All other comparisons in mathematics failed to show predictive value for the purpose of this study.

In a general over-view, Samples A and B had almost the same number of comparisons that had predictive value. The predictive efficiency fluctuated between the comparisons of the two samples. The comparison, however, was helpful in substantiating the findings of either study.

Comparison of Samples A and B as motivators

In reviewing Tables XXI and XLII, Sample A claimed a number of significant differences between the June marks scored by term tested students and continual tested students. On the basis of such difference either one of the systems

was proved to be the better motivator. According to Sample A, term testing was the superior motivator in language, social studies and mathematics for the entire groups and the male groups. Sample A claimed that there was no significant difference between the marks scored by term tested students and continual tested students in science for the entire groups and for the male groups. Sample A also claimed neither of the systems illustrated any superiority as a motivator for the female groups in all four subjects--no significant differences in the June results could be established.

Sample B made a sweeping claim that in all cases, except one, the two systems were equal as motivators. Only in mathematics for the female groups, was there found to be a significant difference between the average June marks scored by the term tested students and the continual tested students. In view of the number of variables in Sample A and the much fewer variables in Sample B, it appeared to be quite significant that Sample B should illustrate such consistency of claim.

CHAPTER VII

THE QUESTIONNAIRE

The two testing systems have been compared as predictors and as motivators. Both of these comparisons were made on an objective basis using accepted statistical procedure. Another method of comparison can be carried out. The questionnaire is a method of compiling the opinions of a group of people. These opinions are then analyzed and conclusions are drawn.

In this study, interested parents, students, teachers and inspectors were frequently involved in discussions on the merits of the two testing systems. The two systems were frequently compared in terms of time involved, attitudes fostered, amount of homework required, and other pertinent points of comparison. The questionnaire was designed to provide answers to these questions.

In order to locate meaningless or ambiguous questions, the rough draft of the questionnaire was submitted to a few students, teachers, and parents for a trial run. The doubtful questions were reworded for greater clarity.

Parents frequently provide the atmosphere of education by either backing or opposing school programs. Consequently, the questionnaires were mailed to the parents as well as distributed to the students, teaching staff, and inspectoral staff. Three hundred twenty-five questionnaires

were mailed to the parents, 415 distributed to the students, 21 to the teachers, and 5 to the inspectors of schools. Of these, the parents returned 213, the students 400, the teachers 18, and the inspectors of schools none. The team of inspectors presented a report on the testing systems to the school board.

The responses were analyzed on the basis of the entire group. This group was further divided into parents, students, teachers and inspectors. These groups were re-grouped to check whether the males and females of the separate groups had any difference of opinions. The responses of these interest groups were then used to compare the two testing systems.

Composition of the Questionnaire

The questionnaire was composed with definite sections in mind even though a consecutive numbering system was retained. Questions pertaining to the same section were not always placed consecutively.

Question one forces the respondent to a committal. Questions two to four establish the system providing the better motivation to complete assignments, to review continually, and to use reference material. Either system may provide the better motivation, but the question, whether the student actually follows through, still remains. This follow-up task is the purpose of questions five to seven.

Questions eight to twenty-three deal with study habits, students attitude, psychological situations, and time factors. The consecutively placed questions, eight to nine, deal with study habits. This placement permits a later check on consistency. Ten, thirteen, fifteen and sixteen deal with student attitude towards the examinations and the systems. Again random position makes a check on consistency possible. The remaining questions of this group, except numbers nineteen, twenty, and twenty-three, deal with psychological effects that the examinations have on the students. Placement of questions as well as rephrasing similar questions forces the respondent to reinforce or weaken his opinion. Questions nineteen, twenty, and twenty-three are concerned with the time factor. The second last block of questions, twenty-four to twenty-seven, deals with the opinions on term and yearly averages. Juxtaposition is necessary at this point in order to arrive at a concise conclusion. This block presents one of the major reasons for the choice in question number one. The same holds true for the final block of questions, twenty-eight to thirty. In summary, question one asks for the preference while the other questions probe the reasons for this preference.

Analysis of the Questionnaire

The writer was interested in comparing the two testing systems. In order to do this, all responses that were

not definitely for term or continual testing were relegated to "no opinion". The results were compared on a percentage basis. Tables (List) show the percentage breakdown. The raw count of the entire group as well as of the sub groups may be found in Tables (List) of the appendix.

The first question constitutes section one. The respondent is required to state a preference between the two testing systems. Any opinion other than a clear preference for either testing system is shunted out of contention under "no opinion". Eighty-four per cent of the entire group prefer continual testing while 11 per cent prefer term testing. The three sub groups of parents, students and faculty favour continual testing 82, 86, and 77 per cent respectively.

The preference has been stated in section one. Sections two to seven probe the reasons for this preference. Questions two to four examine the testing systems in terms of their ability to motivate study. All three sub groups again favor continual testing where motivation to finish assignments and to review continually is queried. Question four selects the system providing the better motivation to study from reference other than the text. Forty-two per cent of the entire group registers no preference between term testing or continual testing in question four. All the sub groups register low percentage of choice for either system.

A testing system may provide motivation but application is required in order to complete a task. Section three, composed of questions five to seven, reviews the carry-through on motivation. The sub groups are consistent in choosing the continual testing system as the system which causes a student to review continually. They are also consistent in admitting that they are not sure whether the continual testing system actually causes students to study out of references other than the text.

Questions eight, nine and fourteen constitute section four. This section probes the area of study habits. A favourable rating should score a high percentage in questions eight and nine and score a low percentage in question fourteen. All groups rate continual testing this way. According to all the sub groups, the continual testing system produces better study habits and causes less cramming. Fifty-two per cent of the entire group claim that term testing causes students to become indifferent to study habits. Only 20 per cent claim that this is the case in continual testing, while 28 per cent show no preference between the two systems on this account.

The attitude of the students to the examination system is also very important. Section five, composed of questions ten, thirteen, fifteen and sixteen, explores this phase of testing. Question ten asks for the testing system

producing the better attitude towards examinations. Sixty-two per cent of the entire group favor the continual testing, 23 per cent prefer term testing, and 15 per cent show no preference. Question thirteen counters with whether either testing system gives a student a false sense of security. Only 26 per cent of the entire group claim that continual testing provides a false sense of security while 46 per cent claim this to be a fault of term testing. A similar decision is supported by all the sub groups. Thirty-seven per cent of the entire group, however, claim that continual testing causes students to become indifferent to examinations. Only 28 per cent claim that this is a weakness of term testing. Thirty-five per cent have no decided opinion on which system causes indifference to examinations. All groups rate question sixteen similar to question fifteen. Neither testing system is rated excessively high as conducive to cheating.

Working and testing conditions put pressures on students. It was assumed that these pressures affected the student either advantageously or disadvantageously. Section six examines the psychological effects of the systems. A battery of six questions, three sets of two consecutive questions, constitute this section. Questions eleven and twelve inquire into which system puts a student at ease during examinations and during the day-to-day work. Eighty

per cent of the entire group claim that the continual testing systems sets a student more at ease during examinations. Only 52 per cent claim that continual testing sets students more at ease in their daily work. Thirty-four per cent claim that term testing sets students more at ease during daily work while 14 per cent fall under the "no opinion" category. Question seventeen is a deliberate rephrasing of the content of questions eleven and twelve, continual testing should have a low percentage rating while the rating for term testing should be high. The actual figures show the entire group rating continual testing and term testing as 21 per cent and 65 per cent respectively. The respondents are consistent in their responses. Question eighteen poses the problem whether either testing system aids the weak student. Eighty-six per cent of the entire group claims that the continual testing system aids the weak student. Questions twenty-one and twenty-two question the emphasis placed on testing and on marks. Low percentages of 41 per cent and 18 per cent are registered for continual testing to questions twenty-one and twenty-two respectively. Thirty per cent claim that term testing places too much emphasis on testing while 64 per cent claim that term testing puts too much emphasis on marks. These percentages included the entire group.

Section seven, made up of questions nineteen, twenty, and twenty-three, probes the time factor. The entire group breaks up with 38 per cent checking continual testing and 25 per cent checking term testing as requiring too much homework. The two systems are rated fairly evenly in this category, neither receiving a high rating. A similar rating is given to question number twenty. Consequently the entire group does not feel that too much time is lost in testing in either system. Neither is too much time spent in checking examination results.

The eighth section deals with the opinions on term and June results. Questions twenty-four and twenty-six respectively ask for opinions on which system produces better term and June results. In both cases the entire group favors the continual testing system. Seventy-five per cent of the entire group backed continual testing for better term results and sixty-three per cent of the entire group favoured continual testing for better June results. Questions twenty-five and twenty-seven probe whether either system produces meaningless term or yearly averages. The entire group claims that the results produced by continual testing is not meaningless. Only 15 per cent claim that the continual testing produces meaningless term results. Eighteen per cent of the entire group claim that meaningless yearly averages are produced by continual testing.

Claims are slightly higher for term testing. Thirty-eight per cent of the entire group feel that term results produced by term testing are meaningless. Forty per cent of the entire group claim that yearly averages produced by term testing are meaningless. On this basis, the entire group favours continual testing. This is corroborated by the sub groups.

Parents, teachers, and students want to know how the students are getting along. The last section, questions twenty-eight to thirty, opens this problem. The entire group endorses continual testing for this section. Ninety, eighty-seven, and eighty-nine per cent of the entire group favour questions twenty-eight, twenty-nine and thirty respectively. Parents, teachers, and students feel that continual testing gives them a better idea of how the student is getting along.

Originally the entire group was to be sub-divided into groups of parents, teachers and students. This was for analytical purposes. The separate groups were to be further divided into male and female groups. A cursory scanning of the percentage scores of these various sub groups reveals why further analysis is unnecessary. The sub groups agree so fully that there is little purpose in carrying the analysis any further. This observation also holds true for the male and female groups.

The entire group, as well as the numerous sub groups, favor continual testing. This analysis is backed in every question. No returns were received from the inspection staff, but they endorsed continual testing in their report to the school board.

CHAPTER VIII

SUMMARY AND CONCLUSIONS

Summary

The purpose of this study was to compare two testing systems used in Grade IX; term testing and continual testing. This study attempted to compare the two testing systems as predictors of June marks. It was reasoned that if school marks could predict June marks within a reasonable margin, June examinations would be unnecessary. To this end both the first term average marks and the year's average marks were compared with the average June marks. This was done for the four core subjects in Grade IX--literature, social studies, science, and mathematics. The students in the experiment were compared as entire groups, male groups, and female groups.

The two testing systems were also compared as methods of motivating students. It was reasoned that, all other conditions being equal, the group scoring the higher average June mark could possibly have been motivated by the testing program. With this in mind, the average June marks scored by the term-tested students were compared with the average June marks scored by the continual-tested students. This was again carried out for all of the four core subjects and for the entire group, which was again sub-divided into the male group and the female group.

A final comparison of the two testing systems was made through the use of a questionnaire. Students, teachers, and parents are important components of any teaching-learning process. It is, therefore, important how they react to a program. The main purpose of this questionnaire was to establish the preference of these key people as well as reasons for this preference.

Sample A supplied data for this study from September, 1962 to June, 1966. The students of this sample were term-tested from September, 1962, to June, 1964. This group consisted of 70 students, 30 boys and 40 girls. Their mean scores for the first term and for the year were compared with their mean scores obtained in the June Department of Education Examinations.

Another 70 students supplied the data from September, 1964, to June, 1966. These students were continual-tested. Their first term and yearly mean scores were also compared to their mean scores which they obtained in the June Department of Education Examinations.

The second group of 70 students were paired with the first group of 70 students. They were paired by sex, age, and intelligence quotient. Their intelligence quotients were obtained by Otis Quick Scoring Intelligence Tests.

Sample B used the same approach as Sample A. This time the students were tested simultaneously. Twenty-seven

students were paired with a second 27 students. They were paired in a similar manner as those of Sample A. Sample B permitted more controlled variables. In the second study, teachers and tests became controlled variables. These groups of students provided data from September, 1966, to June, 1967.

In all cases the first term and yearly mean scores were correlated with the mean scores obtained in the June Department of Education Examinations. These correlation coefficients were then checked to confirm whether the differences in the means were significant. Significant differences of 5% or lower were accepted in this study. The t-test method was used to obtain the significant difference percentage levels. Prediction and motivation comparisons were made based on the correlation coefficient and t-test results.

A COMPARISON OF TERM-TESTING AND CONTINUAL TESTING OF SAMPLE A AND B AS PREDICTORS

In drawing conclusions concerning the prediction of June marks, it should be noted that even a high correlation has only modest predictive value. A correlation of .80, for instance has a predictive ability of 40%--40% improvement over a pure chance guess.[1] This limitation qualifies all claims of being a better predictor of June marks. Such claims are made soley on the raw r value substantiated by the t-test for confidence level.

First Term Results, Sample A

1. Language had predictive value for establishing June results. This was the case for both term and continual testing. Continual testing, however, had a negative correlation coefficient while that of term testing was positive. This held true for the entire group as well as for the two sub-groups. On the basis of a mathematical difference, the term testing system proved to be a slightly better predictor of June results than the continual testing system.

2. Social studies produced no significant differences between the first term means and the June means. This held true for all three groups compared. Consequently neither

---

[1] Best, op. cit., p. 240 - 241

testing system proved of value as a predictor of June results.

3. Science provided prediction statistics for the entire group as well as for the female group. For both groups, term testing proved to be the better predictor of June results. There was no predictive value in either testing system for the male group.

4. Mathematics provided predictive results in all three groups of both testing systems. Continual testing proved to be the better predictor of June results for the entire group and the female group. Term testing provided superior prediction for the male group.

Yearly Results, Sample A

1. Language showed predictive value for all three groups compared. The term testing system proved the superior predictor of June results for the entire group and the male group. Continual testing was the better predictor for the female group.

2. Term testing showed no significant difference between yearly mean scores and mean June scores for all groups in social studies. Continual testing showed predictive results for the entire group and the female group. Continual testing was, therefore, the superior predictor of June results for these two groups. Continual testing failed to have predictive value for the male group.

3. Both term testing and continual testing showed significant difference in the means in the entire group and in the female group in science. Only continual testing showed predictive value for the male group. Term testing proved the superior predictor of June marks for the entire and female groups. Continual testing proved the superior predictor for the male group.

4. Term testing and continual testing could not be compared as predictors of June marks. No significant differences could be established between yearly means and June means for either system.

First Term Results, Sample B

1. Language correlation coefficients had significance at 1% levels of confidence for the entire groups and the male groups in both testing systems. Term testing had a value of r of .82 but the level of confidence was between 40% and 50% while continual testing had a value of r of .63 at the 1-2% level of confidence. Consequently the continual testing system is the superior predictor of June results for the entire group and the female group. Term testing, however, is the superior predictor for the male group.

2. Term testing had r values of .78, .80, and .80 for the entire group, the male group and the female group respectively in social studies. The level of confidence was at 1% for the entire group and the male group. The level

of confidence was at 5% for the female group. Continual testing indicated no significant differences between the means for the entire group and the male group. It showed an r value of .87 at the 1-2% level of confidence for the female group. Term testing, therefore, is the superior predictor of June marks for the entire group and the male group in social studies. Continual testing had superior predictive qualities for the female group.

3. In science, the male group in continual testing was the only group to have a value of r exceeding the 5% level of confidence. All the other groups showed significant differences at the 1% level of confidence. Continual testing had a slightly higher r value than term testing for the entire group, .79 and .64 respectively. Continual testing was the better predictor of June results for the entire group and female group in science. Term testing was the better predictor for the male group, although it had an r value of only .63. For the female group, continual testing was superior to term testing with respective r values of .80 and .72.

4. All of the r values were significant at the 1% level of confidence. Continual testing was the better predictor of June results for the entire groups as well as the

female groups. Term testing was the superior predictor for the male group.

Yearly Results, Sample B

1. In language, the r values were significant at the 1% level of confidence for both terms. The entire group and the male group in continual testing had exceptionally high values of r, .92 and .95 respectively. For these two groups, continual testing was the superior predictor of June results. Even though continual testing had a high r value for the female group, .87, the term tested group's r value was still higher, .93. Term testing was the superior predictor of June results for the female groups.

2. The social study results showed only one comparison significant at the 1% level of confidence. This exception was the entire group in term testing. It had an r value of .92. Term testing was therefore, the superior predictor of June results for the entire group. Since the results for the other groups could have occurred too frequently by chance fluctuation, no prediction comparison was possible on the basis of this study.

3. All of the comparisons in science showed significant differences at the 1% level of confidence. In this case, continual testing proved the superior system as predictor of June marks for all groups. The r value for the female group in continual testing was an impressive .98.

The entire group and male group had r values of .88 and .89 respectively.

4. There were no significant differences for mathematics in the term testing system. All their test values were zero. Continual testing, likewise, showed no significant difference in the comparisons for the entire group and the female group. The male group r value was significant at the 5% level. The r value, however, was only .24. Consequently there were no predictive comparisons in mathematics for the entire group and the female group. Continual testing was the superior predictor of June results for the male group.

Analysis of Tables XLIII and XLIV

The first term results were consistent in having predictive value. Social Studies, was the consistent exception. The results of Sample A and B indicated that first term results have value as predictors of June marks for the entire groups in language, science and mathematics. The Social Study results were too consistently significant at too high a confidence level to indicate predictive value.

On the yearly results, Social Study predictor values were too inconsistent to make even limited claims. Mathematics indicated no predictive value at all. Language and science, however, were consistent in showing predictive value.

COMPARISON OF THE RESULTS OF SAMPLES A AND B AS PREDICTORS

TABLE XLIII

SAMPLE A

ENTIRE GROUP

FIRST TERM RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.70	1%	29%	8	-.67	1%	26%	7
Social Studies	null hypothesis accepted							
Science	.74	1%	33%	10	.60	1%	20%	10
Mathematics	.71	1%	29%	13	.81	1%	41%	9

YEARLY AVERAGE RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.87	1%	51%	6	.80	1%	40%	6
Social St.	null hypothesis accepted				.72	1%	31%	10
Science	.83	1%	45%	8	.80	1%	40%	7
Mathematics	null hypothesis accepted							

Legend

R----------correlation coefficient
C.L.-------Confidence level
EFF.-------prediction efficiency
E.E.-------estimate of error

TABLE XLIV

SAMPLE B

ENTIRE GROUP

FIRST TERM RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.87	1%	51%	5	.89	1%	54%	6
Social St.	.78	1%	36%	10	null hypothesis accepted			
Science	.64	1%	24%	13	.79	1%	38%	9
Mathematics	.74	1%	32%	7	.78	1%	36%	7

YEARLY AVERAGE RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.78	1%	37%	7	.92	1%	61%	5
Social St.	.92	2%	61%	7	null hypothesis accepted			
Science	.78	1%	37%	8	.88	1%	53%	6
Mathematics	null hypothesis accepted							

Language and science consistently showed predictive value for both samples, for both systems and for both first term and yearly averages. This consistency indicated that these findings have value for the prediction of June results. The first term results of mathematics were also consistent for both samples and both systems. Consequently, first term averages in mathematics have value as predictors of June results.

Analysis of Tables XLV and XLVI

The first term results indicated predictive value in language and mathematics. Both samples agreed that this was the case for both testing systems. Sample B also showed predictive value for social studies and science in the term testing system.

When the yearly averages were considered, only language proved to be a consistent predictor in both systems. The consistency held through both samples. Continual testing of sample B showed that science also had predictive value.

Analysis of Tables XLVII and XLVIII

The first term results in both systems had predictive values for science and mathematics. Predictive consistency held through Samples A and B. Sample B, continual testing also claimed predictive value for language and social studies. Term testing of Sample B claimed predictive

TABLE XLV

SAMPLE A

MALE GROUP

FIRST TERM RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.71	5%	29%	8	-.57	1%	18%	7
Social St.	null hypothesis accepted							
Science	null hypothesis accepted							
Mathematics	.83	5%	45%	10	.79	1%	39%	8

YEARLY AVERAGE RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.84	5%	46%	7	.73	1%	32%	6
Social St.	null hypothesis accepted							
Science	null hypothesis accepted				.79	1%	39%	6
Mathematics	null hypothesis accepted							

TABLE XLVI

SAMPLE B

MALE GROUP

FIRST TERM RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.88	1%	53%	6	.86	1%	49%	6
Social St.	.79	1%	38%	13	null hypothesis accepted			
Science	.64	1%	23%	11	null hypothesis accepted			
Mathematics	.84	1%	46%	7	.76	1%	35%	8

YEARLY AVERAGE RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.72	1%	31%	10	.95	1%	69%	4
Social St.	null hypothesis accepted							
Science	.80	1%	40%	7	.89	2%	54%	7
Mathematics	null hypothesis accep..				.24	5%	8%	16

TABLE XLVII

SAMPLE A

FEMALE GROUP

FIRST TERM RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.70	1%	29%	8	-.62	1%	22%	8
Social St.	null hypothesis accepted							
Science	.77	1%	37%	9	.61	1%	21%	10
Mathematics	.62	1%	22%	14	.83	1%	45%	9

YEARLY AVERAGE RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.78	5%	37%	7	.81	1%	41%	4
Social St.	null hypothesis accepted				.75	1%	34%	10
Science	.84	1%	46%	7	.79	1%	39%	6
Mathematics	null hypothesis accepted							

TABLE XLVIII

SAMPLE B

FEMALE GROUP

FIRST TERM RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	null hypothesis accepted				.62	1-2%	22%	9
Social St.	.80	5%	40%	9	.87	1-2%	51%	5
Science	.72	1%	31%	13	.80	1%	40%	7
Mathematics	.39	1%	7%	7	.87	1%	51%	4

YEARLY AVERAGE RESULTS AS PREDICTORS OF JUNE MARKS

	Term Testing				Continual Testing			
	R.	C.L.	EFF.	E.E.	R.	C.L.	EFF.	E.E.
Language	.93	1%	63%	3	.87	1%	51%	5
Social St.	null hypothesis accepted							
Science	.93	1%	63%	5	.98	1%	80%	2
Mathematics	null hypothesis accepted							

value for social studies as well.

The yearly averages of language and science proved to have predictive value for language and science. Both samples indicated consistency in this case. The null hypothesis was accepted in all cases in mathematics and social studies except in continual testing, Sample A, where social studies indicated a predictive value of 34%.

The outstanding exception seemed to be Sample B science. Both continual and term testing showed excellent predictive value. Continual testing, especially, showed high predictive efficiency of 80% with a + or - 2 variation.

Conclusions

Predictive Results - First Term

Language

Tables XLIII to XLVIII indicated that first term results in language had predictive value for June results. Such value held true for all three groups. There was only one exception to this trend, the Sample B female group in term testing. The prediction efficiency for the entire group was not conclusive enough to claim that either testing system was superior. Term testing, however, consistently proved to be the better predictor for the male group. Continual testing was quite consistently the better predictor for the female group, especially in Sample B.

Social Studies

The consistency of social studies indicated that first term averages had little or no predictive value for the entire group and the male group. Both Samples A and B agreed on this point for both testing systems. Sample B, however, claimed that term testing had a prediction efficiency of 36% for the entire group and 38% for the male group. Sample A showed that neither testing system had predictive value for the female group. Sample B, however, claimed continual testing to be superior to term testing for the female group, 40% and 51% respectively.

Science

Science scores were somewhat inconsistent as a predictor. First term results proved to have predictive value for the entire group and the female group. Inconsistencies in Sample A and B, however, would indicate that the two testing systems had equal proficiency as predictors for these two groups. Both samples claimed science first term results had little predictive value for the male group. The only exception was Sample B, term testing, which claimed a prediction efficiency of 23% for the male group.

Mathematics

First term results of mathematics proved to have predictive value for all three groups. This consistency was maintained for both testing systems in both samples. Con-

tinual testing was consistently the better predictor for the entire group and the female group. Term testing was the superior predictor for the male group. Both samples agreed on these points.

Prediction Results - Yearly Averages

Language

The yearly averages indicated predictive value for June results in language. The consistency held true for all the groups, in both samples, for both testing systems. Fluctuations in prediction proficiency between the two systems, however, would indicate that they were equal in this respect.

Social Studies

The null hypothesis was accepted in practically all cases in social studies. The yearly marks in social studies had little or no value as predictors of June results. This held true for all the groups in both samples. The only exceptions to this statement arose for the entire group where Sample A claimed 31% for continual testing and Sample B claimed 61% for term testing. Sample B also claimed 34% for continual testing in the female group.

Science

Yearly averages had predictive value for the entire group and the female group. Sample B also claimed predictive value for the male group. Neither testing system

showed consistently superior proficiency for the entire group and the female group. Sample B, continual testing, proved to be an exception. It claimed an 80% predictive efficiency, with an estimated error of + or - 2, for the female group. Continual testing proved the superior predictor for the male group.

Mathematics

Yearly averages in Mathematics proved of little value as predictors of June results. The null hypothesis was accepted consistently. The only exception, the Sample B continual testing male group, was noted. This sample claimed an 8% prediction with an estimated error of + or - 16 marks.

If the developed regression equations were applied and the estimate of error were taken into consideration, a percentage of the predicted variable for term testing and continual testing would frequently overlap. For example first term language, Sample A term testing, had an efficiency of prediction value of 29% with an estimated error of + or - 8. First term language, Sample A, continual testing, had an efficiency of prediction value of 26% with an estimated error of + or - 7. If a hypothetical case of 75% were chosen as the first term mark for student A, the predicted

mark for this student would be:-

$$X'=(.70)(1.03)(75-68)+63=68 \text{ in June.}$$

With the + or - 8 mark variable, the actual mark should lie between 60 and 76 per cent. In the continual testing situation the student's mark would be:-

$$X'=(-.67)(1.14)(75-72)+58=56 \text{ in June.}$$

With the + or - 7 mark variable, the actual mark should lie between 49 and 62 per cent.

In this case, the two extremes, the lower range of the term tested prediction overlapped with the higher range of the continual tested prediction.

In some cases, a given range of prediction for one system of testing could completely overlap the range of prediction of the second testing system. Such a situation existed when first term mathematics, Sample A, was investigated. There is, however, a mathematical difference as previously claimed in comparing individual cases on their raw r values. An r value of .80 is higher than that of .78 if at an acceptable level of confidence.

Motivation Results

According to Table XLII Sample A accepted term testing as providing superior motivation in language, social studies and mathematics for the entire group and the male group. Sample B accepted the null hypothesis, that no difference existed, for all the subjects for the entire group

and the male group. Both samples agreed that the two systems provided equally good motivation for the female group. Sample B made it possible to control many of the variables that could not be controlled in Sample A. On the basis of these findings, the two testing systems provided equally good motivation to students in the four Grade IX core subjects.

Applications of Results

According to these results, first term results in language and mathematics should be of value in predicting June results. First term science marks should have predictive value for the entire groups and female group. Such predictions could aid the Guidance Director in assisting Grade IX students to make proper Grade X course selections. If the results were favourable, the student could be advised to select the University Entrance Course. If the results showed that a student probably would have difficulty in a Grade X University Entrance Course, the student could be counselled to take the General Course, the Commercial Course or any one of the technical or industrial courses.

Meanwhile it would be possible to make more realistic reports to the parents on their child's performance. Teachers would have some substantiating evidence for their suggestions during parent-teacher consultations.

All persons concerned would, however, have to be aware of the limitations of such predictions. Students, for instance, would have to realize that they would have to continue a calibre of work on par with their first term work. Students, teachers, and parents would have to realize that the statistical analysis was based on a group effort. Any future predictions would be more applicable to a group than to any individual in that group.[2] An individual in a group could always prove to be the exception to that group.[3]

Only language made a consistent claim for predictive value based on the year's average. Science made similar claims for the entire group and the female group. Sample B in science also made this claim for the male group.

The June Department of Education Examinations have been accepted as the basis of comparison throughout this study. The results of these examinations were accepted as being a reliable measurement of a student's achievement. In the case of these two subjects, this study would substantiate the suggestion that it would be reasonable to drop the June Department of Education Examinations. The statistics of this study would not substantiate such suggestion for social studies and mathematics.

---

[2] Ibid. [3] Ibid.

Neither of the two testing systems proved to be superior in providing motivation. No real difference could be established between the average June mark of the term-tested students and that of the continual-tested students. It would, therefore, seem logical to provide that type of a testing program which was more acceptable to the people involved. In the locale of this study, continual testing was accepted by approximately 85% of all concerned. In such a case, the continual testing program should be employed.

It is hoped that further research will be carried on in a search for better predictors. Valid predictors will have to be developed in order to assist an increasing number of students to select appropriate courses of study.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bauernfeind, R. H. Building a School Testing Program. Boston: Houghton Mifflin Company, 1963.

Best, J. W. Research in Education. Englewood Cliffs, N. J., Prentice-Hall, Inc., 1965.

Bou, I. R. and Stovall, F. L. "Relationship Between High School Grades and First Year Achievement in the University of Puerto Rico" Journal of Educational Psychology, May 1950, Vol. 41, pp. 309 - 320.

Brereton, Joseph L. Examinations, Where Next? Pacific Northwest Humanist Publications, 1965.

Cronbach, L. J. Educational Psychology. New York: Harcourt, Brace, and World, Inc., 1963.

Dressel, P. L. et al. Evaluation in Higher Education. Boston: Houghton Mifflin Company, 1961.

Fisher and Yates. Statistical Tables for Biological, Agricultural, and Medical Research. Edinburgh: Oliver and Boyd, Ltd.

Flemming, W. G. "Factors Predicting Accuracy of Ontario Grade XIII Results" Bulletin No. 16, Department of Educational Research, University of Toronto, Toronto, 1955, p. 3

Linquist, E. F. A First Course in Statistics. Boston: Houghton Mifflin Company, 1942.

Mode, Elmer B. Elements of Statistics. Englewood Cliffs, N. J. :Prentice-Hall, Inc., 1961.

Nunnally, J. C. Educational Measurement and Evaluation. Toronto: McGraw-Hill Book Company, 1964.

Rothney, John W. M. "Evaluating and Reporting Pupil Progress." Department of Classroom Teachers, American Educational Research Association of the National Educational Association. May, 1960.

Schmitz, S. B. "Predicting Success in College: A Study of Various Criteria" Journal of Educational Psychology, September, 1937, Vol. 28, pp. 465 - 473.

Smith, Francis F. "The Use of Previous Records in Estimating College Success", Journal of Education Psychology, March 1945, Vol. 36.

Tyler, L. E. Tests and Measurements. Englewood Cliffs, N. J., Prentice-Hall, Inc., 1965.

UNPUBLISHED MATERIALS

Friesen, D. "The Differential Aptitude Tests as Predictors in Education I at the University of Manitoba", Unpublished Master's Thesis, The University of Manitoba, Winnipeg, 1958.

McIntyre, S. B. "A Comparative Study of Collegiate and Department of Education Marks in Manitoba." Unpublished Master's Thesis, The University of Manitoba, Winnipeg, 1952.

Pollock, H. "High School Averages and Supplementals as Predictors of First Year University Success", Unpublished Master's Thesis, The University of Manitoba, Winnipeg, 1959.

Schusler, Marian M. "Prediction of Grades by Computer for High School Students: A Cross Validation and Experimental Placement Study", Unpublished Doctoral Dissertation, University of Pittsburgh, Pennsylvania, 1964.

APPENDIX A

Raw Scores in First Term Results,
in Yearly Averages and June Marks
as well as
Pairing Data
and
Questionnaire Results

TABLE XLIX

PAIRING DATA IN TERMS OF SEX, AGE, AND INTELLIGENCE QUOTIENT

TERM TESTING, 1962 - 1963 CONTINUAL TESTING, 1964 - 1965

SAMPLE A

	MALE			MALE	
	AGE	I. Q.		AGE	I. Q.
	1962 - 1963			1964 - 1965	
1.	15-5	122		15-2	125
2.	15-2	102		15-1	104
3.	15-2	102		15-5	104
4.	15-2	107		15-2	110
5.	15-3	110		15-1	112
6.	14-6	128		14-9	129
7.	14-9	125		14-8	122
8.	14-10	108		14-7	109
9.	14-8	102		14-9	105
10.	14-9	101		14-8	104
11.	15-7	102		15-9	105
12.	15-1	114		15-2	115
13.	14-9	114		14-9	114
14.	15-1	124		14-11	122
15.	15-5	102		15-3	104
	FEMALE			FEMALE	
16.	14-10	98		15-0	100
17.	16-4	96		16-1	99
18.	15-4	106		15-1	104
19.	14-10	113		14-10	116
20.	15-10	95		15-7	98
21.	15-3	101		15-4	104
22.	16-9	100		16-10	98
23.	15-4	125		15-3	122
24.	15-2	95		15-4	92
25.	15-0	107		15-0	110
26.	15-1	125		15-1	122
27.	15-9	91		15-6	91
28.	15-5	113		15-6	110
29.	15-10	119		15-7	116
30.	15-3	109		15-5	108
31.	15-5	110		15-3	107
32.	15-6	120		15-3	123
33.	14-9	113		14-10	115
34.	14-10	111		14-8	108
35.	14-4	112		14-7	113

TABLE L

PAIRING DATA IN TERMS OF
SEX, AGE, AND INTELLIGENCE QUOTIENT

TERM TESTING, 1963 - 1964 CONTINUAL TESTING, 1965 - 1966

SAMPLE A

	MALE			MALE	
	AGE	I. Q.		AGE	I. Q.
	1963 - 1964			1965 - 1966	
1.	15-3	109		15-3	109
2.	14-8	116		14-10	113
3.	14-7	109		14-10	112
4.	15-1	111		15-1	109
5.	15-3	99		15-5	100
6.	16-8	89		16-5	88
7.	15-2	117		15-5	119
8.	16-2	98		16-5	98
9.	14-3	127		14-6	124
10.	16-4	97		16-3	100
11.	16-5	111		16-5	111
12.	15-9	102		15-11	100
13.	15-0	116		15-0	113
14.	15-10	103		15-11	104
15.	14-10	97		15-1	100
	FEMALE			FEMALE	
16.	14-9	112		14-9	115
17.	14-10	116		14-7	119
18.	15-3	124		15-5	127
19.	14-8	106		14-11	109
20.	14-7	108		14-9	111
21.	14-8	111		14-10	114
22.	15-2	105		15-0	105
23.	14-10	113		14-7	114
24.	14-7	108		14-10	110
25.	16-0	104		16-2	101
26.	14-3	98		14-0	101
27.	15-7	113		15-4	115
28.	14-8	102		14-10	99
29.	14-10	114		14-11	111
30.	16-2	99		16-1	98
31.	14-9	111		14-6	113
32.	15-4	102		15-5	100
33.	15-10	95		15-7	96
34.	14-9	113		14-11	115
35.	15-5	107		15-4	106

TABLE LI

PAIRING DATA IN TERMS OF
SEX, AGE, AND INTELLIGENCE QUOTIENT

TERM TESTING, 1966 - 1967 CONTINUAL TESTING, 1966 - 1967

SAMPLE B

	MALE			MALE	
	AGE	I. Q.		AGE	I. Q.
	1966 - 1967			1966 - 1967	
1.	14-7	114		14-7	115
2.	14-5	117		14-5	116
3.	14-6	117		14-3	116
4.	14-0	117		14-2	119
5.	15-5	97		15-6	97
6.	15-3	95		15-0	92
7.	14-3	120		14-1	121
8.	15-1	101		15-2	99
9.	13-11	108		14-0	108
10.	13-9	119		13-9	116
11.	15-0	99		14-11	102
12.	14-4	101		14-3	104
13.	14-11	99		14-9	99
14.	14-0	108		14-1	109
	FEMALE			FEMALE	
15.	14-8	113		14-5	114
16.	14-1	121		14-3	121
17.	14-6	106		14-6	107
18.	13-10	112		14-0	112
19.	14-9	101		14-9	99
20.	14-5	116		14-6	113
21.	15-4	96		15-1	98
22.	14-1	104		14-1	104
23.	13-9	109		14-0	111
24.	14-2	111		14-3	113
25.	14-2	104		14-1	101
26.	14-1	107		14-1	109
27.	14-5	106		14-4	103

TABLE LII

LANGUAGE

RAW SCORES OF THE TESTING PROGRAM OF 1962 - 1963

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	66	64	65
2.	41	41	48
3.	57	56	55
4.	72	70	67
5.	46	48	56
6.	71	64	75
7.	72	73	74
8.	58	55	60
9.	57	57	48
10.	56	51	62
11.	63	66	69
12.	72	73	69
13.	70	66	63
14.	84	79	68
15.	47	45	53

FEMALE

16.	67	57	55
17.	51	46	48
18.	56	53	53
19.	56	52	53
20.	52	47	48
21.	60	58	64
22.	56	55	65
23.	78	74	82
24.	59	57	65
25.	59	60	58
26.	80	78	77
27.	72	67	77
28.	80	85	81
29.	86	77	80
30.	54	56	63
31.	72	69	63
32.	89	88	77
33.	64	65	58
34.	64	56	53
35.	80	71	65

TABLE LIII

LANGUAGE

RAW SCORES OF THE TESTING PROGRAM OF 1963 - 1964

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	63	58	48
2.	67	75	64
3.	53	52	44
4.	63	62	55
5.	63	69	60
6.	64	48	38
7.	80	83	76
8.	70	62	48
9.	86	86	78
10.	50	41	38
11.	86	87	80
12.	66	71	63
13.	69	80	83
14.	55	55	40
15.	53	47	44

FEMALE

16.	63	72	48
17.	84	85	82
18.	84	83	79
19.	80	71	55
20.	67	60	48
21.	87	84	71
22.	67	64	57
23.	79	82	75
24.	78	74	73
25.	72	73	72
26.	61	57	67
27.	81	79	72
28.	78	80	58
29.	78	85	77
30.	67	69	69
31.	89	85	78
32.	67	61	50
33.	73	65	58
34.	72	66	59
35.	79	73	67

TABLE LIV

LANGUAGE

RAW SCORES OF THE TESTING PROGRAM OF 1964 - 1965

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	67	72	66
2.	67	67	55
3.	72	62	40
4.	62	62	49
5.	62	57	51
6.	92	87	81
7.	67	67	68
8.	67	62	54
9.	72	67	57
10.	57	57	45
11.	62	62	44
12.	72	72	57
13.	77	77	66
14.	77	77	71
15.	62	57	49

FEMALE

16.	52	52	44
17.	72	62	45
18.	77	82	71
19.	82	87	80
20.	62	67	57
21.	67	67	60
22.	72	67	55
23.	87	92	86
24.	72	67	44
25.	72	72	50
26.	72	77	65
27.	72	77	65
28.	72	67	60
29.	77	77	60
30.	57	57	48
31.	82	77	76
32.	77	77	58
33.	77	77	68
34.	82	82	69
35.	77	82	75

TABLE LV

RAW SCORES OF THE TESTING PROGRAM OF 1965 - 1966

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	82	70	46
2.	62	67	53
3.	59	58	33
4.	61	57	57
5.	75	70	44
6.	69	60	47
7.	79	76	61
8.	78	73	61
9.	75	72	63
10.	71	67	61
11.	58	57	37
12.	81	75	47
13.	77	80	65
14.	57	56	48
15.	61	57	51

FEMALE

16.	64	64	52
17.	87	83	58
18.	94	92	75
19.	70	64	58
20.	74	74	67
21.	88	80	75
22.	56	58	46
23.	69	64	63
24.	85	79	67
25.	90	76	52
26.	58	55	46
27.	86	82	64
28.	55	56	54
29.	80	83	77
30.	58	55	44
31.	86	86	63
32.	77	75	47
33.	79	75	63
34.	82	77	58
35.	87	82	79

TABLE LVI

LANGUAGE
TERM TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	57	55	48
2.	83	88	69
3.	77	77	68
4.	87	83	73
5.	53	N	N
6.	43	46	40
7.	71	67	72
8.	63	66	59
9.	62	59	42
10.	72	95	58
11.	71	71	58
12.	47	53	47
13.	81	70	66
14.	63	58	53

FEMALE

15.	82	70	60
16.	78	86	75
17.	69	70	58
18.	71	68	58
19.	77	74	66
20.	73	75	61
21.	58	55	43
22.	59	60	54
23.	78	73	67
24.	83	79	75
25.	63	63	51
26.	68	71	63
27.	65	63	46

TABLE LVII

LANGUAGE
CONTINUAL TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	90	88	84
2.	84	82	71
3.	62	60	58
4.	79	80	69
5.	52	N	N
6.	55	47	25
7.	64	63	58
8.	58	50	37
9.	57	59	54
10.	78	73	58
11.	46	46	39
12.	63	63	58
13.	49	52	44
14.	81	78	67

FEMALE

15.	63	70	61
16.	75	75	72
17.	51	54	55
18.	81	77	65
19.	67	71	65
20.	85	90	76
21.	70	63	57
22.	55	62	54
23.	56	59	50
24.	75	75	63
25.	79	78	72
26.	85	79	57
27.	74	79	81

TABLE LVIII

SOCIAL STUDIES

RAW SCORES OF THE TESTING PROGRAM OF 1962 - 1963

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	75	65	60
2.	58	49	32
3.	60	56	57
4.	73	69	73
5.	69	57	58
6.	79	75	82
7.	77	76	80
8.	81	73	80
9.	56	54	79
10.	54	46	39
11.	78	64	60
12.	75	70	75
13.	72	69	66
14.	87	84	84
15.	74	66	59

FEMALE

16.	78	57	48
17.	65	55	39
18.	30	35	19
19.	41	46	43
20.	38	33	25
21.	59	58	70
22.	79	70	65
23.	86	87	82
24.	48	33	31
25.	54	48	44
26.	71	64	58
27.	52	52	58
28.	89	88	89
29.	88	81	89
30.	59	52	62
31.	78	70	67
32.	93	89	85
33.	65	68	61
34.	63	59	51
35.	83	85	81

TABLE LIX

SOCIAL STUDIES

RAW SCORES OF THE TESTING PROGRAM OF 1963 - 1964

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	39	41	35
2.	51	55	56
3.	18	34	22
4.	24	32	32
5.	56	63	65
6.	16	20	21
7.	75	83	90
8.	56	60	51
9.	90	88	87
10.	22	29	27
11.	75	83	76
12.	55	60	56
13.	46	64	78
14.	39	44	44
15.	59	63	58

FEMALE

16.	56	61	63
17.	78	83	83
18.	81	83	86
19.	45	48	40
20.	48	49	47
21.	76	74	67
22.	36	42	24
23.	62	71	67
24.	70	63	61
25.	60	61	57
26.	33	37	44
27.	68	62	57
28.	56	66	67
29.	86	90	84
30.	53	68	70
31.	76	80	80
32.	27	27	31
33.	51	57	53
34.	30	44	68
35.	59	60	45

TABLE LX

SOCIAL STUDIES

RAW SCORES OF THE TESTING PROGRAM OF 1964 - 1965

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	52	52	65
2.	45	38	58
3.	30	40	35
4.	52	57	56
5.	35	30	42
6.	87	84	77
7.	45	46	66
8.	52	57	62
9.	45	52	51
10.	45	52	65
11.	20	39	24
12.	52	67	39
13.	67	72	75
14.	57	63	60
15.	42	51	50

FEMALE

16.	20	30	12
17.	45	42	63
18.	62	57	59
19.	72	82	89
20.	45	52	54
21.	72	67	56
22.	62	62	50
23.	82	77	84
24.	45	40	54
25.	20	30	13
26.	45	57	57
27.	45	52	53
28.	35	40	36
29.	52	57	55
30.	25	30	31
31.	57	57	62
32.	45	52	65
33.	72	72	74
34.	57	57	46
35.	67	74	77

TABLE LXI

SOCIAL STUDIES

RAW SCORES OF THE TESTING PROGRAM OF 1965 - 1966

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	43	46	10
2.	53	50	41
3.	52	55	55
4.	67	66	62
5.	62	70	68
6.	60	61	49
7.	64	66	63
8.	48	54	40
9.	69	71	64
10.	52	56	47
11.	42	50	48
12.	62	63	60
13.	65	62	58
14.	40	45	34
15.	38	36	26

FEMALE

16.	63	62	62
17.	64	62	59
18.	84	86	73
19.	55	52	41
20.	75	78	70
21.	75	75	65
22.	55	59	40
23.	65	65	63
24.	57	52	39
25.	62	67	64
26.	46	45	23
27.	52	51	52
28.	55	58	39
29.	85	86	71
30.	25	32	29
31.	72	72	59
32.	57	62	45
33.	60	58	55
34.	72	73	60
35.	77	79	69

TABLE LXII

SOCIAL STUDIES

TERM TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	34	39	30
2.	82	80	83
3.	51	70	55
4.	53	60	69
5.	50	N	N
6.	15	16	25
7.	54	61	81
8.	81	76	74
9.	52	67	72
10.	37	61	67
11.	50	54	61
12.	32	33	39
13.	19	46	57
14.	21	33	35

FEMALE

15.	61	63	72
16.	51	62	60
17.	48	51	72
18.	45	41	45
19.	51	41	37
20.	44	36	44
21.	24	22	24
22.	25	38	39
23.	37	38	46
24.	82	74	74
25.	32	28	42
26.	44	48	52
27.	58	59	68

TABLE LXIII

SOCIAL STUDIES
CONTINUAL TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	90	88	89
2.	77	73	76
3.	58	58	64
4.	83	81	83
5.	54	55	60
6.	18	23	05
7.	52	54	52
8.	34	31	12
9.	41	47	44
10.	56	53	62
11.	51	52	48
12.	52	53	45
13.	48	41	29
14.	69	73	77

FEMALE

15.	58	61	53
16.	59	57	63
17.	34	29	35
18.	61	53	59
19.	59	63	61
20.	77	81	81
21.	45	46	45
22.	68	66	71
23.	47	55	56
24.	65	67	69
25.	63	70	75
26.	54	56	74
27.	49	54	57

TABLE LXIV

SCIENCE

RAW SCORES OF THE TESTING PROGRAM OF 1962 - 1963

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	72	67	55
2.	46	47	37
3.	48	66	46
4.	81	75	71
5.	52	61	53
6.	59	67	70
7.	86	84	81
8.	84	74	59
9.	73	64	66
10.	57	53	44
11.	68	69	71
12.	80	77	71
13.	82	72	83
14.	74	71	69
15.	69	66	58

FEMALE

16.	79	64	48
17.	74	61	50
18.	41	40	33
19.	59	56	44
20.	50	46	33
21.	69	65	58
22.	88	74	62
23.	87	86	79
24.	52	49	53
25.	57	49	39
26.	72	66	48
27.	82	73	70
28.	91	89	91
29.	94	89	92
30.	70	62	55
31.	80	76	63
32.	94	88	87
33.	71	70	55
34.	55	51	43
35.	77	66	63

TABLE LXV

SCIENCE

RAW SCORES OF THE TESTING PROGRAM OF 1963 - 1964

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	65	63	48
2.	80	75	63
3.	51	41	32
4.	59	54	55
5.	73	69	70
6.	37	28	30
7.	80	84	90
8.	63	61	40
9.	96	82	92
10.	30	31	40
11.	81	83	88
12.	71	70	61
13.	68	75	88
14.	44	52	60
15.	55	60	72

FEMALE

16.	65	71	61
17.	94	89	80
18.	81	77	83
19.	79	69	53
20.	75	66	55
21.	83	83	78
22.	70	56	52
23.	76	79	74
24.	76	74	61
25.	75	68	55
26.	50	55	48
27.	68	63	77
28.	65	69	68
29.	82	85	84
30.	83	76	67
31.	83	82	80
32.	45	43	50
33.	74	62	61
34.	57	55	61
35.	82	72	60

TABLE LXVI

SCIENCE

RAW SCORES OF THE TESTING PROGRAM OF 1964 - 1965

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	57	65	75
2.	52	58	50
3.	38	50	33
4.	67	70	76
5.	39	50	51
6.	72	83	79
7.	57	56	59
8.	57	59	64
9.	49	57	54
10.	45	60	44
11.	57	50	31
12.	57	59	52
13.	72	76	75
14.	67	66	62
15.	49	57	50

FEMALE

16.	62	44	36
17.	57	57	40
18.	62	74	69
19.	87	86	89
20.	57	65	61
21.	87	73	60
22.	77	73	58
23.	92	86	89
24.	52	58	60
25.	57	46	30
26.	62	60	64
27.	57	69	65
28.	57	54	48
29.	67	73	64
30.	62	54	42
31.	77	75	74
32.	67	63	65
33.	87	78	83
34.	87	84	54
35.	67	78	70

TABLE LXVII

SCIENCE

RAW SCORES OF THE TESTING PROGRAM OF 1965 - 1966

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	54	58	39
2.	39	54	44
3.	69	68	66
4.	56	60	62
5.	69	72	67
6.	65	72	68
7.	61	72	64
8.	65	67	33
9.	75	75	76
10.	55	61	59
11.	51	50	36
12.	73	72	66
13.	71	76	67
14.	65	62	51
15.	79	53	45

FEMALE

16.	61	68	64
17.	70	67	55
18.	78	83	84
19.	58	62	51
20.	83	85	86
21.	78	64	77
22.	43	50	42
23.	65	67	58
24.	67	73	68
25.	65	71	78
26.	45	50	31
27.	60	64	52
28.	45	51	54
29.	84	73	83
30.	59	58	48
31.	78	78	58
32.	64	64	31
33.	63	66	61
34.	69	69	53
35.	48	82	79

TABLE LXVIII

SCIENCE
TERM TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	64	67	46
2.	93	89	78
3.	98	78	68
4.	83	80	68
5.	53	N	N
6.	47	50	55
7.	58	78	68
8.	91	92	74
9.	88	79	78
10.	83	78	64
11.	76	74	55
12.	66	62	48
13.	68	68	58
14.	64	62	52

FEMALE

15.	88	81	70
16.	90	84	77
17.	86	81	68
18.	78	73	58
19.	80	77	64
20.	82	71	55
21.	35	39	43
22.	69	60	58
23.	60	61	46
24.	97	86	71
25.	64	56	59
26.	80	70	69
27.	76	70	61

TABLE LXIX

SCIENCE
CONTINUAL TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	95	92	82
2.	82	76	67
3.	58	65	74
4.	86	85	83
5.	54	N	N
6.	33	33	14
7.	63	64	62
8.	49	53	53
9.	54	54	40
10.	65	68	73
11.	53	57	53
12.	60	51	52
13.	76	65	57
14.	85	80	63

FEMALE

15.	72	73	62
16.	77	72	63
17.	47	40	20
18.	71	65	63
19.	75	64	58
20.	95	90	81
21.	67	59	58
22.	72	68	60
23.	54	59	58
24.	80	78	68
25.	73	74	58
26.	75	68	52
27.	69	67	69

TABLE LXX

MATHEMATICS

RAW SCORES OF THE TESTING PROGRAM OF 1962 - 1963

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	76	76	77
2.	76	62	50
3.	81	69	68
4.	93	84	69
5.	95	88	69
6.	94	90	92
7.	86	91	94
8.	62	78	70
9.	92	86	91
10.	87	71	56
11.	79	71	86
12.	98	97	92
13.	92	87	82
14.	92	97	91
15.	95	81	76

FEMALE

16.	66	65	76
17.	61	47	53
18.	51	50	33
19.	71	50	33
20.	42	37	19
21.	78	55	30
22.	63	74	66
23.	95	89	85
24.	88	61	40
25.	69	67	46
26.	84	89	87
27.	74	81	77
28.	96	97	93
29.	100	94	92
30.	92	71	58
31.	90	88	75
32.	98	96	97
33.	82	86	75
34.	76	50	39
35.	85	61	42

TABLE LXXI

MATHEMATICS

RAW SCORES OF THE TESTING PROGRAM OF 1963 - 1964

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	96	84	88
2.	84	86	93
3.	71	66	73
4.	74	76	75
5.	79	63	66
6.	27	32	16
7.	92	88	95
8.	50	60	64
9.	97	87	96
10.	27	27	16
11.	96	91	95
12.	91	81	85
13.	71	76	85
14.	53	58	56
15.	58	55	44

FEMALE

16.	83	74	87
17.	98	90	96
18.	81	83	91
19.	79	63	62
20.	76	70	76
21.	61	64	79
22.	94	83	73
23.	92	83	85
24.	93	81	79
25.	63	67	74
26.	31	35	47
27.	76	71	68
28.	81	74	84
29.	98	86	90
30.	88	85	83
31.	94	86	85
32.	45	51	39
33.	45	49	58
34.	44	52	62
35.	70	62	63

TABLE LXXII

MATHEMATICS

RAW SCORES OF THE TESTING PROGRAM OF 1964 - 1965

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	82	75	85
2.	72	58	72
3.	52	56	58
4.	67	61	57
5.	52	45	47
6.	82	90	87
7.	87	73	84
8.	87	71	79
9.	67	61	63
10.	49	59	69
11.	67	59	45
12.	97	84	90
13.	87	75	73
14.	82	77	69
15.	40	46	37

FEMALE

16.	30	40	22
17.	62	50	44
18.	72	70	81
19.	92	90	95
20.	35	36	27
21.	49	57	64
22.	62	52	57
23.	100	83	99
24.	57	29	28
25.	72	46	52
26.	72	60	61
27.	77	65	77
28.	67	45	58
29.	77	74	79
30.	62	50	54
31.	57	55	54
32.	77	81	78
33.	67	70	64
34.	77	76	57
35.	92	79	88

TABLE LXXIII

MATHEMATICS

RAW SCORES OF THE TESTING PROGRAM OF 1965 - 1966

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	61	51	63
2.	60	59	52
3.	69	62	65
4.	53	50	51
5.	56	52	43
6.	60	51	39
7.	80	77	77
8.	69	64	58
9.	73	71	77
10.	70	67	71
11.	59	50	25
12.	67	64	61
13.	71	72	57
14.	53	50	48
15.	47	51	50

FEMALE

16.	34	44	31
17.	77	69	57
18.	79	77	84
19.	80	77	71
20.	85	82	83
21.	69	80	63
22.	38	42	50
23.	59	66	65
24.	80	78	73
25.	65	62	63
26.	51	50	39
27.	68	70	85
28.	54	50	58
29.	70	83	89
30.	54	45	25
31.	55	50	42
32.	73	70	46
33.	67	65	71
34.	63	61	50
35.	89	87	88

TABLE LXXIV

MATHEMATICS
TERM TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	60	52	53
2.	86	86	91
3.	90	87	80
4.	94	86	87
5.	55	N	N
6.	53	41	39
7.	73	68	68
8.	81	65	54
9.	68	61	70
10.	83	75	72
11.	68	57	48
12.	61	51	53
13.	75	66	70
14.	56	44	53

FEMALE

15.	87	73	59
16.	79	69	72
17.	75	65	59
18.	75	66	61
19.	82	62	77
20.	73	56	70
21.	78	62	62
22.	75	61	59
23.	71	62	64
24.	94	83	76
25.	72	71	74
26.	61	51	61
27.	73	68	57

TABLE LXXV

MATHEMATICS
CONTINUAL TESTING
RAW SCORES OF THE TESTING PROGRAM OF 1966 - 1967

MALE

	FIRST TERM	YEAR'S AVERAGE	JUNE MARK
1.	96	89	87
2.	84	78	83
3.	71	66	67
4.	92	88	87
5.	47	57	16
6.	71	47	05
7.	69	59	63
8.	65	44	26
9.	67	57	51
10.	92	84	77
11.	48	32	10
12.	73	67	08
13.	75	60	45
14.	85	75	84

FEMALE

15.	83	72	67
16.	78	62	67
17.	79	63	66
18.	84	70	71
19.	80	70	80
20.	99	95	92
21.	66	53	61
22.	75	66	67
23.	81	66	63
24.	88	74	74
25.	90	87	81
26.	91	78	80
27.	89	76	74

TABLE LXXVI

QUESTIONNAIRE PLAN

Section	Question	Content
1	1	commital
2	2, 3, 4	better motivation
3	5, 6, 7	follow through on motivation
4	8, 9, 14	study habits
5	10, 13, 15, 16	attitude towards exams and system
6	11, 12, 17, 18, 21, 22	psychological effects
7	19, 20, 23	consider time factor
8	24, 25, 26, 27	opinions on term and June Results
9	28, 29, 30	information to parent, teacher and student on how the student is getting along

TABLE LXXVII

QUESTIONNAIRE RESULTS

TOTALS

	CN	TN	NOP	TN	C%	T%	NO
1.	521	71	28	620	84	11	5
2.	510	50	57	617	82	9	9
3.	558	34	26	618	90	5	5
4.	259	103	257	619	42	16	42
5.	500	51	69	620	80	8	12
6.	562	36	22	620	90	6	4
7.	251	80	284	615	41	14	45
8.	500	63	58	621	80	11	9
9.	487	88	45	620	78	14	8
10.	387	146	92	625	62	23	15
11.	493	61	59	613	80	10	10
12.	317	205	83	605	52	34	14
13.	163	288	167	618	26	46	28
14.	127	323	164	614	20	52	28
15.	229	173	216	618	37	28	35
16.	232	159	218	609	39	27	34
17.	135	411	81	627	21	65	14
18.	532	37	50	619	86	6	8
19.	237	149	226	612	38	25	37
20.	225	152	238	615	37	24	39
21.	253	177	178	608	41	30	29
22.	109	393	109	611	18	64	18
23.	260	103	247	610	42	18	40
24.	465	78	74	617	75	15	14
25.	94	234	283	611	15	38	47
26.	392	109	119	620	63	17	20
27.	111	243	250	604	18	40	42
28.	553	29	34	616	90	4	6
29.	539	21	54	614	87	4	9
30.	544	33	37	614	89	5	6

CN continual numerical
TN term numerical
NOP no opinion numerical
TN total numerical
C% continual percentage
T% term percentage
NO no opinion

TABLE LXXVIII

QUESTIONNAIRE RESULTS (STUDENTS)

	FC	MC	CT	C%	FT	MT	TT	T%	FNO	MNO	NOT	NO%	TN
1.	175	157	332	86	16	28	45	11	5	8	13	3	390
2.	161	156	317	83	14	18	32	8	18	20	38	9	387
3.	183	177	360	94	4	11	15	3	5	7	12	3	387
4.	95	85	180	46	27	41	68	17	74	66	140	37	388
5.	152	151	303	77	17	19	36	10	25	25	50	13	389
6.	181	175	356	91	7	12	19	5	6	8	14	4	389
7.	84	86	170	44	29	32	61	17	78	75	153	39	384
8.	165	150	315	80	17	22	39	10	14	22	36	10	390
9.	160	138	298	76	25	42	67	18	11	13	24	6	389
10.	112	115	227	57	43	60	103	26	40	25	65	17	395
11.	156	156	312	80	26	16	42	10	15	21	36	10	390
12.	98	97	195	51	67	71	138	36	18	27	45	13	378
13.	52	71	123	31	94	68	162	41	48	54	102	28	387
14.	38	56	94	24	112	82	194	50	44	53	97	26	385
15.	86	79	165	43	60	55	115	29	48	60	108	29	388
16.	89	89	178	47	40	55	95	26	57	48	105	27	378
17.	46	57	103	25	125	129	254	63	23	17	40	12	397
18.	168	163	331	85	13	15	28	7	12	18	30	8	389
19.	95	91	186	48	36	53	89	24	58	50	108	28	383
20.	76	90	166	45	45	52	97	26	68	54	102	28	365
21.	100	84	184	49	44	62	106	27	40	52	92	24	382
22.	38	46	84	21	126	116	242	64	26	31	57	15	383
23.	95	93	188	48	30	50	80	20	65	54	119	32	387
24.	154	144	298	77	23	24	47	12	16	28	44	11	389
25.	22	45	67	17	83	68	151	39	84	82	166	44	384
26.	140	124	264	67	28	37	65	17	25	36	61	16	390
27.	31	53	84	21	84	70	154	40	74	72	146	39	384
28.	179	167	346	88	6	10	16	4	9	20	29	8	391
29.	177	170	347	90	3	9	12	12	3	18	30	7	389
30.	174	172	346	88	6	13	19	4	8	17	25	8	390

FC female continual
MC male continual
CT continual total
C% continual percentage
FT female term
MT male term
TT total term
T% term percentage
FNO female no opinion
MNO male no opinion
NOT no opinion total
NO% no opinion percentage
TN total number

TABLE LXXIX

QUESTIONNAIRE RESULTS (PARENTS)

	CF	CM	CT	C%	TF	TM	TT	T%	NOF	NOM	NO%	NOT	TR
1.	71	104	175	82	7	16	23	10	7	8	8	15	213
2.	73	107	180	84	5	13	18	8	8	7	8	15	213
3.	77	105	182	84	2	15	17	8	7	7	8	14	213
4.	28	47	75	35	10	21	31	14	52	55	51	107	213
5.	74	109	183	84	5	10	15	8	5	10	8	15	213
6.	73	117	190	89	2	14	16	8	0	7	3	7	213
7.	30	46	76	35	8	8	16	8	51	70	57	121	213
8.	71	104	175	83	9	12	21	9	6	11	8	17	213
9.	70	105	175	83	8	13	21	9	7	10	8	17	213
10.	65	89	154	73	13	24	37	18	8	14	9	22	213
11.	73	98	171	84	6	11	17	8	6	11	8	17	205
12.	50	63	113	54	21	39	60	28	14	22	18	36	209
13.	12	23	35	17	46	70	116	55	28	34	28	62	213
14.	13	18	31	14	47	76	123	59	24	33	28	57	211
15.	18	39	57	26	17	41	58	26	51	47	48	98	213
16.	20	27	47	22	18	45	63	29	49	54	49	103	213
17.	12	17	29	13	62	89	151	72	10	22	15	32	212
18.	76	109	185	88	2	6	8	3	4	15	9	19	212
19.	21	27	48	22	18	41	59	27	48	57	51	105	212
20.	18	35	53	24	23	29	52	24	45	63	52	108	213
21.	21	42	63	30	24	45	69	33	38	39	37	77	209
22.	10	13	23	10	57	85	142	68	17	29	22	46	211
23.	22	47	69	34	9	13	22	10	56	60	56	116	217
24.	67	90	157	74	9	16	25	12	6	22	14	28	210
25.	9	18	27	12	30	44	74	36	47	61	52	108	209
26.	54	69	123	57	13	27	40	18	17	33	25	50	213
27.	10	17	27	12	41	39	80	38	36	60	50	106	213
28.	76	114	190	91	4	8	12	5	1	4	3	5	207
29.	69	106	175	84	4	4	8	3	8	16	13	24	207
30.	71	110	181	88	5	9	14	6	5	7	6	12	207

CF continual female
CM continual male
CT continual total
C% continual percentage
TF term female
TM term male
TT term total
T% term percentage
NOF no. of female
NOM no. of male
NO% no opinion percentage
NOT no opinion total
TR total responses

TABLE LXXX

QUESTIONNAIRE RESULTS (FACULTY)

	CF	CM	T	C%	TF	TM	T%	T	NOF	NOM	NO%	T	TR
1.	1	13	14	77	1	3	23	4	0	0		0	18
2.	1	12	13	77	0	0		0	0	4	23	4	17
3.	2	14	16	88	0	2	12	2	0	0		0	18
4.	1	3	4	22	0	4	22	4	1	9	55	10	18
5.	1	13	14	77	0	0		0	1	3	23	4	18
6.	1	15	16	88	0	1	11	1	1	0	11	1	18
7.	1	4	5	27	0	3	18	3	1	9	55	10	18
8.	1	9	10	55	0	3	18	3	1	4	27	5	18
9.	2	12	14	77	0	0		0	1	3	23	4	18
10.	0	6	6	36	0	6	36	6	1	4	28	5	17
11.	1	9	10	55	1	1	11	2	0	6	34	6	18
12.	0	9	9	50	1	6	38	7	1	1	12	2	18
13.	0	5	5	27	1	9	55	10	1	2	18	3	18
14.	0	2	2	12	1	5	33	6	1	10	55	11	18
15.	1	6	7	42	0	0		0	1	9	58	10	17
16.	1	6	7	38	0	1	7	1	1	9	55	10	18
17.		3	3	17	1	5	33	6	1	8	50	9	18
18.	2	14	16	88	0	1	6	1	0	1	6	1	18
19.	0	3	3	16	0	1	7	1	2	11	77	13	17
20.	2	4	6	36	0	3	17	3	0	8	47	8	17
21.	2	4	6	36	0	2	11	2	0	9	53	9	17
22.	0	2	2	11	1	8	53	9	1	5	36	6	17
23.	1	2	3	18	1	0	7	1	0	12	75	12	16
24.	2	8	10	55	0	6	33	6	0	2	12	2	18
25.	0				1	8	50	9	1	8	50	9	18
26.	0	5	5	30	2	2	23	4	0	8	47	8	17
27.	0	0	0		1	8	53	9	1	7	47	8	17
28.	2	15	17	95	0	1	5	1	0	0	0	0	18
29.	2	15	17	95	0	1	5	1	0	0		0	18
30.	2	15	17	95	0	1	5	1	0	0		0	18

CF continual female
CM continual male
T total
C% continual percentage
TF term female
TM term male
T% term percentage
T total
NOF no opinion female
NOM no opinion male
NO% no opinion percentage
T total
TR total responses

TABLE LXXXI

QUESTIONNAIRE RESULTS
MALE GROUPS

	CN	TN	NON	T	CP	TP	NOP
1.	274	47	16	337	81	13	6
2.	275	31	31	337	82	9	9
3.	296	28	14	338	87	9	4
4.	135	66	130	331	41	19	40
5.	273	29	38	340	80	8	12
6.	307	27	15	349	87	8	5
7.	136	43	154	333	40	13	47
8.	263	37	37	337	78	11	11
9.	255	55	26	336	75	16	9
10.	210	90	43	343	62	27	11
11.	263	28	38	329	79	8	13
12.	169	116	50	335	50	34	16
13.	99	147	90	336	29	43	28
14.	76	163	95	334	22	48	30
15.	124	96	116	336	36	30	34
16.	122	101	111	334	36	30	34
17.	77	223	47	347	23	65	12
18.	286	22	34	342	83	6	11
19.	121	95	118	334	37	28	35
20.	129	84	125	338	39	25	36
21.	130	109	100	339	38	33	29
22.	61	209	65	335	19	62	19
23.	142	63	126	331	42	20	38
24.	242	46	52	340	72	13	15
25.	63	120	151	334	18	35	47
26.	198	66	77	341	58	19	23
27.	70	117	139	326	21	35	44
28.	296	19	24	339	87	5	8
29.	291	14	34	339	85	4	11
30.	297	23	24	344	86	7	7

CN	continual numerical
TN	term numerical
NON	no opinion numerical
T	total
CP	continual percentage
TP	term percentage
NOP	no opinion percentage

TABLE LXXXII

QUESTIONNAIRE RESULTS
FEMALE GROUPS

	CN	TN	NON	TN	C%	T%	NOP
1.	247	24	12	283	88	8	4
2.	235	19	26	280	83	6	11
3.	262	6	12	280	91	3	6
4.	124	37	127	288	43	12	45
5.	227	22	31	280	81	7	12
6.	255	9	7	271	94	4	2
7.	115	37	130	282	40	13	47
8.	237	26	21	284	83	9	8
9.	232	33	19	284	81	11	8
10.	177	56	49	282	63	19	18
11.	230	33	21	284	80	11	9
12.	148	89	33	270	54	32	14
13.	64	141	77	282	22	50	28
14.	51	160	69	280	19	57	24
15.	105	77	100	282	38	28	34
16.	110	58	107	275	40	21	39
17.	58	188	34	280	20	67	13
18.	246	15	16	277	88	6	5
19.	116	54	108	278	41	19	40
20.	98	68	113	277	34	24	42
21.	123	68	78	269	45	26	29
22.	48	184	44	276	17	66	17
23.	118	40	121	279	42	15	43
24.	223	32	22	277	80	11	9
25.	31	114	132	277	12	41	47
26.	194	43	42	279	70	15	15
27.	41	126	111	278	14	46	40
28.	257	10	10	277	92	4	4
29.	248	7	20	275	91	1	8
30.	247	10	13	270	92	4	4

CN continual numerical
TN term numerical
NON no opinion numerical
TN total numerical
C% continual percentage
T% term percentage
NOP no opinion percentage

TABLE LXXXIII

PERCENTAGE FAVORING TERM TESTING

QUEST.	PARENTS	STUDENTS	FACULTY	GROUP
1	10%	11%	23%	11%
2	8	8	0	9
3	8	3	12	5
4	14	17	22	16
5	8	10	0	8
6	8	5	11	6
7	8	17	18	14
8	99	10	18	11
9	9	18	0	14
14	59	50	33	52
10	18	26	36	23
13	55	41	55	46
15	26	29	0	28
16	29	26	7	27
11	8	10	11	10
12	28	36	38	34
17	72	63	33	65
18	3	7	6	6
21	33	27	11	30
22	68	64	53	64
19	27	24	7	25
20	24	26	17	24
23	10	20	7	18
24	12	12	33	13
25	36	39	50	38
26	18	17	23	17
27	38	40	53	40
28	6	4	5	4
29	3	3	5	4
30	6	4	5	5

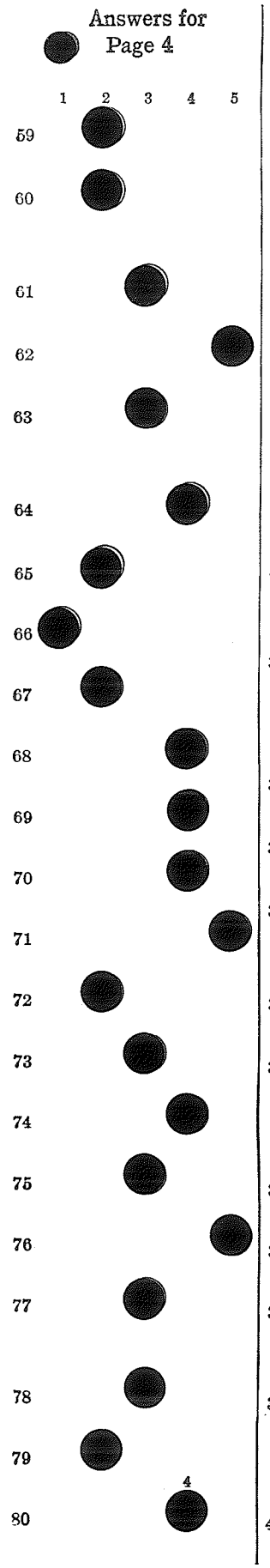
TABLE LXXXIV

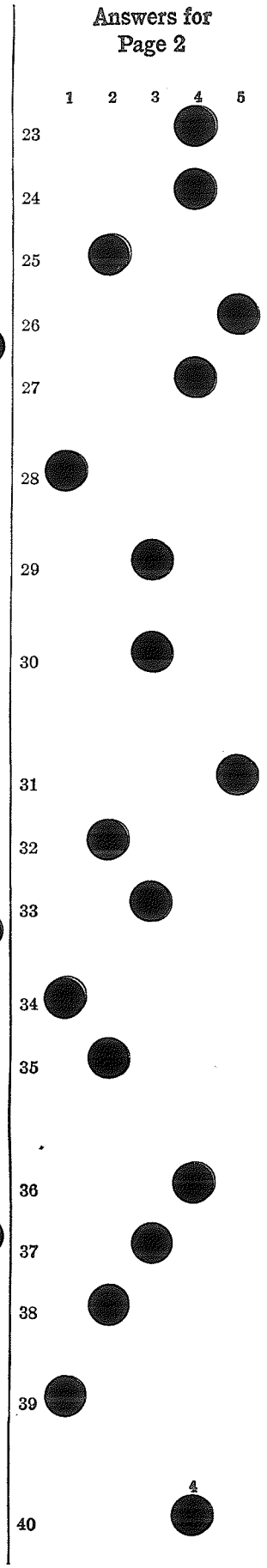
PERCENTAGE FAVORING CONTINUAL TESTING

QUEST.	PARENTS	STUDENTS	FACULTY	GROUP
1	82	86	77	84
2	84	83	77	82
3	84	94	88	90
4	35	46	22	42
5	84	77	77	80
6	89	91	88	90
7	35	44	27	41
8	83	80	55	80
9	83	76	77	78
14	14	24	12	20
10	73	57	36	62
13	17	31	27	26
15	26	43	42	37
16	22	47	38	39
11	84	80	55	80
12	54	51	50	52
17	13	25	17	21
18	88	85	88	86
21	30	49	36	41
22	10	21	11	18
19	22	48	16	38
20	24	46	36	37
23	34	48	18	42
24	74	77	55	75
25	12	17	0	15
26	57	67	30	63
27	12	21	0	18
28	91	88	95	90
29	84	90	95	87
30	88	88	95	89

APPENDIX B

Intelligence Tests and the Questionnaire





OTIS QUICK-SCORING MENTAL ABILITY TESTS

KEY FOR BETA TEST: FORM A

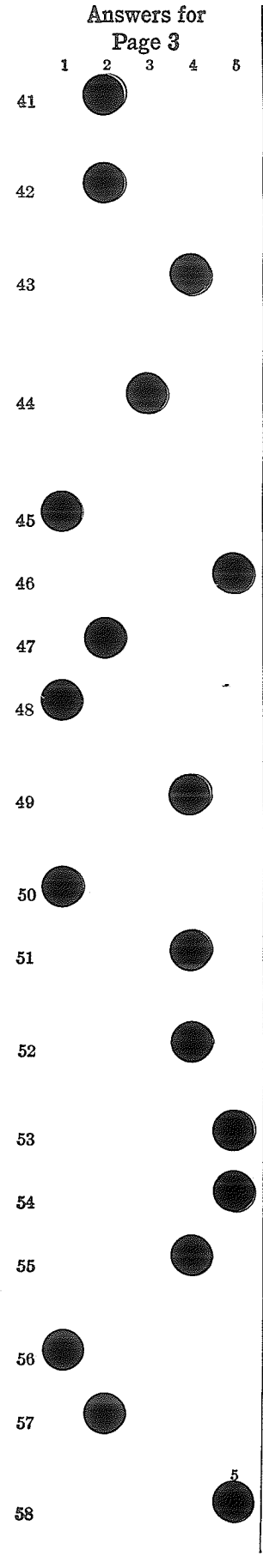
DIRECTIONS

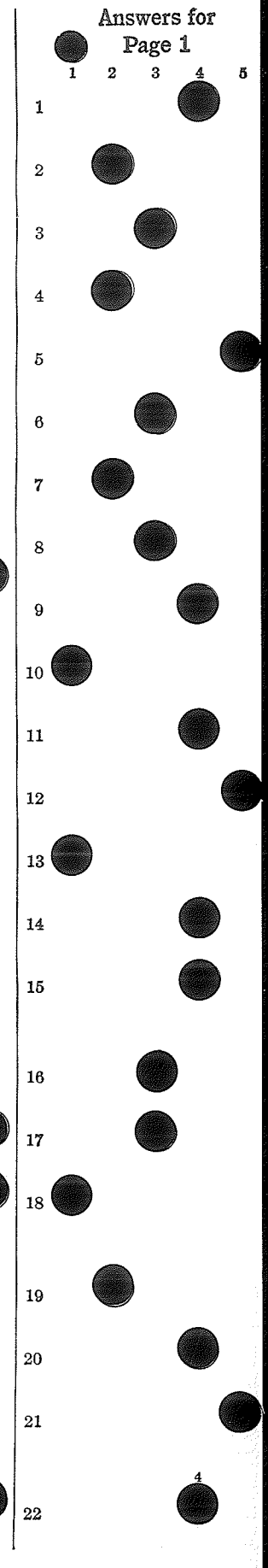
To score a test paper, open it so that the four columns of answers are in view. Notice the two black circles with crosses in them (⊕) at the top of the columns of answers for pages 1 and 4. Lay the Key on the test paper so that these two circles show through the holes in the corresponding positions in the Key. Then the Key will be in the correct position so that all crosses that are in the right circles will show through the holes in the Key.

Then merely count the crosses that can be seen through the holes. The number of crosses so counted is the pupil's score. Write the score in the space provided in the upper right-hand corner of the front page of the test.

The pupils are instructed to be sure not to put more than one cross in any row of five circles. However, if in any instance a pupil has put two crosses in the same row of circles, give no credit for that item. Some scorers glance over each paper before scoring it to see that only one response has been given to each item.

See the section "Directions for Scoring," in the Manual.





Edition ***a***

OTIS QUICK-SCORING MENTAL ABILITY TESTS

By ARTHUR S. OTIS, PH.D.
Formerly Development Specialist with Advisory Board, General Staff, United States War Department

Beta
A

BETA TEST: FORM A

IQ..... For Grades 4–9 Score.....

Read this page. Do what it tells you to do.

Do not open this booklet, or turn it over, until you are told to do so.
Fill these blanks, giving your name, age, birthday, etc. Write plainly.

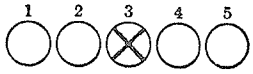
Name.....Age last birthday .....years
First name, initial, and last name

Birthday.....Teacher.....Date.....19.....
Month Day

Grade.....School.....City.....

This is a test to see how well you can think. It contains questions of different kinds. Here is a sample question already answered correctly. Notice how the question is answered:

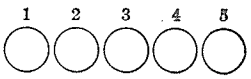
Sample: Which one of the five things below is soft?
1 glass 2 stone 3 cotton 4 iron 5 ice.....



The right answer, of course, is *cotton;* so the word *cotton* is underlined. And the word *cotton* is No. 3; so a heavy cross has been put in the 3d circle. This is the way you are to answer the questions.

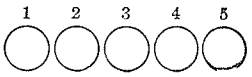
Try this sample question yourself. Do not write the answer; just draw a line under it and then put a heavy cross in the right circle.

Sample: A robin is a kind of —
1 plant 2 bird 3 worm 4 fish 5 flower .....



The answer is *bird;* so you should have drawn a line under the word *bird* and put a heavy cross in the 2d circle. Try this one:

Sample: Which one of the five numbers below is larger than 55?
1 53 2 48 3 29 4 57 5 16 .....



The answer, of course, is 57; so you should have drawn a line under 57 and put a heavy cross in the 4th circle.

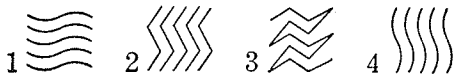
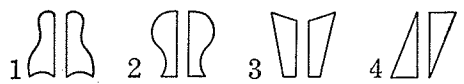
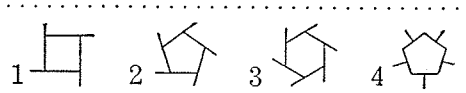
The test contains 80 questions. You are not expected to be able to answer all of them, but do the best you can. You will be allowed half an hour after the examiner tells you to begin. Try to get as many right as possible. Be careful not to go so fast that you make mistakes. Do not spend too much time on any one question. No questions about the test will be answered by the examiner after the test begins. Lay your pencil down.

Do not turn this booklet until you are told to begin.

PRINTED IN U.S.A.

Examination begins here.

Page 1

1. Which one of the five things below does not belong with the others?
   1 potato 2 turnip 3 carrot 4 stone 5 onion
2. Which one of the five answers below tells best what a sword is?
   1 to cut 2 a weapon 3 an officer 4 a tool 5 to fight
3. Which one of the five words below means the opposite of north?
   1 east 2 star 3 south 4 pole 5 equator
4. The peeling is to a banana and the husk is to an ear of corn the same as a shell is to what?
   1 an apple 2 an egg 3 juice 4 a peach 5 a hen
5. A child who knows he is guilty of doing wrong should feel —
   1 bad 2 sick 3 better 4 afraid 5 ashamed
6. Which one of the five things below is the largest?
   1 knee 2 toe 3 leg 4 ankle 5 foot
7. Which one of the five words means the opposite of strong?
   1 man 2 weak 3 small 4 short 5 thin
8. Three of the four designs at the right are alike. Which one is not like the other three?
   
9. Which one of the five things below is most like these three: a chair, a bed, and a stove?
   1 a chimney 2 a stick 3 a window 4 a table 5 a floor
10. A knee is to a leg as an elbow is to what?
    1 an arm 2 a shoulder 3 a bone 4 a wrist 5 a hand
11. Which word means the opposite of joy?
    1 sickness 2 bad 3 happiness 4 sorrow 5 cry
12. If I find a kind of plant that was never seen before, I have made —
    1 an invention 2 an adoption 3 a creation 4 a novelty 5 a discovery
13. A sculptor is to a statue as an author is to a —
    1 book 2 man 3 name 4 bookcase 5 pen
14. At 3 cents each, how many pencils can be bought for 27 cents?
    1 27 2 81 3 3 4 9 5 30
15. Three of the four designs at the right are alike. Which one is not like the other three?
    
16. Which is the most important reason that words in the dictionary are arranged alphabetically?
    1 It is the simplest arrangement. 2 It puts the shortest words first.
    3 It enables us to find any word quickly. 4 It is a custom. 5 The printing is easier.
17. Which one of the five things below is most like these three: a saw, a hammer, and a file?
    1 a bottle 2 a pen 3 a screw driver 4 a fork 5 a carpenter
18. If the following words were arranged in order, which word would be in the middle?
    1 luncheon 2 dress 3 undress 4 supper 5 breakfast
19. The saying, "Don't count your chickens before they are hatched," means —
    1 Don't hurry. 2 Don't be too sure of the future. 3 Haste makes waste.
    4 Don't gamble. 5 Don't raise chickens.
20. Three of the four designs at the right are alike. Which one is not like the other three?
    
21. A boy who often tells big stories about what he can do is said to —
    1 lie 2 fake 3 cheat 4 joke 5 brag
22. Which tells best just what a colt is?
    1 an animal with hoofs 2 an awkward little beast 3 an animal that runs fast
    4 a young horse 5 a little animal that eats hay

(*Go right on to Page 2.*)

Page 2

23. Which of the five things below is most like these three: a horse, a pigeon, and a cricket?
    1 a stall 2 a saddle 3 a feather 4 a goat 5 a wing
24. Railroad tracks are to a locomotive as what is to an automobile?
    1 tires 2 steam 3 speed 4 the road 5 gasoline
25. Which word means the opposite of pretty?
    1 good 2 ugly 3 bad 4 crooked 5 nice
26. Which one of the words below would come first in the dictionary?
    1 tramp 2 saint 3 razor 4 quart 5 grass
27. An event which is sure to happen is said to be —
    1 probable 2 possible 3 doubtful 4 certain 5 delayed
28. One number is wrong in the following series. What should that number be? 7 1 7 2 7 3 7 4 7 5 7 6 7 8
    1 7 2 6 3 8 4 4 5 5
29. Which of these series contains a wrong number?
    1 3–6–9–12–15 2 2–5–8–11–14 3 1–4–7–10–12
    4 2–4–6–8–10 5 1–3–5–7–9
30. Which one of the five things below is most like these three: a ship, a bicycle, and a truck?
    1 a sail 2 a wheel 3 a train 4 the ocean 5 a tire
31. Which statement tells best just what a hallway is?
    1 a small room 2 a place to hang your hat and coat
    3 it is long and narrow 4 where to say good-by
    5 a passage leading from one room to another
32. Steam is to water as water is to —
    1 hot 2 ice 3 an engine 4 a solid 5 gas
33. Which of these words would come last in the dictionary?
    1 health 2 juggle 3 normal 4 never 5 grateful
34. If George is taller than Frank and Frank is taller than James, then George is (?) James. 1 taller than
    2 shorter than 3 just as tall as 4 (cannot say which)
35. A man who betrays his country is called a —
    1 thief 2 traitor 3 enemy 4 coward 5 slacker
36. Count each 7 below that has a 5 next after it.
    7 5 3 0 9 7 3 5 8 7 7 4 2 1 7 5 7 3 2 4 7 0 9 3 7 5 5 7 2 5 7 7 5 4 7 1
    How many such 7's did you count?
    1 11 2 2 3 3 4 4 5 12
37. The daughter of my mother's brother is my —
    1 sister 2 niece 3 cousin 4 aunt 5 granddaughter
38. Peace is to war as (?) is to confusion.
    1 explosion 2 order 3 armistice 4 riot 5 police
39. If Paul is older than Herbert and Paul is younger than Robert, then Robert is (?) Herbert. 1 older than
    2 younger than 3 just as old as 4 (cannot say which)
40. If the following words were arranged in order, with what letter would the middle word begin?
    Week Year Hour Second Day Month Minute
    1 W 2 S 3 H 4 D 5 M

(*Go right on to Page 3.*)

Page 3

41. A quantity which grows larger is said to —
1 prosper 2 increase 3 fatten 4 rise 5 burst

42. A bicycle is to a motorcycle as a wagon is to what?
1 an engine 2 an automobile 3 a horse 4 slower
5 an airplane .....

43. Which of the five things below is most like these three: a tent, a flag, and a sail?
1 a shoe 2 a ship 3 a staff 4 a towel 5 a rope..

44. What is the most important reason that we use clocks?
1 To wake us up in the morning. 2 To help us catch trains.
3 To regulate our daily lives. 4 They are ornamental.
5 So that children will get to school on time. .....

45. If the following words were rearranged to make a good sentence, with what letter would the third word of the sentence begin?
houses stone built of men wood and
1 h 2 s 3 b 4 m 5 w.....

46. Which of these expressions is the most definite?
1 soon 2 early 3 later 4 morning 5 ten A.M.

47. A vase is to flowers as (?) is to milk.
1 a cow 2 a pitcher 3 white 4 drink 5 cream

48. A lamp is to a light as (?) is to a breeze.
1 a fan 2 bright 3 a sailboat 4 a window 5 blow

49. If the following words were arranged in order, which word would be in the middle?
1 good 2 excellent 3 wretched 4 fair 5 poor

50. If Henry is taller than Tom and Henry is shorter than George, then George is (?) Tom. 1 taller than 2 shorter than
3 just as tall as 4 (cannot say which) .....

51. A king is to a kingdom as a president is to what? 1 queen
2 vice-president 3 senate 4 republic 5 democrat

52. John is the fifth child from each end of a row. How many pupils are there in a row?
1 ten 2 eleven 3 seven 4 nine 5 five.....

53. Which tells best what an automobile is? 1 a thing with tires
2 something to travel in 3 an engine mounted on wheels
4 a horseless carriage 5 a vehicle propelled by an engine

54. Brick is to a wall as (?) is to a table.
1 a chair 2 red 3 eat 4 a kitchen 5 wood

55. A wire is to electricity as (?) is to gas.
1 a flame 2 a spark 3 hot 4 a pipe 5 a stove

56. An object or institution that is designed to last only a short time is said to be — 1 temporary 2 changeable
3 unsound 4 worthless 5 unstable .....

57. Which word means the opposite of humility?
1 joy 2 pride 3 dry 4 funny 5 recklessness

58. A word that means suitable, fit, or proper is —
1 grotesque 2 odd 3 inadequate 4 superfluous
5 appropriate.....

(Go on to Page 4 under Page 2.)

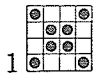
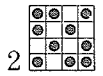
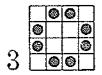
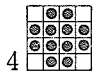
Answers for Page 3

41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58

Answers for Page 1

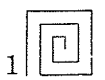
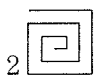
Page 4

59. If the words below were rearranged to make a good sentence, the third word of the sentence would begin with what letter? men high the a wall built stone
1 m 2 b 3 h 4 w 5 s

60. Three of the four designs at the right are alike. Which one is not like the other three?
1  2  3  4 

61. There is a saying, " Any port in a storm." This means —
1 **Ships should not venture out to sea in storms.** 2 **Stormy weather causes large waves in harbors.**
3 **In emergencies any aid is acceptable.** 4 **Ships usually sink in storms.**

62. Which one of the five things below is most like these three: cannon ball, wire, and penny?
1 **dollar bill** 2 **bone** 3 **string** 4 **pencil** 5 **key**

63. Three of the four designs at the right are alike. Which one is not like the other three?
1  2  3  4 

64. There is a saying, " Don't look a gift horse in the mouth." This means —
1 **It is not safe to look into the mouth of a horse.** 2 **You cannot judge the age of a gift horse by his teeth.**
3 **Don't accept a horse as a gift.** 4 **Although you question the value of a gift, accept it graciously.**

65. A boy is to a man as (?) is to a sheep.
1 **wool** 2 **a lamb** 3 **a goat** 4 **a shepherd** 5 **a dog**

66. If the following words were arranged to make a good sentence, with what letter would the last word of the sentence begin? usually are of made tables wood 1 w 2 d 3 a 4 t 5 m

67. An agreement reached in which both sides yield somewhat in their demands is called —
1 **a promise** 2 **a compromise** 3 **an understanding** 4 **a deadlock** 5 **an armistice**

68. Three of the four designs at the right are alike. Which one is not like the other three?
1  2  3  4 

69. What is the letter that follows the letter that comes next after M in the alphabet?
1 L 2 M 3 N 4 O 5 P

70. One number is wrong in the following series: 1 2 4 8 24 32 64 What should that number be?
1 **6** 2 **12** 3 **3** 4 **16** 5 **48**

71. If I have a large box with two small boxes in it and five very small boxes in each small box, how many boxes are there in all? 1 **eight** 2 **seven** 3 **ten** 4 **twelve** 5 **thirteen**

72. There is a saying, " An ounce of practice is worth a pound of preaching." This means —
1 **Don't preach.** 2 **Deeds count more than words.** 3 **Preaching takes practice.** 4 **Don't practice.**

73. If a photograph that is 2 inches wide and 3 inches long is enlarged to be 10 inches wide, how many inches long will it be? 1 **11** 2 **12** 3 **15** 4 **20** 5 **30**

74. One number is wrong in this series: 1 2 4 5 7 8 10 11 12 14 What should that number be?
1 **9** 2 **6** 3 **3** 4 **13** 5 **16**

75. When the time by a clock was 8 minutes past 10, the hands were interchanged. The clock then said about —
1 **10 min. past 8** 2 **8 min. past 10** 3 **8 min. of 2** 4 **8 min. of 10** 5 **8 min. past 2**

76. A car owner uses a mixture in his radiator containing 1 quart of alcohol to every 2 quarts of water. How many quarts of alcohol are needed for 15 quarts of the mixture? 1 **7½** 2 **½** 3 **14** 4 **30** 5 **5**

77. What letter in the following series appears a third time nearest the beginning?
A E C B A D D E C F B C D A E E B D 1 A 2 B 3 C 4 D 5 E

78. In a foreign language
pira numa bega means *very deep snow; pira seco* means *white snow; numa copa* means *very well.*
What word means deep? 1 **pira** 2 **numa** 3 **bega** 4 **copa** 5 **seco**

79. Which of the five words below does not belong with the others?
1 **brave** 2 **clever** 3 **honest** 4 **generous** 5 **loyal**

80. A boy is now three times as old as his sister. In 4 years he will be only twice as old. How many years old is his sister? 1 **1** 2 **2** 3 **3** 4 **4** 5 **5**

Answers for Page 4

59 1 2 3 4 **5**
60 1 2 3 4
61 1 2 3 **4**
62 1 2 3 4 **5**
63 1 2 3 **4**
64 1 2 3 **4**
65 1 2 3 4 **5**
66 1 2 3 4 **5**
67 1 2 3 4 **5**
68 1 2 3 **4**
69 1 2 3 4 **5**
70 1 2 3 4 **5**
71 1 2 3 4 **5**
72 1 2 3 4
73 1 2 3 4 **5**
74 1 2 3 4 **5**
75 1 2 3 4 **5**
76 1 2 3 4 **5**
77 1 2 3 4 **5**
78 1 2 3 4 **5**
79 1 2 3 4 **5**
80 1 2 3 4 **5**

Answers for Page 2

23 1 2 3 4 **5**
24 1 2 3 4 **5**
25 1 2 3 4 **5**
26 1 2 3 4 **5**
27 1 2 3 4 **5**
28 1 2 3 4 **5**
29 1 2 3 4 **5**
30 1 2 3 4 **5**
31 1 2 3 4 **5**
32 1 2 3 4 **5**
33 1 2 3 4 **5**
34 1 2 3 4
35 1 2 3 4 **5**
36 1 2 3 4 **5**
37 1 2 3 4 **5**
38 1 2 3 4 **5**
39 1 2 3 4
40 1 2 3 4 **5**

OTIS QUICK-SCORING MENTAL ABILITY TESTS

CLASS RECORD FOR BETA OR GAMMA TEST

Test used (underline) Beta Gamma Form used.....

Grade.....Examiner.....

Teacher.....School.....

City.....Date of exam.....19....

Name	Age		Score	IQ	Add'l data (?)	
	Yrs.	Mos.				
Class Medians						

Printed in U.S.A.

(*Continued on reverse side*)

NOTE. See under "Reporting to the Author" in the Manual of Directions, regarding a request for data.

CLASS RECORD FOR BETA OR GAMMA TEST — *Continued*

Name	Age		Score	IQ	Add'l data (?)	
	Yrs.	Mos.				

DIRECTIONS. To distribute the scores of a class, make a mark in the second column for each score opposite the interval within which the score falls. In the "Frequency" column write the number of tally marks that fell in each interval.

Scores	Distributing Marks	Freq.
80–84		
75–79		
70–74		
65–69		
60–64		
55–59		
50–54		
45–49		
40–44		
35–39		
30–34		
25–29		
20–24		
15–19		
10–14		
5–9		
0–4		
Total		
Median		

OTIS QUICK-SCORING MENTAL ABILITY TESTS

By ARTHUR S. OTIS, PH.D.
Formerly Development Specialist with Advisory Board, General Staff, United States War Department

MANUAL OF DIRECTIONS FOR BETA TEST
FORMS A AND B

THE QUICK-SCORING SERIES

The Otis Quick-Scoring Mental Ability Tests comprise three tests, called Alpha, Beta, and Gamma. The three tests are designed for grades as follows:

Alpha Test . . . . Grades 1–4
Beta Test . . . . . Grades 4–9
Gamma Test . . . High Schools and Colleges

The Alpha Test consists entirely of pictures and is completely new. The Beta and Gamma Tests are revisions and extensions of the Intermediate and Higher Examinations, respectively, of the Otis Self-Administering Tests of Mental Ability.

PURPOSE OF THE TESTS

The purpose of the three tests in the series is to measure mental ability — thinking power or the degree of maturity of the mind.

It should be understood from the outset that it is not possible to measure mental ability directly. It is possible only to measure the effect mental ability has had in enabling the pupil to acquire certain knowledge and mental skill. Of course the answering of some types of questions depends less upon schooling and more upon mental ability than the answering of others, and in making up the test the aim has been for the most part to choose that kind of question which depends as little as possible on schooling and as much as possible on thinking.

However, in the interest of variety it has been found necessary and even advantageous to include in verbal tests of mental ability such as the Beta and Gamma Tests certain questions which might seem at first glance to be mere measures of achievement. This type includes questions on vocabulary, arithmetic reasoning, etc. It must be remembered, however, that any test which involves the use of language can measure mental ability only to the extent to which we may assume that pupils of the same age have had approximately the same opportunity to learn. Consequently, if a pupil has grown up with a limited educational opportunity, especially with reference to language, his mental ability is not fairly measured by any test involving language. But in a given community in which all children have approximately the same educational opportunity, it is reasonable to assume that a pupil who progresses rapidly in school and learns much has greater mental ability for his age than one who progresses less rapidly and learns less. To this extent, therefore, certain achievement questions such as vocabulary and arithmetic-reasoning questions, even though depending on language, do measure mental ability.

Alternative Forms

There are at present two forms of the Beta Test (Forms A and B), similar in construction and in difficulty but different in content. Two other forms, C and D, are in preparation.

Special Features

The tests are self-administering in the same sense as the Self-Administering Tests of Mental Ability, in that it is necessary merely to pass out the booklets, allow the pupils time to study the first page with a minimum of directions, and then let them go ahead and take the test. A single examiner may administer the tests to all the classes of a moderate-sized school in a day, by devoting a few minutes to start one class taking the test, leaving the class in care of the teacher, and going on to the next class, etc. This is a good way to assure reasonable uniformity of procedure in the giving of the tests.

In addition to the ease of administration which these tests afford by virtue of their single time limit, a new method of scoring is provided by which the tests may be scored even more rapidly than the Self-Administering Tests.

It will be observed that provision is made for the pupils to indicate their answers by putting crosses in circles, that when taking the test the circles corresponding to each item are directly opposite the items to avoid any possibility of a pupil putting a cross in the wrong row of circles, and that when the test is opened up, the four columns of circles all show at once. This enables the scorer to score the paper with one application of the scoring Key.

Moreover, the Key has holes in it, so placed that when it is properly adjusted over the test paper, the crosses that the pupil has put in the right circles of the test paper will show through the holes in the Key.

To score the paper, therefore, it is necessary merely to count the crosses that appear through the holes in the Key. Experience shows that this is the quickest possible method of scoring a test "by hand," so to speak. Its principal advantage is that the scorer does not have to look at each answer to see whether the cross is in or not in a given square or circle — he disregards all wrong answers completely and merely counts right ones.

It is by reason of this new scoring feature that the tests are called "Quick-Scoring Tests."

Directions for Administering

To administer Form A or Form B of the Beta Test, address the pupils as follows: (Give all directions slowly and distinctly, with a pause after each sentence.)

" We are now going to give you some tests that measure your ability to think. I will pass out the test papers and as soon as you receive one, read the first page and do what it tells you to do ; that is, fill the blanks, giving your name, age, etc., and read the sample questions.

" Do not open or turn over the booklet. Part of the test is to see if you can follow directions."

Have the test papers passed, one to each pupil, right side up ; that is, with the title page up. See that every pupil is supplied with two pencils and an eraser. It is better not to have the pencils too sharp, principally because it is desired to have the pupils make wide marks, since these are easy to see.

Allow reasonable time for all to finish reading the first page and studying the samples. A few laggards may be disregarded.

Then say: **"As it says on the first page of the booklet, the test contains 80 questions. You are not expected to be able to answer all of them, but do the best you can.**

"You will be allowed half an hour for the test. Try to get as many right as possible.

"Be careful not to go so fast that you make mistakes. Do not spend too much time on any one question.

"No questions about the test will be answered after the test begins.

"Make your crosses heavy so that they can be easily seen and *be sure not to put more than one cross in any row of circles.*

"Is there anyone who does not understand the first page?" This is the time to answer any questions that the pupils may have about the test, and the examiner should be satisfied that the pupils understand the samples and how to put a cross in the proper circle so as to indicate the correct answer to each item.

Then say: **"Now turn the paper over. Open the flap at the right so that you can see the rows of circles in which you are to put the answers for page 1.**

"As soon as you finish page 1, you are to open the booklet and do pages 2, 3, and 4 in the same way.

"Now take your pencils and begin."

No further directions are necessary.

Note the exact time immediately and write it on the blackboard, together with the time it will be in exactly one half hour, when the pupils are to stop work. Or set the hands of your watch exactly on the hour and stop the work when your watch is at exactly half-past the hour.

It should be understood by the examiner (and by the teacher if the teacher is left in charge of the pupils while they are taking the test) that no questions about the test are to be answered which might give the pupils the slightest help in answering the questions; that is, the examiner or teacher may not explain the meaning of any word or give any hints. It is permissible at the beginning of the examination for the examiner or teacher to move quietly about the room to make sure that the pupils are indicating their answers in the proper manner, and if during the examination a pupil becomes confused on account of the unusual folding of the booklet, it is permissible, of course, to explain to him how to proceed. Thereafter it is better for the teacher to remain seated at her desk so that the room is quiet and the pupils may work undisturbed.

The one in charge of timing the test should be particularly impressed with the need to watch the time carefully, for it is very easy to forget the time and let the pupils work more than half an hour. When the pupils have worked exactly half an hour, the examiner or teacher should say: **"Time is up. Everyone stop. Close the test booklet."** The test papers should then be collected.

Directions for Scoring

A Key for scoring the test is included in each package of tests. In preparing to score the papers of a class, each paper should be opened by picking it up by the flap and then laid in a pile with the four columns of answers showing.

To score a test paper, lay the Key over the paper in such a way that the heavy circles which are printed with crosses in them at the top of the test appear

through the appropriate holes at the top of the Key. The Key will then be adjusted so that all the crosses that the pupils have made in the right circles will show through the holes in the Key. It is necessary then merely to count the crosses that appear through the holes. The number of crosses so appearing is the pupil's score. This should be written in the space provided at the top of the title page.

The pupils have been instructed to be sure not to put more than one cross in any row of circles. However, if in the case of any item two crosses have been put in the same row of circles, no credit is given for that item.

There is no need to mark the answers right or wrong in this test, but merely to count the right answers, for only the total score is of significance.

It is not necessary to fold up the booklets completely after scoring. It will be found convenient, as each paper is scored, to turn over the right-hand page, by lifting up the flap, in order to write the score on the title page and lay the paper aside without entirely refolding it, for when the scores have been transcribed from the test papers onto the Class Record, the papers may not need to be consulted again.

In the interest of accuracy it is well for each paper to be scored independently by two persons. If this is done, the score obtained by the first scorer may be written at the foot of the column of answers for page 1 without turning over the page and the booklet may be left opened out flat. Then, after the next scorer has scored the paper and compared his count with that made by the first scorer and found it to check, the page may be turned and the checked score written on the title page.

If it is not possible for two persons to score the papers, it is advisable for the scorer to check his count of correct answers by counting the circles without crosses in them to see that the sum is 80. (If the number right is 40, record it, then continue counting, 41, 42, etc.)

Directions for Recording Scores

In each package of tests there is included a Class Record which provides for the recording of scores of a class.

Before entering the scores, arrange the papers either in alphabetical order or in order of magnitude of score, according to preference. Then enter the name of each pupil, his age in years and months, and his score.

Note that provision is made on the Class Record for entering later the IQ of the pupil and any additional data, such as percentile rank in the class or school, classification designation, etc., and for entering the median age, median score, etc., if desired.

Provision is made at the foot of page 2 of the Class Record for distributing the scores of a class or a school. To distribute the scores of a class, make a mark in the second column of the table for each pupil's score, putting the mark opposite the interval within which the score falls. Thus, if the first pupil has made a score of 63, put a mark opposite 60–64. Draw each fifth mark across the preceding four like this, 𝍸 𝍸. This makes it easier to count the marks.

After the marks are all entered, count those in each interval and write the number in the column headed "Freq." (Frequency).

To find the median (middle) score, count from either end of the distribution to the middle mark. If the middle mark falls, say, in the interval 50–54, sort out

the papers whose scores fall in this interval, and, if the median is the third mark in the interval, find the score on the third paper in that bunch of papers. That score is the median score of the class. (See Chapter II of Otis: *Statistical Method in Educational Measurement*,[1] or a similar text, for detailed explanations of other methods for finding the median.)

Reporting to the Author

To assist in making the norms more comprehensive, the author would appreciate the favor of receiving from each school system using 100 tests or more the following data for each grade:

Test used (Beta)
Form used (A or B)
Grade
Date of the test
Median Age (when each age has been recorded in years and months)
Median Score

That is, the author wishes to know the median age in years and months and the median score of all the pupils in the school system who are in the fourth grade, the same for all who are in the fifth grade, etc., to the ninth grade (whatever grades were tested). Address Dr. Arthur S. Otis, care of World Book Company, Yonkers, New York. This courtesy will be appreciated.

Distributions of Scores

Table 1 shows the distributions of scores by ages of 12,983 sixth-grade pupils. About half are from a large city in Ohio and about half from towns and villages of New York State. The median age of these pupils was 12 years and 4 months and the median score 42 points.

The table is read as follows: The column headed 12 contains the distributions of scores of the 5017 sixth-grade pupils whose age last birthday was 12 years, and

TABLE 1

Distributions of Scores by Ages of 12,983 Sixth-Grade Pupils in the Otis Quick-Scoring Mental Ability Tests: Beta Test (*Tests given in June of 1936*)

Score Intervals	Age Last Birthday									Totals
	9	10	11	12	13	14	15	16	17	
75–79			2	1						3
70–74		1	21	11						33
65–69		7	90	53	5	2				157
60–64	1	17	243	130	18		1			410
55–59		37	475	342	28	13				895
50–54		62	692	605	106	32	2	1		1500
45–49		53	860	842	260	54	7	2		2078
40–44	2	43	818	1065	410	121	20	4		2483
35–39	1	30	593	914	482	162	27	3		2212
30–34		13	321	562	437	174	33	9	1	1550
25–29		12	149	293	255	134	24	11	1	879
20–24	1	5	62	132	140	75	22	8	1	446
15–19		2	21	50	82	53	12	4	2	226
10–14	1	2	3	15	29	26	8	2		86
5–9		1		2	10	10				23
0–4							1	1		2
Totals	6	285	4350	5017	2262	856	157	45	5	12,983
Median age: 12 yr., 4 mo. Median score: 42										

[1] Published by World Book Company.

whose ages therefore range from 12 years to 13 years at the time of the test. It shows that, of those pupils, 1 made a score that fell in the interval 75–79, 11 made scores that fell in the interval 70–74, etc.

Table 2 shows the distributions of scores by ages of 2657 urban eighth-grade pupils in South Carolina who took the test in April, 1937. The median age of these pupils was 14 years 5 months and the median score, 46 points.

TABLE 2

DISTRIBUTIONS OF SCORES BY AGES OF 2657 EIGHTH-GRADE PUPILS IN THE BETA TEST
(*Given in April, 1937*)

SCORE INTERVALS	AGE LAST BIRTHDAY									TOTALS
	11	12	13	14	15	16	17	18	19	
80										
75–79										
70–74		1	16	4	1					22
65–69		4	49	25	2	1				81
60–64		10	96	58	14	3				181
55–59		14	143	130	31	22	6			336
50–54		14	159	137	63	25	6	4	1	409
45–49	1	13	152	166	95	36	12	4		479
40–44		13	128	145	103	45	8	6	1	449
35–39	1	6	68	115	87	42	16	6		341
30–34		4	24	71	47	41	14	5	3	209
25–29		1	14	19	25	20	9	4	1	93
20–24		2	5	4	7	8	6	1	1	34
15–19			1	2	3	4		2		12
10–14				4	3	1				8
5–9					1			1		2
0–4						1				1
Totals	2	82	855	880	482	239	77	33	7	2657
Median Age: 14 yr. 5 mo. Median Score: 46										

These tables are given partly in order to show what wide ranges of ages and ability are found in a single grade. Of course the average classroom does not show quite as wide a range of ages and scores, but nearly so. The need for dividing the pupils of such a grade into more homogeneous groups and the method of doing so are given below under the heading "Application of Results" (see page 11).

NORMS

If a large number of 12-year pupils take a test and the scores are arranged in order, the median or middle score is considered just normal for 12-year pupils and is said to be the *norm* for the age of 12 years. Table 3 gives the norms for the various ages of pupils taking Beta.

Table 3 is read as follows: The norm for the age of 8 years 0 months is 13 points of score; the norm for the age of 11 years 3 months is 36 points, etc.

The norms in Table 3 are based in part on the scores of 16,242 pupils in Beta and in part on a comparison of scores in Beta and scores in the Intermediate Examination of the Otis Self-Administering Tests of Mental Ability made by means of an experiment in which 3259 pupils in Grades 4 to 9 took Beta: Forms A and B, and Form A of the Intermediate Examination, in part on a comparison between Beta and Alpha, Nonverbal, in which 612 pupils in Grades 4 and 5 took both these tests, in part on a comparison between Beta and Gamma, in which

TABLE 3

REVISED (1939) AGE NORMS FOR BETA: FORMS A AND B

YEARS→		8	9	10	11	12	13	14	15	16	17	18 or over
MONTHS	0	13	20	27	34	40	45	49	52	54	56	57
	1	14	21	28	35	41	45	49	52	54	56	
	2	14	21	29	35	41	45	49	52	55	56	
	3	15	22	29	36	41	46	49	53	55	56	
	4	16	23	30	36	42	46	50	53	55	56	
	5	16	23	30	37	42	46	50	53	55	56	
	6	17	24	31	37	43	47	50	53	55	57	
	7	17	24	31	38	43	47	51	53	55	57	
	8	18	25	32	38	43	47	51	54	55	57	
	9	18	26	33	39	44	48	51	54	56	57	
	10	19	26	33	39	44	48	51	54	56	57	
	11	20	27	34	40	44	48	52	54	56	57	

742 pupils in Grades 7, 8, and 9 took these tests, and in part on a comparison between Beta and the Pintner General Ability Test (1661 scores).

These norms apply to a first test. If a pupil takes a second form of the test later, it is necessary to make a correction for familiarity with the test before using Table 3. (See "Practice Effect" below.)

PRACTICE EFFECT

When a pupil takes a second form of a test within a short time after the first form, he tends to make a better score on the second test. This increase in score is generally called "practice effect."

It was found that when a second form of Beta was given two days after the first form, the practice effect was about 4 points. This means that to render the second score of a pupil comparable to the first score if the tests were taken two days apart, 4 points should be subtracted from the second score.

Practice effect decreases, of course, as the length of time between tests increases. Possibly the amount of practice effect would drop to about 3 points if the interval were a week; to 2 points if the interval were a month; to 1 point if the interval were three months or more.

Whenever it is desired to find a Mental Age or IQ (see below) from the score of a pupil in a second test, the proper correction should be made for practice effect in the second score before comparing it with the norm for the pupil's age in Table 3 or before finding the pupil's Mental Age.

MENTAL AGES

Some examiners wish to express scores in terms of Mental Age. The term "Mental Age" originally meant the age for which a pupil's score was normal or median. Thus, if a pupil makes a score just normal or median for pupils 10 years old, he is said to have a Mental Age (MA) of 10 years.

This method of interpretation has a serious limitation, since mental growth slows down along with physical growth, and pupils reach a mental maturity in their teens. Thus the highest norm for any age in the Beta Test is 57 points, as shown in Table 3. This means that some pupils make scores that are above what

is normal for any age. In order to express degrees of mental ability which are above the norm for adults in terms of Mental Age, it is customary to proceed as though mental growth did not slow down but kept on increasing at approximately the same rate. According to this supposition, artificial mental ages are assigned to scores above age 13. This is called "extrapolation." This extrapolation method is used also with the Binet Scale.

According to the above method the following table of Mental Ages (Table 4) has been drawn up.

Table 4 is read as follows: A score of 1 in Beta: Form A or Form B, denotes a Mental Age of 6 years 5 months; a score of 61 may be treated as denoting a Mental Age of 16 years 0 months (though actually it is 4 points above the norm for adults).

TABLE 4

MENTAL AGES CORRESPONDING TO SCORES IN BETA: FORMS A AND B

SCORE	MA	SCORE	MA	SCORE	MA	SCORE	MA
1	6–5	21	9–2	41	12–3	61	16–0
2	6–6	22	9–3	42	12–5	62	16–2
3	6–7	23	9–5	43	12–8	63	16–4
4	6–9	24	9–7	44	12–10	64	16–6
5	6–11	25	9–8	45	13–0	65	16–8
6	7–0	26	9–10	46	13–3	66	16–10
7	7–2	27	10–0	47	13–5	67	17–0
8	7–4	28	10–1	48	13–8	68	17–2
9	7–5	29	10–3	49	13–11	69	17–4
10	7–7	30	10–5	50	14–2	70	17–6
11	7–9	31	10–7	51	14–4	71	17–8
12	7–10	32	10–8	52	14–6	72	17–10
13	8–0	33	10–10	53	14–8	73	18–0
14	8–2	34	11–0	54	14–10	74	18–2
15	8–3	35	11–2	55	15–0	75	18–4
16	8–5	36	11–4	56	15–2	76	18–6
17	8–7	37	11–6	57	15–4	77	18–8
18	8–9	38	11–8	58	15–6	78	18–10
19	8–10	39	11–10	59	15–8	79	19–0
20	9–0	40	12–0	60	15–10	80	19–2

MEASURING BRIGHTNESS

Pupils making the same score in the test are presumed to have the same mental ability or, as we say, the same Mental Age even though their actual ages (spoken of as "chronological ages") are not the same. That is, as explained above, a pupil who makes a score equal to the norm for the age of 10 years is said to have a Mental Age of 10 years, whether the pupil is 10 years old or 9 years old or 11 years old.

A 10-year pupil who has a Mental Age of 11 years is brighter than normal, and a measure of his brightness is often found by dividing his Mental Age of 11 years by his "chronological age" of 10 years ($11 \div 10 = 1.10$). The decimal point is then dropped and the 110 is called the pupil's Intelligence Quotient (IQ). Intelligence Quotients so found cluster most thickly around 100, but in a few instances go above 150 and below 50. They are distributed according to the "law of normal distribution."

A study of the dispersion of IQ's of various populations aggregating 100,000 pupils, tested by various group tests of mental ability, showed standard deviations

of IQ's ranging from 10 to 19 points of IQ for the various populations, the median value of the standard deviation of IQ's being between 15 and 16 points; hence theoretically about $\frac{1}{10}$ of 1% of pupils make IQ's of 150 or over, $\frac{1}{2}$ of 1% of pupils make IQ's of 142 or over, and so on as shown in Table 5.

TABLE 5

PER CENTS OF PUPILS MAKING VARIOUS IQ'S

This per cent of pupils	make these IQ's:	This per cent of pupils	make these IQ's:
$\frac{1}{10}$ of 1%	150 or over	$\frac{1}{10}$ of 1%	50 or less
$\frac{1}{2}$ of 1%	142 or over	$\frac{1}{2}$ of 1%	58 or less
1%	136 or over	1%	64 or less
5%	126 or over	5%	74 or less
10%	121 or over	10%	79 or less
25%	111 or over	25%	89 or less
$33\frac{1}{3}$%	107 or over	$33\frac{1}{3}$%	93 or less
50%	100 or over	50%	100 or less

A measure of brightness comparable to the IQ can be found from scores of pupils in the Beta Test according to the method below. Although the measures are not quotients, they are called "Beta IQ's" because they are comparable to IQ's.

HOW TO FIND A PUPIL'S "BETA IQ"

To find a pupil's "Beta IQ," proceed as follows:

1. Find the norm for the pupil's age from Table 3.
2. Find the amount by which the pupil's score exceeds (or falls below) the norm for his age. Call this his "deviation of score."
3. Add the pupil's deviation of score to 100 (or subtract from 100 if the deviation is downward). The result is the pupil's "Beta IQ."
4. If a pupil's score is above 70, it is to be augmented before proceeding with Steps 2 and 3 above. Treat a score of 71 as though it were 72. Treat a score of 72 as though it were 74, etc., according to Table 6.

TABLE 6

FOR AUGMENTING HIGH SCORES

Treat a score of	71	72	73	74	75	76	77	78	79	80
as though it were	72	74	76	78	80	82	84	86	88	90

As a sample of Step 4, suppose a pupil of 16 years 6 months makes a score of 75 in Form A. The norm for 16 years 6 months is 55. To find his deviation of score, treat the score of 75 as though it were 80, subtract 55 from 80 (answer 25), and add 25 to 100, yielding a "Beta IQ" of 125.

Various determinations of the dispersion of "Beta IQ's" yield standard deviations of "IQ" of from 10 to 17 points for various populations. It is believed that "Beta IQ's" tend to be *somewhat less* dispersed than IQ's obtained by the division method from group tests in general (that is, they tend to be somewhat nearer to 100). Therefore allowance should be made for this fact when comparing "Beta IQ's" with ordinary IQ's from other tests.

However, the above method is recommended as yielding measures of brightness that are more consistent and constant for a given individual than ordinary IQ's.

Reliability and Validity of the Beta Test

By "reliability" is meant the degree of precision with which a test measures what it measures.

One common measure of the reliability of a test is the coefficient of correlation between two forms of the test. Table 7 gives the coefficients of correlation between Forms A and B in Grades 4 to 9 of a large school system, the average number of pupils per coefficient being 86. The average of the 12 coefficients is .79. For Grades 4 to 9 combined the coefficient is .96.

TABLE 7

Reliability Coefficients (Form A *vs.* Form B)

	Grades						Grades Combined
	4	5	6	7	8	9	
A (1st)–B (2d)	.730	.979	.826	.711	.833	.665	.96
B (1st)–A (2d)	.764	.842	.859	.869	.688	.651	

Another measure of reliability is the coefficient of correlation between odd and even items of a single test. This is virtually a correlation between two forms of a short test each half as long as the full test, the two tests being given, we might say, simultaneously.

It is customary, then, to correct the coefficients of correlation between the half tests by the Spearman-Brown formula to obtain the corresponding coefficient for two full-length tests given under the same circumstances.

The coefficients of correlation for the odd and even items of one test are as shown in Table 8.

TABLE 8

Reliability Coefficients (Odd *vs.* Even Items) Corrected by Spearman-Brown Formula

Grades	4	5	6	7	8	9
Corrected coefficients	.81	.92	.90	.87	.86	.79

The average of the six corrected coefficients in Table 8 is .86, which is 7 points higher than .79, the average of the coefficients of Table 7. This deficiency of 7 points in the coefficients of Table 7 is due to the instability of the pupils themselves. That is if pupils remained as constant in ability from day to day as from moment to moment, so to speak, the coefficients in Table 7 would be as high as the coefficients in Table 8.

By validity of a test is meant the degree to which it measures the ability it is designed to measure. Or we might say, it is the degree to which it serves its purpose.

Now the purpose of the Beta Test is most generally that of finding the degree of brightness of a pupil; that is, obtaining some measure (such as the IQ) that indicates the probable rate of progress the pupil will make in school. This being the case, it follows that actual rate of progress of pupils through school is the most appropriate criterion of the validity of the Beta Test.

This criterion is the one that was used in the standardization of the Otis Intermediate Examination, from which most of the items of the Beta Test were taken. The method is described in the Manual for the Otis Self-Administering Tests of Mental Ability (page 3). The determination of the validity of each item consisted of comparing the number of passes of that item by a group of pupils who were making rapid progress through school with the number of passes of the item by a group of pupils who were making slow progress through school. Only those items were used which showed a distinct gain in number of passes of the rapid-progress pupils over the number of passes of the slow-progress pupils. Each item justified its inclusion, therefore, because it contributed definitely to the capacity of the test to measure brightness as reflected in rate of progress through school.

Probable Error of a Score

Another measure of reliability which is entirely independent of the degree of heterogeneity of the group is the *probable error* of a score. By "probable error of a score" is meant the median amount by which any pupil's actual score differs from his true score. While we do not know the true score of any pupil (by which is meant the average of a great many scores found under identical conditions), we can tell from the differences between scores pupils make in two forms what this probable error is.

In the case of 465 pupils in Grades 4 to 9 the median amount of difference between two scores of the same pupil was 3.8 points, from which it follows theoretically that the probable error of a score is 2.7 points. $(3.8 \div \sqrt{2} = 2.7)$

That is, a pupil's score will be in error only between 0 and 2.7 points in 50% of cases, and so on as shown in Table 9.

TABLE 9

Errors of Scores in Beta

In this per cent of cases	the pupil's score will probably be in error
50%	between 0 and 2.7 points
32%	between 2.7 and 5.4 points
16%	between 5.4 and 8.1 points
2%	over 8.1 points

Application of Results

Purposes of mental ability tests. The principal purposes for which mental tests are given are these:

1. For teaching purposes, to discover which pupils are bright and capable of doing better school work than they are doing and to discover which pupils are dull and may be attempting work beyond their capacity.

2. For administrative purposes, to regrade pupils so that the pupils in any one grade will be more homogeneous in mental ability and therefore able to progress at more nearly the same rate than otherwise.

3. For administrative purposes, to classify pupils into separate groups within grades in order that the brighter or the more mature pupils may be given an enriched curriculum and in order that the duller or the less mature pupils may be allowed to progress at a slower rate.

Such classifying is sometimes done on the basis of score (dividing the pupils on the basis of mental maturity) and sometimes on the basis of IQ (dividing the pupils on the basis of brightness). The first of these methods is recommended.

4. For research purposes, to obtain two or more groups of equal mental ability or brightness which may be given different methods of instruction for the purpose of determining which method is superior.

5. For guidance purposes, to assist pupils to choose wisely in planning their educational, recreational, and vocational programs.

6. For administrative purposes, to determine the comparative mental status of pupils of different schools or localities.

Distributing scores. For any one of the purposes mentioned above it is desirable to distribute the scores of a class. This is usually done by finding the intervals 0–4, 5–9, etc., into which the scores fall. Provision is made for so distributing the scores of a class on the Class Record, a copy of which is enclosed in each package of tests.

Classifying pupils according to score. If desired to divide the pupils of a grade into classes according to score, the scores of all the pupils of the grade may be entered in one distribution on a Class Record or the test papers may be arranged in order of score. The scores may then be divided into an upper third, middle third, and lower third, or in any other convenient way, and the pupils classified accordingly.

It will be found that pupils so grouped are much more alike in their ability to learn than the pupils of the whole group and can be taught together much more easily.

ACKNOWLEDGMENTS

Thanks are due to Dr. Leon N. Neulen, Superintendent of Schools, Camden, New Jersey, and to the teachers of Camden; to Dr. W. C. McCall, University Personnel Bureau, University of South Carolina; to Dr. William L. Connor, formerly head of the Bureau of Research at Cleveland; and to Dr. Leo J. Brueckner of the University of Minnesota, for kind coöperation in furnishing scores for standardizing the Beta Test.

Thanks are due also to Dr. A. L. Maxon, Director of Research, Department of Public Instruction, Schenectady, New York, for coöperation in the standardization of Beta, Form CM.

OTIS QUICK-SCORING MENTAL ABILITY TESTS

By ARTHUR S. OTIS, PH.D.

Formerly Development Specialist with Advisory Board, General Staff, United States War Department

Beta
DM

BETA TEST: FORM DM

IQ..... For Grades 4–9 Score.....

Read this page. Do what it tells you to do.

Do not open this booklet, or turn it over, until you are told to do so.
Fill these blanks, giving your name, age, birthday, etc. Write plainly.

Name..... Age last birthday .....years
(First name, initial, and last name)

Birthday..... Teacher..... Date.....19.....
(Month, Day)

Grade..... School..... City and state.....

This is a test to see how well you can think. It contains questions of different kinds. Here are three sample questions. Five answers are given under each question. Read each question and decide which of the five answers below it is the right answer.

Sample a: Which one of the five things below is soft?

① glass ② stone ③ cotton ④ iron ⑤ ice.....

1 2 3 4 5

The right answer, of course, is *cotton;* so the word *cotton* is underlined. And the word *cotton* is No. 3; so a heavy mark has been put in the space under the 3 at the right. This is the way you are to answer the questions.

Try the next sample question yourself. Do not write the answer; just draw a line under it and then put a heavy mark in the space under the right number.

Sample b: A robin is a kind of —

⑥ plant ⑦ bird ⑧ worm ⑨ fish ⑩ flower....

6 7 8 9 10

The answer is *bird;* so you should have drawn a line under the word *bird*, and *bird* is No. 7; so you should have put a heavy mark in the space under the 7. Try this one:

Sample c: Which one of the five numbers below is larger than 55?

⑪ 53 ⑫ 48 ⑬ 29 ⑭ 57 ⑮ 16.....

11 12 13 14 15

The answer, of course, is 57; so you should have drawn a line under 57, and that is No. 14; so you should have put a heavy mark in the space under the 14.

The test contains 80 questions. You are not expected to be able to answer all of them, but do the best you can. You will be allowed half an hour after the examiner tells you to begin. Try to get as many right as possible. Be careful not to go so fast that you make mistakes. Do not spend too much time on any one question. No questions about the test will be answered by the examiner after the test begins. Lay your pencil down.

Do not turn this booklet until you are told to begin.

PRINTED IN U.S.A. BETA:DM 20

ANSWER SHEET

Otis Quick-Scoring : Beta

Page 6

64	61	62	63	64	65
65	66	67	68	69	70
66	71	72	73	74	75
67	1	2	3	4	5
68	6	7	8	9	10
69	11	12	13	14	15
70	16	17	18	19	20
71	21	22	23	24	25
72	26	27	28	29	
73	31	32	33	34	35
74	36	37	38	39	40
75	41	42	43	44	45
76	46	47	48	49	50
77	51	52	53	54	55
78	56	57	58	59	60
79	61	62	63	64	65
80	66	67	68	69	70

Page 5

45	41	42	43	44	45
46	46	47	48	49	
47	51	52	53	54	55
48	56	57	58	59	
49	61	62	63	64	65
50	66	67	68	69	
51	71	72	73	74	75
52	1	2	3	4	5
53	6	7	8	9	10
54	11	12	13	14	15
55	16	17	18	19	20
56	21	22	23	24	25
57	26	27	28	29	30
58	31	32	33	34	35
59	36	37	38	39	
60	41	42	43	44	45
61	46	47	48	49	50
62	51	52	53	54	
63	56	57	58	59	

Page 4

25	26	27	28	29	30
26	31	32	33	34	35
27	36	37	38	39	40
28	41	42	43	44	45
29	46	47	48	49	50
30	51	52	53	54	
31	56	57	58	59	60
32	61	62	63	64	65
33	66	67	68	69	
34	71	72	73	74	75
35	76	77	78	79	80
36	81	82	83	84	85
37	1	2	3	4	5
38	6	7	8	9	10
39	11	12	13	14	15
40	16	17	18	19	
41	21	22	23	24	25
42	26	27	28	29	30
43	31	32	33	34	35
44	36	37	38	39	40

Page 3

1	1	2	3	4	5
2	6	7	8	9	10
3	11	12	13	14	15
4	16	17	18	19	20
5	21	22	23	24	
6	26	27	28	29	30
7	31	32	33	34	35
8	36	37	38	39	
9	41	42	43	44	45
10	46	47	48	49	50
11	51	52	53	54	55
12	56	57	58	59	60
13	61	62	63	64	65
14	66	67	68	69	70
15	71	72	73	74	75
16	76	77	78	79	
17	81	82	83	84	85
18	86	87	88	89	90
19	91	92	93	94	95
20	1	2	3	4	5
21	6	7	8	9	10
22	11	12	13	14	15
23	16	17	18	19	20
24	21	22	23	24	25

NOTE. *This Answer Sheet is not intended for machine scoring.*

Examination begins here.

1. Which one of the five things below does not belong with the others?
   ① horse ② dog ③ camel ④ fish ⑤ bear

2. Which one of the five answers below tells best what a gun is?
   ⑥ shoot ⑦ a weapon ⑧ a tool ⑨ an apparatus ⑩ a thing

3. Which one of the five words below means the opposite of east?
   ⑪ north ⑫ pole ⑬ west ⑭ equator ⑮ south

4. A hat is to a head and a glove is to a hand as a shoe is to what?
   ⑯ leather ⑰ a foot ⑱ a shoestring ⑲ walk ⑳ a toe

5. A child who has accidentally broken another child's toy should —
   ㉑ say he didn't do it ㉒ say nothing ㉓ throw it away ㉔ say, "I'm sorry"

6. Which one of the five things below is the smallest?
   ㉖ an ankle ㉗ a leg ㉘ a toe ㉙ a knee ㉚ a foot

7. Which word means the opposite of fail?
   ㉛ lose ㉜ succeed ㉝ rise ㉞ recede ㉟ give up

8. Three of the four designs at the right are alike. Which one is not like the other three?
   ㊱  ㊲  ㊳  ㊴ 

9. Which one of the five things below is most like these three: a violin, a radio, a harp?
   ㊶ music ㊷ a chair ㊸ a stove ㊹ a piano ㊺ a bow

10. An elbow is to an arm as a knee is to what?
    ㊻ an ankle ㊼ a leg ㊽ trousers ㊾ a bone ㊿ a man

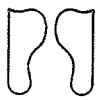
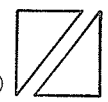
11. What word means the opposite of comfort?
    (51) uncomfort (52) hard (53) sickness (54) discomfort (55) painful

12. If you are sure you are right, you have —
    (56) pride (57) doubt (58) safety (59) confusion (60) confidence

13. Which answer tells best just what a gate is?
    (61) a hole in a fence (62) something to swing on (63) It has hinges.
    (64) a door in a fence (65) It opens and shuts

14. A parasol is to sunshine as an umbrella is to what?
    (66) the sun (67) rain (68) night (69) winter (70) black

15. At 6 cents each, how many pencils can be bought for 48 cents?
    (71) 48 (72) 54 (73) 42 (74) 8 (75) 288

16. Three of the four designs at the right are alike. Which one is not like the other three?
    (76)  (77)  (78)  (79) 

17. Which one of the five things below is most like these three: an apple, a peach, and a pear?
    (81) a seed (82) a tree (83) a plum (84) a bud (85) a peel

18. Feathers are to a bird as fur is to what?
    (86) a coat (87) a rabbit (88) a swan (89) a glove (90) an ostrich

19. If the words below were rearranged to make a good sentence, with what letter would the last word of the sentence begin?
    nuts from squirrels trees the gather
    (91) s (92) f (93) t (94) n (95) g

20. Which one of the five words below means the opposite of easy?
    ① simple ② difficult ③ tough ④ slow ⑤ baffling

21. If a person walking in a quiet place suddenly hears a loud sound, he is likely to be —
    ⑥ stopped ⑦ struck ⑧ made deaf ⑨ startled ⑩ angered

22. One number is wrong in the following series: 6 1 6 2 6 3 6 4 6 5 6 7
    What should that number be? ⑪ 8 ⑫ 6 ⑬ 7 ⑭ 4 ⑮ 5

23. Which one of the five things below is most like these three: a snake, a cow, and a sparrow?
    ⑯ a tree ⑰ a doll ⑱ a feather ⑲ a pig ⑳ a skin

24. Steam is to a locomotive as what is to a sailboat?
    ㉑ the ocean ㉒ a whistle ㉓ a rudder ㉔ the wind ㉕ a mast

(Go right on to the next page.)

64. If the following words were arranged in order, which word would be in the middle?
(61) Youth (62) Infancy (63) Manhood (64) Childhood (65) Birth

65. Which tells best just what a foot is?
(66) to wear a shoe and stocking on (67) Both feet are the same size.
(68) It has five toes and a heel. (69) the part of the body on which an animal stanc
(70) Men have larger feet than women.

66. A statement which expresses just the opposite of that which another statement expresses is said to be a —
(71) lie (72) contradiction (73) falsehood (74) correction (75) explanation

67. Which one of the words below would come last in the dictionary?
(1) graft (2) leader (3) lively (4) gallop (5) know

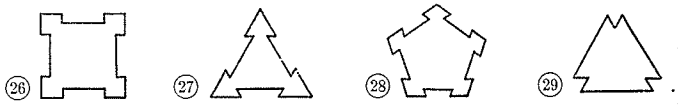
68. What is the letter that precedes the letter that comes just before O in the alphabet?
(6) O (7) N (8) M (9) P (10) Q

69. One number is wrong in this series. 1 2 4 8 16 36 64 What should that number
(11) 6 (12) 12 (13) 13 (14) 32 (15) 24

70. If I have a large box with 3 small boxes in it and 4 very small boxes in each of the small boxes, how many boxes are there in all?
(16) 8 (17) 12 (18) 13 (19) 15 (20) 16

71. There is a saying, "All is not gold that glitters." This means that —
(21) Some gold has a dull finish. (22) Appearances are sometimes deceptive.
(23) Diamonds sparkle more than gold. (24) Don't wear cheap jewelry.
(25) Some people like to make a show of wealth.

72. Three of the four designs at the right are alike. Which one is not like the other three?



73. If a photograph that is 4 in. long and 3 in. wide is enlarged to be 20 in. long, how many inches wide will it be?
(31) 19 (32) 18 (33) 15 (34) $6\frac{2}{3}$ (35) 5

74. One number is wrong in this series.
4 5 8 9 12 13 16 18 20 21
What should that number be?
(36) 6 (37) 7 (38) 10 (39) 17 (40) 14

75. When the time by a clock was 2 minutes past 4, the hands were interchanged. The clock then said about —
(41) 4 minutes past 2 (42) 2 minutes past 4 (43) 20 minutes past 12
(44) 2 minutes of 4 (45) 20 minutes of 12

76. A car owner uses a mixture in his radiator containing 1 quart of alcohol to every 2 quarts of water. How many quarts of alcohol are needed for 21 quarts of the mixture?
(46) $11\frac{1}{2}$ (47) 42 (48) 20 (49) 14 (50) 7

77. What letter in the following series appears a third time nearest the beginning?
A C E B A D D C E F B E D A C C B I
(51) A (52) C (53) E (54) D (55) B

78. In a foreign language *para misa tela* means *very hard ground*, *para fola* means *soft ground*, and *misa roga* means *very many*. What word means hard?
(56) para (57) misa (58) tela (59) fola (60) roga

79. Which one of the five words below does not belong with the others?
(61) efficiency (62) authority (63) accuracy (64) utility (65) durability

80. A boy is now three times as old as his sister. In 5 years he will be only twice as old. How many years old is he now?
(66) 9 (67) 6 (68) 3 (69) 15 (70) 12

45. A library is to books as (?) is to money.
(41) a store (42) a school (43) knowledge (44) a bank (45) gold

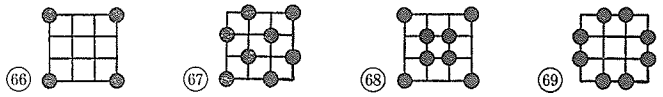
46. There is a saying, "All's well that ends well." This means that —
(46) All comes out well in the end.
(47) The success of anything is judged by the final result.
(48) Stick to a job until it is finished. (49) Don't worry about how things will turn out.

47. If the following words were arranged in order, which word would be in the middle?
(51) foot (52) inch (53) mile (54) yard (55) rod

48. If Harry is older than William and William is just as old as Charles, then Charles is (?) Harry.
(56) older than (57) younger than (58) just as old as (59) (cannot say which)

49. A pitcher is to milk as (?) is to flowers.
(61) a stem (62) a leaf (63) a vase (64) water (65) a root

50. Three of the four designs at the right are alike. Which one is not like the other three?



51. The feeling of a father for his children is usually —
(71) contempt (72) affection (73) joy (74) pity (75) reverence

52. Which tells best just what a lie is?
(1) a mistake (2) an exaggeration (3) an accidental false statement
(4) a malicious false statement (5) a wrong answer

53. Wood is to a table as (?) is to a knife.
(6) cutting (7) a chair (8) a fork (9) a handle (10) steel

54. If the words below were rearranged to make a good sentence, the third word of the sentence would begin with what letter?
cook chocolate the a cake made layer
(11) t (12) c (13) l (14) m (15) a

55. Which word means the opposite of guilty?
(16) tarnished (17) brave (18) unselfish (19) cordial (20) innocent

56. If a man has walked east from his home 8 blocks and then walked west 3 blocks, how many blocks is he from home?
(21) 11 (22) 8 (23) 3 (24) 4 (25) 5

57. If an act conforms to recognized principles or standards, it is said to be —
(26) legislative (27) wicked (28) legitimate (29) harmonious (30) wrong

58. A captain is to a ship as a mayor is to what?
(31) a state (32) a city (33) a council (34) a boss (35) a lawyer

59. There is a saying, "People who live in glass houses should not throw stones." This means that —
(36) The stones thrown are likely to break the glass in the houses. (37) People should not live in glass houses. (38) Those who have faults should not criticize others.
(39) People who live in glass houses need all the stones they have.

60. Which of the five words below is most like these three: large, red, good?
(41) very (42) size (43) color (44) apple (45) heavy

61. A revolver is to a man as what is to a bee?
(46) a wing (47) honey (48) flying (49) a sting (50) wax

62. There is a saying, "Birds of a feather flock together." This means that —
(51) Birds fly in large flocks. (52) People associate with others like themselves.
(53) Birds in a flock have the same color.
(54) People settle in cities to be near others

63. Three of the four designs at the right are alike. Which one is not like the other three?

(Go right on to the next page

25. Which one of the words below would come first in the dictionary?
(26) march (27) ocean (28) horse (29) elbow (30) paint

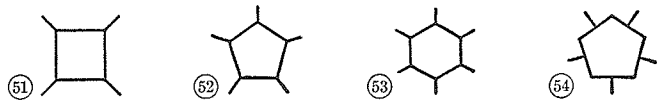
26. An automobile is to a wagon as a motorcycle is to what?
(31) a bicycle (32) a horse (33) a buggy (34) a train (35) walking

27. Which tells best just what a horse is?
(36) a large, four-legged animal (37) It has a tail.
(38) a thing that works and eats (39) a live thing (40) something to pull a wagon.

28. Which of these series contains a wrong number?
(41) 3–6–9–12–15 (42) 2–5–8–11–14 (43) 1–3–5–7–8 (44) 2–4–6–8–10
(45) 1–4–7–10–13

29. Which one of the five things below is most like these three: a skate, a baseball, and a jump rope?
(46) a shoe (47) a club (48) a scooter (49) a string (50) a hammer

30. Three of the four designs at the right are alike. Which one is not like the other three?



31. If the words below were rearranged to make a good sentence, the first word of the sentence would begin with what letter?
heavier lead cork is than
(56) t (57) h (58) c (59) i (60) l

32. A hospital is to the sick as what is to criminals?
(61) a doctor (62) an asylum (63) a judge (64) a prison (65) criminals

33. If George is older than Frank and Frank is older than James, then George is (?) James.
(66) younger than (67) older than (68) just as old as (69) (cannot say which).

34. Count each 6 below that has a 9 next after it.
Tell how many 6's you count.
9 6 4 6 9 3 4 9 6 7 9 9 3 6 9 4 5 9 9 6 3 1 9 6 9 0 4 9 3 6 2 9 1 7 6 9
(71) 6 (72) 2 (73) 3 (74) 4 (75) 5

35. An event which might happen is said to be —
(76) doubtful (77) possible (78) certain (79) probable (80) unreasonable

36. The daughter of my mother's brother is my —
(81) niece (82) aunt (83) cousin (84) stepsister (85) granddaughter

37. Better is to good as worse is to what?
(1) very good (2) bad (3) medium (4) much worse (5) best

38. A government in which there are graft and bribery is said to be —
(6) anarchistic (7) corrupt (8) autocratic (9) inefficient (10) disorganized.

39. If the following words were arranged in order, the middle word would begin with what letter?
One Six Ten Eight Four Nine Zero
(11) 0 (12) E (13) S (14) F (15) N

40. If Harry is shorter than William and Harry is taller than Charles, then Charles is (?) William.
(16) shorter than (17) taller than (18) just as tall as (19) (cannot say which).

41. Which tells best what a wheel is?
(21) something that turns (22) It goes round.
(23) a circular rim and hub connected by spokes
(24) a round thing to put on an automobile (25) A bicycle always has two of them.

42. Which one of the five things below is most like these three: a king, a general, and a dictator?
(26) a war (27) a servant (28) a command (29) a president (30) a monarchy.

43. What is the most important reason that glass is used in windows?
(31) It is cheaper than wood. (32) It permits light to pass through the window.
(33) It keeps out the rain and snow. (34) It does not collect dust and germs.
(35) The people inside can watch their friends go by outside

44. Which one of the words below would come first in the dictionary?
(36) bully (37) button (38) broad (39) brass (40) breakable

(Go right on to the nex

Tables for Deriving IQ's on

OTIS QUICK-SCORING MENTAL ABILITY TEST

BETA: FORM C_M or D_M

Directly from Score and Chronological Age

This three-page table provides a means of getting IQ's on Forms C_M and D_M of the Beta test more simply than by the three-step process described in the Manual of Directions. (The two methods result in identical IQ's.)

To get a student's IQ: Note the column whose heading includes his CA and the row corresponding to his obtained score. Read off his IQ from the intersection of this row and column. (Illustration: CA = 9 – 11, Score = 31. The IQ is 107.)

If a large number of IQ's are to be obtained, it will save time to sort the answer sheets according to the CA ranges shown in the column headings, then get all the IQ's for one CA-column before proceeding to the next. (Folding the table to bring the Score scale adjacent to each successive CA-column is suggested.)

When the number of tests is exceedingly large, it will pay to make a second sort, that by score within each CA-grouping. The IQ for a given CA – Score combination is obtained only once, then recorded on all appropriate answer sheets.

TEST DEPARTMENT
HARCOURT, BRACE & WORLD, INC.

IQ's on Otis Quick-Scoring Mental Ability Test: Beta, Form C_M or D_M
from pupil's score and chronological age

Score ↓ CA →	8–0	8–1 8–2	8–3	8–4 8–5	8–6 8–7	8–8 8–9	8–10	8–11 9– 0	9–1 9–2	9–3	9–4 9–5	9–6 9–7	9–8	9– 9 9–10	9–11 10– 0	10–1
80															164	163
79															162	161
78													164	163	160	159
77											164	163	162	161	159	157
76									164	163	162	161	160	159	158	156
75							164	163	162	161	160	159	158	157	156	155
74					164	163	162	161	160	159	158	157	156	155	154	153
73			164	163	162	161	160	159	158	157	156	155	154	153	152	151
72	164	163	162	161	160	159	158	157	156	155	154	153	152	151	150	149
71	162	161	160	159	158	157	156	155	154	153	152	151	150	149	148	147
70	160	159	158	157	156	155	154	153	152	151	150	149	148	147	146	145
69	159	158	157	156	155	154	153	152	151	150	149	148	147	146	145	144
68	158	157	156	155	154	153	152	151	150	149	148	147	146	145	144	143
67	157	156	155	154	153	152	151	150	149	148	147	146	145	144	143	142
66	156	155	154	153	152	151	150	149	148	147	146	145	144	143	142	141
65	155	154	153	152	151	150	149	148	147	146	145	144	143	142	141	140
64	154	153	152	151	150	149	148	147	146	145	144	143	142	141	140	139
63	153	152	151	150	149	148	147	146	145	144	143	142	141	140	139	138
62	152	151	150	149	148	147	146	145	144	143	142	141	140	139	138	137
61	151	150	149	148	147	146	145	144	143	142	141	140	139	138	137	136
60	150	149	148	147	146	145	144	143	142	141	140	139	138	137	136	135
59	149	148	147	146	145	144	143	142	141	140	139	138	137	136	135	134
58	148	147	146	145	144	143	142	141	140	139	138	137	136	135	134	133
57	147	146	145	144	143	142	141	140	139	138	137	136	135	134	133	132
56	146	145	144	143	142	141	140	139	138	137	136	135	134	133	132	131
55	145	144	143	142	141	140	139	138	137	136	135	134	133	132	131	130
54	144	143	142	141	140	139	138	137	136	135	134	133	132	131	130	129
53	143	142	141	140	139	138	137	136	135	134	133	132	131	130	129	128
52	142	141	140	139	138	137	136	135	134	133	132	131	130	129	128	127
51	141	140	139	138	137	136	135	134	133	132	131	130	129	128	127	126
50	140	139	138	137	136	135	134	133	132	131	130	129	128	127	126	125
49	139	138	137	136	135	134	133	132	131	130	129	128	127	126	125	124
48	138	137	136	135	134	133	132	131	130	129	128	127	126	125	124	123
47	137	136	135	134	133	132	131	130	129	128	127	126	125	124	123	122
46	136	135	134	133	132	131	130	129	128	127	126	125	124	123	122	121
45	135	134	133	132	131	130	129	128	127	126	125	124	123	122	121	120
44	134	133	132	131	130	129	128	127	126	125	124	123	122	121	120	119
43	133	132	131	130	129	128	127	126	125	124	123	122	121	120	119	118
42	132	131	130	129	128	127	126	125	124	123	122	121	120	119	118	117
41	131	130	129	128	127	126	125	124	123	122	121	120	119	118	117	116
40	130	129	128	127	126	125	124	123	122	121	120	119	118	117	116	115
39	129	128	127	126	125	124	123	122	121	120	119	118	117	116	115	114
38	128	127	126	125	124	123	122	121	120	119	118	117	116	115	114	113
37	127	126	125	124	123	122	121	120	119	118	117	116	115	114	113	112
36	126	125	124	123	122	121	120	119	118	117	116	115	114	113	112	111
35	125	124	123	122	121	120	119	118	117	116	115	114	113	112	111	110
34	124	123	122	121	120	119	118	117	116	115	114	113	112	111	110	109
33	123	122	121	120	119	118	117	116	115	114	113	112	111	110	109	108
32	122	121	120	119	118	117	116	115	114	113	112	111	110	109	108	107
31	121	120	119	118	117	116	115	114	113	112	111	110	109	108	107	106
30	120	119	118	117	116	115	114	113	112	111	110	109	108	107	106	105
29	119	118	117	116	115	114	113	112	111	110	109	108	107	106	105	104
28	118	117	116	115	114	113	112	111	110	109	108	107	106	105	104	103
27	117	116	115	114	113	112	111	110	109	108	107	106	105	104	103	102
26	116	115	114	113	112	111	110	109	108	107	106	105	104	103	102	101
25	115	114	113	112	111	110	109	108	107	106	105	104	103	102	101	100
24	114	113	112	111	110	109	108	107	106	105	104	103	102	101	100	99
23	113	112	111	110	109	108	107	106	105	104	103	102	101	100	99	98
22	112	111	110	109	108	107	106	105	104	103	102	101	100	99	98	97
21	111	110	109	108	107	106	105	104	103	102	101	100	99	98	97	96
20	110	109	108	107	106	105	104	103	102	101	100	99	98	97	96	95
19	109	108	107	106	105	104	103	102	101	100	99	98	97	96	95	94
18	108	107	106	105	104	103	102	101	100	99	98	97	96	95	94	93
17	107	106	105	104	103	102	101	100	99	98	97	96	95	94	93	92
16	106	105	104	103	102	101	100	99	98	97	96	95	94	93	92	91
15	105	104	103	102	101	100	99	98	97	96	95	94	93	92	91	90
14	104	103	102	101	100	99	98	97	96	95	94	93	92	91	90	89
13	103	102	101	100	99	98	97	96	95	94	93	92	91	90	89	88
12	102	101	100	99	98	97	96	95	94	93	92	91	90	89	88	87
11	101	100	99	98	97	96	95	94	93	92	91	90	89	88	87	86
10	100	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85
9	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84
8	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83
7	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82
6	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81
5	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80
4	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79
3	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78
2	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77
1	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76

IQ's on Otis Beta, Form C_M or D_M (Cont'd)

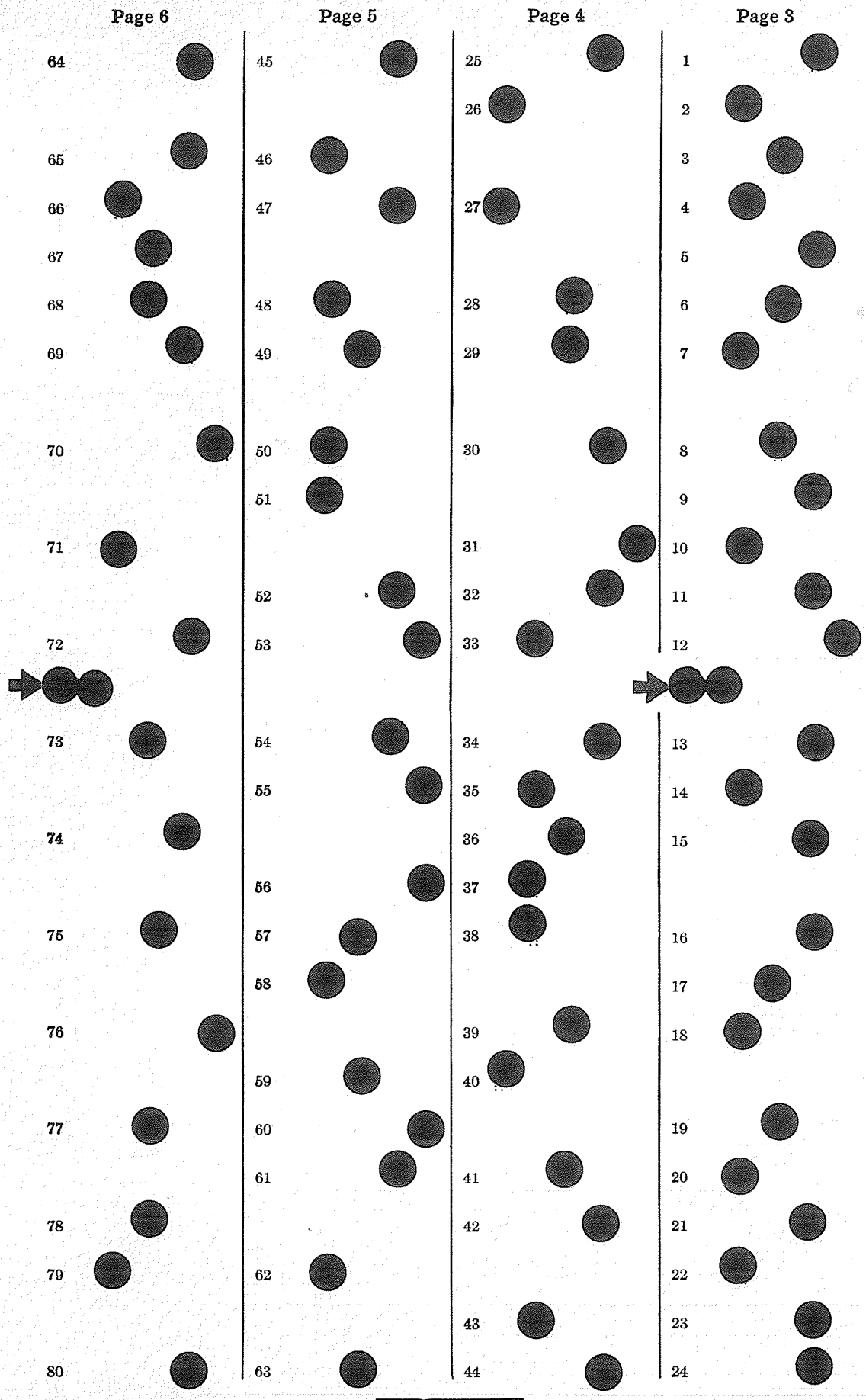
10–4 10–5	10–6 10–7	10–8	10– 9 10–10	10–11 11– 0	11–1 11–2	11–3 11–4	11–5 11–6	11–7 11–8	11– 9 11–10	11–11 12– 0	12–1 12–2	12–3 12–4	12–5 12–6	12–7 12–8	12– 9 12–10	CA / Score
163	162	161	160	159	158	157	156	155	154	153	152	151	150	149	148	80
161	160	159	158	157	156	155	154	153	152	151	150	149	148	147	146	79
159	158	157	156	155	154	153	152	151	150	149	148	147	146	145	144	78
157	156	155	154	153	152	151	150	149	148	147	146	145	144	143	142	77
155	154	153	152	151	150	149	148	147	146	145	144	143	142	141	140	76
153	152	151	150	149	148	147	146	145	144	143	142	141	140	139	138	75
151	150	149	148	147	146	145	144	143	142	141	140	139	138	137	136	74
149	148	147	146	145	144	143	142	141	140	139	138	137	136	135	134	73
147	146	145	144	143	142	141	140	139	138	137	136	135	134	133	132	72
145	144	143	142	141	140	139	138	137	136	135	134	133	132	131	130	71
143	142	141	140	139	138	137	136	135	134	133	132	131	130	129	128	70
142	141	140	139	138	137	136	135	134	133	132	131	130	129	128	127	69
141	140	139	138	137	136	135	134	133	132	131	130	129	128	127	126	68
140	139	138	137	136	135	134	133	132	131	130	129	128	127	126	125	67
139	138	137	136	135	134	133	132	131	130	129	128	127	126	125	124	66
138	137	136	135	134	133	132	131	130	129	128	127	126	125	124	123	65
137	136	135	134	133	132	131	130	129	128	127	126	125	124	123	122	64
136	135	134	133	132	131	130	129	128	127	126	125	124	123	122	121	63
135	134	133	132	131	130	129	128	127	126	125	124	123	122	121	120	62
134	133	132	131	130	129	128	127	126	125	124	123	122	121	120	119	61
133	132	131	130	129	128	127	126	125	124	123	122	121	120	119	118	60
132	131	130	129	128	127	126	125	124	123	122	121	120	119	118	117	59
131	130	129	128	127	126	125	124	123	122	121	120	119	118	117	116	58
130	129	128	127	126	125	124	123	122	121	120	119	118	117	116	115	57
129	128	127	126	125	124	123	122	121	120	119	118	117	116	115	114	56
128	127	126	125	124	123	122	121	120	119	118	117	116	115	114	113	55
127	126	125	124	123	122	121	120	119	118	117	116	115	114	113	112	54
126	125	124	123	122	121	120	119	118	117	116	115	114	113	112	111	53
125	124	123	122	121	120	119	118	117	116	115	114	113	112	111	110	52
124	123	122	121	120	119	118	117	116	115	114	113	112	111	110	109	51
123	122	121	120	119	118	117	116	115	114	113	112	111	110	109	108	50
122	121	120	119	118	117	116	115	114	113	112	111	110	109	108	107	49
121	120	119	118	117	116	115	114	113	112	111	110	109	108	107	106	48
120	119	118	117	116	115	114	113	112	111	110	109	108	107	106	105	47
119	118	117	116	115	114	113	112	111	110	109	108	107	106	105	104	46
118	117	116	115	114	113	112	111	110	109	108	107	106	105	104	103	45
117	116	115	114	113	112	111	110	109	108	107	106	105	104	103	102	44
116	115	114	113	112	111	110	109	108	107	106	105	104	103	102	101	43
115	114	113	112	111	110	109	108	107	106	105	104	103	102	101	100	42
114	113	112	111	110	109	108	107	106	105	104	103	102	101	100	99	41
113	112	111	110	109	108	107	106	105	104	103	102	101	100	99	98	40
112	111	110	109	108	107	106	105	104	103	102	101	100	99	98	97	39
111	110	109	108	107	106	105	104	103	102	101	100	99	98	97	96	38
110	109	108	107	106	105	104	103	102	101	100	99	98	97	96	95	37
109	108	107	106	105	104	103	102	101	100	99	98	97	96	95	94	36
108	107	106	105	104	103	102	101	100	99	98	97	96	95	94	93	35
107	106	105	104	103	102	101	100	99	98	97	96	95	94	93	92	34
106	105	104	103	102	101	100	99	98	97	96	95	94	93	92	91	33
105	104	103	102	101	100	99	98	97	96	95	94	93	92	91	90	32
104	103	102	101	100	99	98	97	96	95	94	93	92	91	90	89	31
103	102	101	100	99	98	97	96	95	94	93	92	91	90	89	88	30
102	101	100	99	98	97	96	95	94	93	92	91	90	89	88	87	29
101	100	99	98	97	96	95	94	93	92	91	90	89	88	87	86	28
100	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	27
99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	26
98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	25
97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	24
96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	23
95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	22
94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	21
93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	20
92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	19
91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	18
90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75	17
89	88	87	86	85	84	83	82	81	80	79	78	77	76	75	74	16
88	87	86	85	84	83	82	81	80	79	78	77	76	75	74	73	15
87	86	85	84	83	82	81	80	79	78	77	76	75	74	73	72	14
86	85	84	83	82	81	80	79	78	77	76	75	74	73	72	71	13
85	84	83	82	81	80	79	78	77	76	75	74	73	72	71	70	12
84	83	82	81	80	79	78	77	76	75	74	73	72	71	70	69	11
83	82	81	80	79	78	77	76	75	74	73	72	71	70	69	68	10
82	81	80	79	78	77	76	75	74	73	72	71	70	69	68	67	9
81	80	79	78	77	76	75	74	73	72	71	70	69	68	67	66	8
80	79	78	77	76	75	74	73	72	71	70	69	68	67	66	65	7
79	78	77	76	75	74	73	72	71	70	69	68	67	66	65	64	6
78	77	76	75	74	73	72	71	70	69	68	67	66	65	64	63	5
77	76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	4
76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	3
75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	2
74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	59	1

IQ's on Otis Beta, Form C_M or D_M (Cont'd)

Score \ CA	12–11 thru 13– 1	13–2 thru 13–3	13–4 thru 13–6	13–7 thru 13–9	13–10 thru 14– 0	14–1 thru 14–3	14–4 thru 14–6	14–7 thru 14–9	14–10 thru 15– 0	15–1 thru 15–3	15–4 thru 15–7	15–8 thru 16–1	16–2 thru 16–8	16–9 thru 17–5	17–6 and over	Score
80	147	146	145	144	143	142	141	140	139	138	137	136	135	134	133	80
79	145	144	143	142	141	140	139	138	137	136	135	134	133	132	131	79
78	143	142	141	140	139	138	137	136	135	134	133	132	131	130	129	78
77	141	140	139	138	137	136	135	134	133	132	131	130	129	128	127	77
76	139	138	137	136	135	134	133	132	131	130	129	128	127	126	125	76
75	137	136	135	134	133	132	131	130	129	128	127	126	125	124	123	75
74	135	134	133	132	131	130	129	128	127	126	125	124	123	122	121	74
73	133	132	131	130	129	128	127	126	125	124	123	122	121	120	119	73
72	131	130	129	128	127	126	125	124	123	122	121	120	119	118	117	72
71	129	128	127	126	125	124	123	122	121	120	119	118	117	116	115	71
70	127	126	125	124	123	122	121	120	119	118	117	116	115	114	113	70
69	126	125	124	123	122	121	120	119	118	117	116	115	114	113	112	69
68	125	124	123	122	121	120	119	118	117	116	115	114	113	112	111	68
67	124	123	122	121	120	119	118	117	116	115	114	113	112	111	110	67
66	123	122	121	120	119	118	117	116	115	114	113	112	111	110	109	66
65	122	121	120	119	118	117	116	115	114	113	112	111	110	109	108	65
64	121	120	119	118	117	116	115	114	113	112	111	110	109	108	107	64
63	120	119	118	117	116	115	114	113	112	111	110	109	108	107	106	63
62	119	118	117	116	115	114	113	112	111	110	109	108	107	106	105	62
61	118	117	116	115	114	113	112	111	110	109	108	107	106	105	104	61
60	117	116	115	114	113	112	111	110	109	108	107	106	105	104	103	60
59	116	115	114	113	112	111	110	109	108	107	106	105	104	103	102	59
58	115	114	113	112	111	110	109	108	107	106	105	104	103	102	101	58
57	114	113	112	111	110	109	108	107	106	105	104	103	102	101	100	57
56	113	112	111	110	109	108	107	106	105	104	103	102	101	100	99	56
55	112	111	110	109	108	107	106	105	104	103	102	101	100	99	98	55
54	111	110	109	108	107	106	105	104	103	102	101	100	99	98	97	54
53	110	109	108	107	106	105	104	103	102	101	100	99	98	97	96	53
52	109	108	107	106	105	104	103	102	101	100	99	98	97	96	95	52
51	108	107	106	105	104	103	102	101	100	99	98	97	96	95	94	51
50	107	106	105	104	103	102	101	100	99	98	97	96	95	94	93	50
49	106	105	104	103	102	101	100	99	98	97	96	95	94	93	92	49
48	105	104	103	102	101	100	99	98	97	96	95	94	93	92	91	48
47	104	103	102	101	100	99	98	97	96	95	94	93	92	91	90	47
46	103	102	101	100	99	98	97	96	95	94	93	92	91	90	89	46
45	102	101	100	99	98	97	96	95	94	93	92	91	90	89	88	45
44	101	100	99	98	97	96	95	94	93	92	91	90	89	88	87	44
43	100	99	98	97	96	95	94	93	92	91	90	89	88	87	86	43
42	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	42
41	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	41
40	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	40
39	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	39
38	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	38
37	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	37
36	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	36
35	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	35
34	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	34
33	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	33
32	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75	32
31	88	87	86	85	84	83	82	81	80	79	78	77	76	75	74	31
30	87	86	85	84	83	82	81	80	79	78	77	76	75	74	73	30
29	86	85	84	83	82	81	80	79	78	77	76	75	74	73	72	29
28	85	84	83	82	81	80	79	78	77	76	75	74	73	72	71	28
27	84	83	82	81	80	79	78	77	76	75	74	73	72	71	70	27
26	83	82	81	80	79	78	77	76	75	74	73	72	71	70	69	26
25	82	81	80	79	78	77	76	75	74	73	72	71	70	69	68	25
24	81	80	79	78	77	76	75	74	73	72	71	70	69	68	67	24
23	80	79	78	77	76	75	74	73	72	71	70	69	68	67	66	23
22	79	78	77	76	75	74	73	72	71	70	69	68	67	66	65	22
21	78	77	76	75	74	73	72	71	70	69	68	67	66	65	64	21
20	77	76	75	74	73	72	71	70	69	68	67	66	65	64	63	20
19	76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	19
18	75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	18
17	74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	17
16	73	72	71	70	69	68	67	66	65	64	63	62	61	60	59	16
15	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58	15
14	71	70	69	68	67	66	65	64	63	62	61	60	59	58	57	14
13	70	69	68	67	66	65	64	63	62	61	60	59	58	57	56	13
12	69	68	67	66	65	64	63	62	61	60	59	58	57	56	55	12
11	68	67	66	65	64	63	62	61	60	59	58	57	56	55	54	11
10	67	66	65	64	63	62	61	60	59	58	57	56	55	54	53	10
9	66	65	64	63	62	61	60	59	58	57	56	55	54	53	52	9
8	65	64	63	62	61	60	59	58	57	56	55	54	53	52	51	8
7	64	63	62	61	60	59	58	57	56	55	54	53	52	51	50	7
6	63	62	61	60	59	58	57	56	55	54	53	52	51	50	50–	6
5	62	61	60	59	58	57	56	55	54	53	52	51	50	50–		5
4	61	60	59	58	57	56	55	54	53	52	51	50	50–			4
3	60	59	58	57	56	55	54	53	52	51	50	50–				3
2	59	58	57	56	55	54	53	52	51	50	50–					2
1	58	57	56	55	54	53	52	51	50	50–						1

OTIS QUICK-SCORING MENTAL ABILITY TESTS

KEY FOR BETA TEST: FORMS CM, DM, EM, and FM



Published by World Book Company, Yonkers-on-Hudson, New York, and Chicago, Illinois

OTIS QUICK-SCORING MENTAL ABILITY TESTS

CLASS RECORD FOR BETA TEST: FORMS CM, DM, EM, & FM AND GAMMA TEST: FORMS AM, BM, EM, & FM

Test used (underline) Beta Gamma

Form used.....Examiner.....

Grade.....Teacher.....School.....

City.....Date of exam.....19....

Name	Age		Score	IQ	Add'l data (?)	
	Yrs.	Mos.				
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
Class Medians						

(*Continued on reverse side*)

NOTE. See under "Reporting to the Author" in the Manual of Directions, regarding a request for data. *b*

PRINTED IN U.S.A.

CLASS RECORD FOR BETA TEST: FORMS CM, DM, EM, AND FM AND GAMMA TEST: FORMS AM, BM, EM, AND FM — *Continued*

	Name	Age		Score	IQ	Add'l data (?)	
		Yrs.	Mos.				
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							

DIRECTIONS. To distribute the scores of a class, make a mark in the second column for each score opposite the interval within which the score falls. In the "Frequency" column write the number of tally marks that fell in each interval.

Scores	Distributing Marks	Freq.
80		
75–79		
70–74		
65–69		
60–64		
55–59		
50–54		
45–49		
40–44		
35–39		
30–34		
25–29		
20–24		
15–19		
10–14		
5–9		
0–4		
Total		
Median		

Otis Quick-Scoring Mental Ability Tests

by **Arthur S. Otis**

MANUAL OF DIRECTIONS FOR BETA TEST
Forms CM and DM and New Edition: Forms EM and FM

The Quick-Scoring Series

The *Otis Quick-Scoring Mental Ability Tests* comprise three tests, called Alpha, Beta, and Gamma. The three tests are designed for grades as follows:

Alpha Test.....Grades 1–4
Beta Test.....Grades 4–9
Gamma Test...High Schools and Colleges

The Alpha Test, both in the regular and the short form, consists entirely of pictures. The Beta and Gamma Tests originally were revisions and extensions of the Intermediate and Higher Examinations, respectively, of the *Otis Self-Administering Tests of Mental Ability*. New forms EM and FM of both the Beta and the Gamma Tests have been equated to the older forms.

Purpose of the Tests

The purpose of the three tests in the series is to measure mental ability — thinking power or the degree of maturity of the mind.

It should be understood from the outset that it is not possible to measure mental ability directly. It is possible only to measure the effect mental ability has had in enabling the pupil to acquire certain knowledge and mental skill. Of course, the answering of some types of questions depends less upon schooling and more upon mental ability than the answering of others, and in making up the test the aim has been for the most part to choose that kind of question which depends as little as possible on schooling and as much as possible on thinking.

However, in the interest of variety it has been found necessary and even advantageous to include in verbal tests of mental ability, such as the Beta and Gamma Tests, certain questions which might seem at first glance to be mere measures of achievement. This type includes questions on vocabulary, arithmetic reasoning, etc. It must be remembered, however, that any test which involves the use of language can measure mental ability only to the extent to which we may assume that pupils of the same age have had approximately the same opportunity to learn. Consequently, if a pupil has grown up with limited educational opportunities, especially with reference to language, his mental ability is not fairly measured by any test involving language. But in a given community in which all children have approximately the same educational opportunities, it is reasonable to assume that a pupil who progresses rapidly in school and learns much has greater mental ability for his age than one who progresses less rapidly and learns less. To this extent, therefore, certain achievement questions such as vocabulary and arithmetic-reasoning questions, even though depending on language, do measure mental ability.

Alternative Forms

There are six forms of the Beta Test. Forms A and B are published in a smaller size for hand scoring only. Forms CM, DM, EM, and FM are for machine scoring or hand scoring.

Special Features

The tests are self-administering. It is necessary merely to pass out the booklets, allow the pupils time to study the first page with a minimum of directions, and then let them go ahead and take the test. A single examiner may administer the tests to all the classes of a moderate-sized school in a day, by devoting a few minutes to start one class taking the test, leaving the class in care of the teacher, and going on to the next class, etc. This is a good way to assure reasonable uniformity of procedure in the giving of the tests.

PRINTED IN U.S.A.

In addition to the ease of administration which these tests afford by virtue of their single time limit, a method of stencil scoring is provided by which the tests may be rapidly scored.

Provision is made in Forms CM, DM, EM, and FM for the pupil to put his answers to all the questions on one sheet of the test booklet. This sheet is called the Answer Sheet and appears as page 2. To use the Answer Sheet, the pupil tears it off from the rest of the booklet and slips it under the booklet in such a way that the spaces for the answers appear just to the right of the test page.

A row of 5 spaces like this:  on the Answer Sheet corresponds to each question. The spaces are numbered consecutively and arranged so as to align perfectly with the questions on the test paper in order to make sure the pupil will not put his answer mark in the wrong row of spaces.

To indicate his answer to a question, the pupil makes a vertical mark in the space that has the same number as the answer he has chosen, like this: 

The Answer Sheet is then scored by a stencil key containing holes so spaced that if the pupil has put his mark in the right space it will show through the hole in the Key; otherwise not. To score the paper, it is necessary merely to count the marks that can be seen through the holes in the Key. One application of the Key is sufficient, of course, to score the whole test.

Experience shows that this is the quickest possible method of scoring a test "by hand," so to speak. Its principal advantage is that the scorer does not have to look at each answer to see whether a cross is in or not in a given square or circle — he disregards all wrong answers completely and merely counts right ones. It is by reason of this scoring feature that the tests are called "Quick-Scoring Tests."

The test may be scored also by the International Business Machines Corporation scoring machine. For this purpose a special Separate Answer Sheet must be used. It is used in the same way as the Attached Answer Sheet but is printed and sold separately. Special mechanical pencils must be used by the pupils when marking the machine-scored Answer Sheet. Special Directions for Administering with the Machine Scoring Answer Sheet are given on the next page.

Directions for Administering

Two separate sets of directions for administering are furnished — one for use with the Attached Answer Sheet (see col. 2, this page) and the other for use with the Machine Scoring Answer Sheet (see col. 1, page 3). Be sure to use the appropriate directions. Give all directions slowly and distinctly, with a pause after each sentence.

To administer Beta, Form CM, DM, EM, or FM, address the pupils as follows:

Use the following directions with the Attached Answer Sheet. (Directions for Machine Scoring Answer Sheet are given on next page.)

"We are now going to give you some tests that measure your ability to think. I will pass out the test papers and as soon as you receive one, read the first page and do what it tells you to do; that is, fill the blanks, giving your name, age, etc., and answer the sample questions.

"Do not open or turn over the booklet. Part of the test is to see if you can follow directions."

Have the test papers passed, one to each pupil, right side up; that is, with the title page up.

Allow a reasonable time for all to finish reading the first page; then say: "Is there anyone who does not understand how to answer the samples?" Be sure all do.

Instruct the pupils to tear the Answer Sheet off from the rest of the booklet. See that every pupil is supplied with two pencils and an eraser. It is better not to have the pencils too sharp, principally because it is better to have the pupils make wide marks.

Then say: "You are to put your marks in the spaces on the Attached Answer Sheet.

"Slip the Answer Sheet under the edge of page 3 so that the column of spaces marked 'Page 3' is alongside page 3 like this. (Show by holding up page 3 with the "Page 3" column of the Answer Sheet close to page 3 of the booklet.) Notice that the arrow tips on the Answer Sheet point directly toward the arrow tips on page 3. In answering the first question, you put a mark in one of the spaces in the first row, and so on.

"When you finish page 3, pull out the Answer Sheet a little way like this (Show.) so that you can see the column of answers for page 4, and do page 4. Always keep the Answer Sheet shoved under the booklet so that the column of the Answer Sheet on which you are working is close to the test paper.

"When you come to page 5, fold page 6 under like this (Show how.) so that you can get the 'Page 5' column of the Answer Sheet close to page 5 of the booklet like this. (Show.)

"Never put more than one mark in any row of spaces.

"Is there anyone who does not understand what to do?" (Answer any questions about how to take the test.) Then say:

"As explained in the paragraph below the samples, the test contains eighty questions. You are not expected to be able to answer all of them, but do the best you can. You will be allowed a half hour. Try to get as many right as possible. Be careful not to go so fast that you make mistakes. Do not spend too much time on any one question. No questions about the test will be answered after the test begins.

"Now go ahead and answer the questions. Remember to make heavy black marks." (Continue with directions on page 3, column 2.)

Use the following directions with the Machine Scoring Answer Sheet.

"We are now going to give you some tests that measure your ability to think. I will pass out the test papers with the Separate Answer Sheets inserted. As soon as you receive the Answer Sheet, fill the blanks here (Point to place on Answer Sheet.), **giving your name, age, etc. Do not write anything on the test booklet.**

"Do not open or turn over the test booklet. Part of the test is to see if you can follow directions."

Give each pupil a test paper with the Separate Answer Sheet inserted. Pass out the mechanical pencils, and then say:

"Read this front page of the test booklet carefully. You see that there are spaces here for recording answers. (Hold up booklet and point to the spaces.) **DO NOT put the answers to the samples in these spaces. You are not to mark the test booklet in any way.**

"If you look in the upper left-hand corner of the Answer Sheet (Hold up an Answer Sheet and point to the spaces for answers to sample.), **you will see spaces for the answers to sample questions *a*, *b*, and *c*. Put your answers to the samples in these spaces. Read the front page of the test booklet and answer the sample questions."** Allow a reasonable time for all to finish reading the first page; then say: **"Is there anyone who does not understand how to answer the samples?"** Be sure all do.

Then say: **"All your answers are to be marked in the spaces on the Answer Sheet.**

"Slip the Answer Sheet under the edge of page 3 so that the column of spaces marked 'Page 3' is alongside page 3 like this. (Show by holding up page 3 with the "Page 3" column of the Answer Sheet close to page 3 of the booklet.) **Notice that the arrow tips on the Answer Sheet point directly toward the arrow tips on page 3. In answering the first question, you put a mark in one of the spaces in the first row, and so on.**

"When you finish page 3, pull out the Answer Sheet a little way like this (Show.) **so that you can see the column of answers for page 4, and do page 4. Always keep the Answer Sheet shoved under the booklet so that the column of the Answer Sheet on which you are working is close to the test paper.**

"When you come to page 5, fold page 6 under like this (Show how.) **so that you can get the 'Page 5' column of the Answer Sheet close to page 5 of the booklet like this.** (Show.)

"Never put more than one mark in any row of spaces.

"In making your marks on the Answer Sheet, move your pencil up and down two or three times so as to make a heavy black mark filling the space between the two dotted lines in each case.

"Is there anyone who does not understand what to do?" (Answer any questions about how to take the test.) Then say:

"As explained in the paragraph below the samples, the test contains eighty questions. You are not expected to be able to answer all of them, but do the best you can. You will be allowed a half hour. Try to get as many right as possible. Be careful not to go so fast that you make mistakes. Do not spend too much time on any one question. No questions about the test will be answered after the test begins.

"Now go ahead and answer the questions. Remember to make heavy black marks."

(*Continue here.*)

Write immediately on the board the exact time when the pupils begin to take the test. It is helpful to write on the board also the time the pupils must stop work. Thus, if pupils are started at 1 : 17, write this on the board and under it write 1 : 47. Or set your watch exactly on the hour and when it is exactly half past the hour by your watch, the time will be up.

It should be understood by the examiner that no questions about the test are to be answered which might give the pupils the slightest help in answering the questions; that is, the examiner or teacher may not explain the meaning of any word or give any hints. It is permissible at the beginning of the examination for the examiner to move quietly about the room to make sure that the pupils are indicating their answers in the proper manner (making heavy black marks), and if during the examination a pupil becomes confused on account of the use of the Separate Answer Sheet, it is permissible, of course, to explain to him how to proceed. Thereafter it is better for the teacher to remain seated at her desk so that the room is quiet and the pupils may work undisturbed.

The one in charge of timing the test should be particularly impressed with the need to watch the time carefully, for it is very easy to forget the time and let the pupils work more than the time allowed.

After exactly *30 minutes*, say: **"Stop! Lay your pencil down."**

If the Attached Answer Sheet is used, have the Answer Sheets collected, then have the pupils write their names at the top of page 3 of the booklet, and have the booklets collected. If the Machine Scoring Answer Sheet is used, have the Separate Answer Sheets and the test booklets collected at once.

Directions for Scoring

Hand scoring

A Key for scoring the Attached Answer Sheet is included in each package of tests.

The pupils have been instructed to be sure not to put more than one mark in any row of spaces. However, if in the case of any item two marks have been put in the same row of spaces, draw a colored line through the row of answer spaces and give no credit for that item.

To score an Answer Sheet, lay the Key over the Answer Sheet in such a way that two of the heavy arrows on the Answer Sheet show through the holes of the Key and point directly toward the two arrows on the Key. The Key will then be adjusted so that all the marks that the pupils have made in the right spaces will show through the holes. The number of marks so appearing is the pupil's score. This score should be written in the space provided at the top of the title page.

Ordinarily in scoring this test there is no need to mark the answers right or wrong but merely to count them, for only the total score is of significance. To avoid errors in counting, after the numbers right have been counted, continue the count with the wrong and omitted items and make sure that you end with 80.

In the interest of accuracy it is well for each Answer Sheet to be scored independently by two persons. If this is done, the score obtained by the first scorer may be written at the foot of the page. Then, after the next scorer has scored the paper and compared his count with that made by the first scorer and found it to check, the sheet may be turned and the checked score written on the title page.

Machine scoring

It is assumed here that all persons attempting to score the Otis Answer Sheet on the International Test Scoring Machine will have thoroughly familiarized themselves with the scoring techniques described in the various International Business Machines publications, particularly as they concern the manipulation of the machine itself. To insure scoring of satisfactory accuracy, the following steps are suggested:

1. Adjust the machine properly, according to the manufacturer's directions.
2. Set the master switch on A and the formula switch on R.
3. Scan each Answer Sheet carefully before it is scored. Where more than one answer has been marked for an item, erase all marks for the item. Erase any stray pencil marks made in an answer space, inasmuch as even very small and light marks are sometimes sensed by the machine. If the pupil has failed to make complete erasures, make a clean erasure. If the marks are too light, go over them with one of the special lead pencils.
4. Check carefully by hand a certain proportion of the Answer Sheets to insure maximum accuracy.
5. Enter the raw score in the box provided for it on the Answer Sheet.

Directions for Recording Scores

In each package of tests there is included a Class Record which provides for the recording of scores of a class. Before entering the scores on the Class Record, arrange the Answer Sheets either in alphabetical order or in order of magnitude of score, according to preference. Then enter the name of each pupil, his age in years and months, and his score.

Note that provision is made on the Class Record for entering later the IQ of the pupil and any additional data, such as percentile rank in the class or school, classification designation, etc., and for entering the median age, median score, etc., if desired.

Provision is made at the foot of page 2 of the Class Record for distributing the scores of a class or a school. After the marks are all entered, count those in each interval and write the number in the column headed "Freq." (Frequency).

To find the median (middle) score, count from either end of the distribution to the middle mark. If the middle mark falls, say, in the interval 50–54, sort out the papers whose scores fall in this interval, and, if the median is the third mark in the interval, find the score on the third paper in that bunch of papers. That score is the median score of the class.

Distributions of Scores

Table 1 shows the distributions of scores by ages of 12,983 sixth-grade pupils in Form A. About half are from a large city in Ohio and about half from towns and villages of New York State. The median age of these pupils was 12 years and 4 months and the median score 42 points.

TABLE 1. *Distributions of Scores by Ages of 12,983 Sixth-Grade Pupils in the Otis Quick-Scoring Mental Ability Tests: Beta Test, Form A*

SCORE INTERVALS	AGE LAST BIRTHDAY									TOTALS
	9	10	11	12	13	14	15	16	17	
75–79			2	1						3
70–74		1	21	11						33
65–69		7	90	53	5	2				157
60–64	1	17	243	130	18		1			410
55–59		37	475	342	28	13				895
50–54		62	692	605	106	32	2	1		1500
45–49		53	860	842	260	54	7	2		2078
40–44	2	43	818	1065	410	121	20	4		2483
35–39	1	30	593	914	482	162	27	3		2212
30–34		13	321	562	437	174	33	9	1	1550
25–29		12	149	293	255	134	24	11	1	879
20–24	1	5	62	132	140	75	22	8	1	446
15–19		2	21	50	82	53	12	4	2	226
10–14	1	2	3	15	29	26	8	2		86
5–9		1		2	10	10				23
0–4							1	1		2
Totals	6	285	4350	5017	2262	856	157	45	5	12,983
Median age: 12 yr., 4 mo. Median score: 42										

The table is interpreted as follows: The column headed 12 contains the distributions of scores of the 5017 sixth-grade pupils whose age last birthday was 12 years, and whose ages therefore ranged from 12 years to 13 years at the time of the test. It shows that, of those pupils, 1 made a score that fell in the interval 75–79, 11 made scores that fell in the interval 70–74, etc.

This table is given partly to show what wide ranges of ages and ability are found in a single grade. Of course, the average classroom does not show quite as wide a range of ages and scores, but nearly so. The need for dividing the pupils of such a grade into more homogeneous groups and the method of doing so are given below under the heading "Application of Results" (page 8).

Norms

If a large number of 12-year pupils take a test and the scores are arranged in order, the median or middle score is considered just normal for 12-year pupils and is said to be the *norm* for the age of 12 years. Table 2 a gives the norms for the various ages of pupils taking Beta, Form CM or DM. Table 2 b gives the norms for pupils taking Beta, Form EM or FM.

TABLE 2 a. *Age Norms for Beta:* **Forms CM and DM**

YEARS→		8	9	10	11	12	13	14	15	16	17	18 or over
MONTHS	0	10	17	24	31	37	43	47	51	54	56	57
	1	11	18	25	32	38	43	48	52	54	56	
	2	11	18	26	32	38	44	48	52	55	56	
	3	12	19	26	33	39	44	48	52	55	56	
	4	13	20	27	33	39	45	49	53	55	56	
	5	13	20	27	34	40	45	49	53	55	56	
	6	14	21	28	34	40	45	49	53	55	57	
	7	14	21	28	35	41	46	50	53	55	57	
	8	15	22	29	35	41	46	50	54	55	57	
	9	15	23	30	36	42	46	50	54	56	57	
	10	16	23	30	36	42	47	51	54	56	57	
	11	17	24	31	37	43	47	51	54	56	57	

See Table 5 for Augmenting scores over 70.

TABLE 2 b. *Age Norms for Beta:* **Forms EM and FM**

YEARS→		8	9	10	11	12	13	14	15	16	17	18 or over
MONTHS	0	9	14	21	29	35	41	46	50	53	56	57
	1	9	15	22	30	36	41	47	51	53	56	
	2	9	15	23	30	36	42	47	51	54	56	
	3	10	16	23	31	37	43	47	51	54	56	
	4	10	17	24	31	37	44	48	52	54	56	
	5	10	17	24	32	38	44	48	52	55	56	
	6	11	18	25	32	38	44	48	52	55	57	
	7	11	18	25	33	39	45	49	52	55	57	
	8	12	19	26	33	39	45	49	53	55	57	
	9	12	20	27	34	40	45	49	53	56	57	
	10	13	20	27	34	40	46	50	53	56	57	
	11	14	21	28	35	41	46	50	53	56	57	

See Table 5 for Augmenting scores over 70.

Table 2 a is read as follows: The norm for the age of 8 years 0 months on Beta CM or DM is 10 points of score; the norm for the age of 11 years 3 months is 33 points, etc. Table 2 b is read in a similar manner.

The norms in Table 2 a are based in part on the scores of 16,242 pupils in Beta, Form A; in part on a comparison of scores in Beta and scores in the Intermediate Examination of the *Otis Self-Administering Tests of Mental Ability* made by means of an experiment in which 3259 pupils in Grades 4 to 9 took Beta, Forms A and B, and Form A of the Intermediate Examination; in part on a comparison between Beta, Form A, and Alpha, Nonverbal, in which 612 pupils in Grades 4 and 5 took both these tests; in part on a comparison between Beta, Form CM, and Gamma in which 742 pupils in Grades 7, 8, and 9 took both these tests; and in part on two experiments in which Beta, Form CM, was compared with Beta, Form A, using groups of 780 and 1068 pupils in Grades 4 to 9. The norms in Table 2 b are based on a comparison of scores on Beta EM and FM with CM, by means of an experiment in which 3107 pupils in Grades 5 to 9 took part.

Local norms for different localities differ markedly. The norms in Tables 2 a and 2 b, therefore, should not be thought of as necessarily representative of any particular section of the country but rather as representative of the country as a whole.

These norms apply to a first test. If a pupil takes a second form of the test later, it is necessary to make a correction for familiarity with the test before using Tables 2 a and 2 b. (See "Practice Effect" below.)

Practice Effect

When a pupil takes a second form of a test within a short time after the first form, he tends to make a better score on the second test. This increase in score is generally called "practice effect."

It was found that when a second form of Beta was given two days after the first form, the practice effect was about 4 points. This means that to render the second score of a pupil comparable to the first score if the tests were taken two days apart, 4 points should be subtracted from the second score.

Practice effect decreases, of course, as the length of time between tests increases. Possibly the amount of practice effect would drop to about 3 points if the interval were a week; to 2 points if the interval were a month; to 1 point if the interval were three months or more.

Whenever it is desired to find a Mental Age or IQ (see below) from the score of a pupil in a second test, the proper correction should be made for practice effect in the second score before comparing it with the norm for the pupil's age in Tables 2 a and 2 b or before finding the pupil's Mental Age.

Mental Ages

Some examiners wish to express scores in terms of Mental Age. The term "Mental Age" originally meant the age for which a pupil's score was normal or median. Thus, if a pupil makes a score just normal or median for pupils 10 years old, he is said to have a Mental Age (MA) of 10 years.

TABLE 3 a. *Mental Ages Corresponding to Scores in Beta:* Forms CM and DM

SCORE	MA	SCORE	MA	SCORE	MA	SCORE	MA
1	6–8	21	9–7	41	12–8	61	16–0
2	6–10	22	9–8	42	12–10	62	16–2
3	7–0	23	9–10	43	13–0	63	16–4
4	7–1	24	10–0	44	13–2	64	16–6
5	7–3	25	10–1	45	13–4	65	16–8
6	7–5	26	10–3	46	13–6	66	16–10
7	7–7	27	10–5	47	13–8	67	17–0
8	7–8	28	10–7	48	13–10	68	17–2
9	7–10	29	10–8	49	14–0	69	17–4
10	8–0	30	10–10	50	14–2	70	17–6
11	8–2	31	11–0	51	14–4	71	17–8
12	8–3	32	11–2	52	14–6	72	17–10
13	8–5	33	11–4	53	14–8	73	18–0
14	8–7	34	11–6	54	14–10	74	18–2
15	8–9	35	11–8	55	15–0	75	18–4
16	8–10	36	11–10	56	15–2	76	18–6
17	9–0	37	12–0	57	15–4	77	18–8
18	9–2	38	12–2	58	15–6	78	18–10
19	9–3	39	12–4	59	15–8	79	19–0
20	9–5	40	12–6	60	15–10	80	19–2

TABLE 3 b. *Mental Ages Corresponding to Scores in Beta:* Forms EM and FM

SCORE	MA	SCORE	MA	SCORE	MA	SCORE	MA
1	7–1	21	10–0	41	13–0	61	16–0
2	7–3	22	10–1	42	13–2	62	16–2
3	7–5	23	10–3	43	13–3	63	16–4
4	7–7	24	10–5	44	13–4	64	16–6
5	7–8	25	10–7	45	13–6	65	16–8
6	7–10	26	10–8	46	13–8	66	16–10
7	8–0	27	10–10	47	13–10	67	17–0
8	8–2	28	11–0	48	14–0	68	17–2
9	8–3	29	11–1	49	14–2	69	17–4
10	8–5	30	11–2	50	14–4	70	17–6
11	8–7	31	11–4	51	14–6	71	17–8
12	8–9	32	11–6	52	14–8	72	17–10
13	8–10	33	11–8	53	14–10	73	18–0
14	9–0	34	11–10	54	15–0	74	18–2
15	9–2	35	12–0	55	15–1	75	18–4
16	9–3	36	12–2	56	15–2	76	18–6
17	9–5	37	12–4	57	15–4	77	18–8
18	9–7	38	12–6	58	15–6	78	18–10
19	9–8	39	12–8	59	15–8	79	19–0
20	9–10	40	12–10	60	15–10	80	19–2

This method of interpretation has a serious limitation, since mental growth slows down along with physical growth, and pupils reach a mental maturity in their teens. Thus the highest norm for any age in the Beta Test, Form CM or DM, is 57 points, as shown in Table 2 a. This means that some pupils make scores that are above what is normal for any age. In order to express degrees of mental ability which are above the norm for adults in terms of Mental Age, it is customary to proceed as though mental growth did not slow down but kept on at about the rate it is increasing between the ages of 12 and 13, which in the Beta Test is approximately 1 point in score for each two months of age. According to this assumption, artificial mental ages are assigned to scores above age 13. This is called "extrapolation." This extrapolation method is used also with the Binet Scale.

According to the above method tables of Mental Ages (Tables 3 a and 3 b) have been drawn up.

Table 3 a is read as follows: A score of 1 in Beta CM or DM denotes a Mental Age of 6 years 8 months; a score of 61 may be treated as denoting a Mental Age of 16 years 0 months (though actually it is 4 points above the norm for adults). Table 3 b is read in a similar manner.

Measuring Brightness

Pupils making the same score in the test are presumed to have the same mental ability or, as we say, the same Mental Age even though their actual ages (spoken of as "chronological ages") are not the same. That is, as explained above, a pupil who makes a score equal to the norm for the age of 10 years is said to have a Mental Age of 10 years, whether the pupil is 10 years old or 9 years old or 11 years old.

A 10-year pupil who has a Mental Age of 11 years is brighter than normal, and a measure of his brightness is often found by dividing his Mental Age of 11 years by his "chronological age" of 10 years (11 ÷ 10 = 1.10). The decimal point is then dropped and the 110 is called the pupil's Intelligence Quotient (IQ). Intelligence Quotients so found cluster most thickly around 100, but in a few instances go above 150 or below 50. They are distributed according to the "law of normal distribution."

A study of the dispersion of IQ's of various populations aggregating 100,000 pupils tested by various group tests of mental ability showed standard deviations of IQ's ranging from 10 to 19 points of IQ for the various populations, the median value of the standard deviations of IQ being between 15 and 16 points; hence theoretically about $\frac{1}{10}$ of 1% of pupils make IQ's of 150 or over, $\frac{1}{2}$ of 1% of pupils make IQ's of 142 or over, and so on as shown in Table 4.

TABLE 4. *Per Cents of Pupils Making Various IQ's*

This per cent of pupils	make these IQ's:	This per cent of pupils	make these IQ's:
$\frac{1}{10}$ of 1%	150 or over	$\frac{1}{10}$ of 1%	50 or less
$\frac{1}{2}$ of 1%	142 or over	$\frac{1}{2}$ of 1%	58 or less
1%	136 or over	1%	64 or less
5%	126 or over	5%	74 or less
10%	121 or over	10%	79 or less
25%	111 or over	25%	89 or less
$33\frac{1}{3}$%	107 or over	$33\frac{1}{3}$%	93 or less
50%	100 or over	50%	100 or less

A measure of brightness comparable to the IQ can be found from scores of pupils in the Beta Test according to the method below. Although the measures are not quotients, they are called "Beta IQ's" because they are comparable to IQ's.

How to Find a Pupil's "Beta IQ"

To find a pupil's "Beta IQ," proceed as follows:

1. Find the norm for the pupil's age from Table 2a or Table 2b, depending upon the form taken.

2. Find the amount by which the pupil's score exceeds (or falls below) the norm for his age. Call this his "deviation of score."

3. Add the pupil's deviation of score to 100 (or subtract from 100 if the deviation is downward). The result is the pupil's "Beta IQ."

4. If a pupil's score is above 70, it is to be augmented before proceeding with Steps 2 and 3 above. Treat a score of 71 as though it were 72. Treat a score of 72 as though it were 74, etc., according to Table 5.

TABLE 5. *For Augmenting High Scores*

Treat a score of	71	72	73	74	75	76	77	78	79	80
as though it were	72	74	76	78	80	82	84	86	88	90

As a sample of Step 4, suppose a pupil of 16 years 6 months makes a score of 75. The norm for 16 years 6 months is 55. To find his deviation of score, treat the score of 75 as though it were 80, subtract 55 from 80 (answer 25), and add 25 to 100, yielding an "IQ" of 125.

Various determinations of the dispersions of "Beta IQ's" yield standard deviations of "IQ" of from 10 to 17 points for various populations. The standard deviation of "IQ's" of 32,139 pupils of Pittsburgh derived from scores in the Intermediate Examination (similar to Beta) was 16.2 points. It is believed that "Beta IQ's" tend to be somewhat less dispersed than IQ's obtained by the division method from group tests in general (that is, they tend to be somewhat nearer to 100); therefore allowance should be made for this fact when comparing "Beta IQ's" with ordinary IQ's from other tests.

However, the above method is recommended as yielding measures of brightness that are more consistent and constant for a given individual than ordinary IQ's.

Reliability and Validity of the Beta Test

By "reliability" is meant the degree of precision with which a test measures what it measures.

One common measure of the reliability of a test is the coefficient of correlation between two forms of the test. Table 6 gives the coefficients of correlation between Forms A and B in Grades 4 to 9 of a large school system, the average number of pupils per coefficient being 86. The average of the 12 coefficients is .79. For Grades 4 to 9 combined the coefficient is .96.

Another measure of reliability is the coefficient of correlation between odd and even items of a single test. This is virtually a correlation between two forms of a short test each half as long as the full test, the two tests being given, we might say, simultaneously.

TABLE 6. *Reliability Coefficients (Form A vs. Form B)*

	GRADES						GRADES COMBINED
	4	5	6	7	8	9	
A (1st)–B (2d)	.730	.979	.826	.711	.833	.665	.96
B (1st)–A (2d)	.764	.842	.859	.869	.688	.651	

It is customary, then, to correct the coefficients of correlation between the half tests by the Spearman-Brown formula to obtain the corresponding coefficient for two full-length tests given under the same circumstances.

The coefficients of correlation for the odd and even items of one test (Form CM) are as shown in Table 7a.

TABLE 7a. *Reliability Coefficients (Odd vs. Even Items) for Form CM Corrected by Spearman-Brown Formula*

GRADES	4	5	6	7	8	9
CORRECTED COEFFICIENTS	.81	.92	.90	.87	.86	.79

The average of the six corrected coefficients in Table 7a is .86, which is 7 points higher than .79, the average of the coefficients of Table 6. This deficiency of 7 points in the coefficients of Table 6 is due to the instability of the pupils themselves. That is, if pupils remained as constant in ability from day to day as from moment to moment, so to speak, the coefficients in Table 6 would be as high as the coefficients in Table 7a.

TABLE 7b. *Reliability Coefficients (Odd vs. Even Items) for Form EM Corrected by Spearman-Brown Formula*

GRADES	5	6	7	8	9
CORRECTED COEFFICIENTS	.89	.84	.94	.93	.95

Table 7b shows the coefficients of correlation for odd and even items for Form EM.

Another measure of reliability which is entirely independent of the degree of heterogeneity of the group is the *standard error of measurement.* By "standard error of measurement" is meant the amount by which any pupil's actual score may differ from his "true" score in two cases out of three.

In the case of 465 pupils in Grades 4 to 9 the standard error of measurement was 4.0 points.

That is, a pupil's score will be in error not more than 4.0 points in $66\frac{2}{3}\%$ of cases.

By validity of a test is meant the degree to which it measures the ability it is designed to measure. Or we might say, it is the degree to which it serves its purpose.

Now the purpose of the Beta Test is most generally that of finding the degree of brightness of a pupil; that is, obtaining some measure (such as the IQ) that indicates the probable rate of progress the pupil will make in school. This being the case, it follows that actual rate of progress of pupils through school is the most appropriate criterion of the validity of the Beta Test.

This criterion is the one that was used in the development of the Otis Intermediate Examination, from which most of the items of Forms CM and DM of the Beta Test were taken. The method is described in the Manual for *Otis Self-Administering Tests of Mental Ability* (page 3). The determination of the validity of each item consisted of comparing the number of passes of that item by a group of pupils who were making rapid progress through school with the number of passes of the item by a group of pupils who were making slow progress through school. Only those items were used which showed a distinct gain in number of passes of the rapid-progress pupils over the number of passes of the slow-progress pupils. Each item justified its inclusion, therefore, because it contributed definitely to the capacity of the test to measure brightness as reflected in rate of progress through school.

When Forms EM and FM were prepared, difficulty and validity indices [1] were computed for each item in these new forms. Since all pupils in the item-analysis experiment took Form CM as well as one of the new forms, difficulty and validity indices were also computed for the items in the older Form CM. The final items in EM and FM were selected to match those in CM in terms of difficulty, validity, and item type. The mean difficulty for Grades 6 and 7 combined on each of the three forms was found to be approximately 60%. The mean validity index of the test items in each form was approximately .45.

Since the *Otis Quick-Scoring Mental Ability Tests* will be used mainly for the prediction of scholastic success, it is important that there be some objective evidence of the relationship between performance on the Otis test and school achievement. In Table 8 are shown the correlations between Otis scores and scores on the subtests of Form J of the *Stanford Achievement Test* for single grade ranges.

The Otis tests and the *Stanford Achievement Test* used in these correlations were administered within a month of one another. However, it seems reasonable to assume that if the Otis tests had been administered sometime previous to the achievement test the correlations would not vary greatly.

[1] Difficulty values for each item were computed by averaging the per cents passing each item in the upper and lower 27% of the item-analysis population. Validity indices are approximations of the item-total score correlations obtained from the upper-lower 27% groups by means of the Flanagan table.

TABLE 8. *Correlations between Otis Quick-Scoring Mental Ability Tests, Beta Test, and Stanford Achievement Test, Form J*

OTIS SCORE AND	GRADE 5 $N = 396$	GRADE 8 $N = 398$
1. Paragraph Meaning	.770	.770
2. Word Meaning	.827	.819
3. Spelling	.748	.623
4. Language	.698	.731
5. Arithmetic Reasoning	.673	.723
6. Arithmetic Computation	.564	.685
7. Social Studies	.779	.742
8. Science	.761	.765
9. Study Skills	.716	.760

Application of Results

Purposes of mental ability tests. The principal purposes for which mental tests are given are these:

1. For teaching purposes, to discover which pupils are bright and capable of doing better school work than they are doing and to discover which pupils are dull and may be attempting work beyond their capacity.

2. For administrative purposes, to regrade pupils so that the pupils in any one grade will be more homogeneous in mental ability and therefore able to progress at more nearly the same rate than otherwise.

3. For administrative purposes, to classify pupils into separate groups within grades in order that the brighter or the more mature pupils may be given an enriched curriculum and in order that the duller or the less mature pupils may be allowed to progress at a slower rate.

Such classifying is sometimes done on the basis of score (dividing the pupils on the basis of mental maturity) and sometimes on the basis of IQ (dividing the pupils on the basis of brightness). The first of these methods is recommended.

4. For research purposes, to obtain two or more groups of equal mental ability or brightness which may be given different methods of instruction for the purpose of determining which method is superior.

5. For guidance purposes, to assist pupils to choose wisely in planning their educational, recreational, and vocational programs.

6. For administrative purposes, to determine the comparative mental status of pupils of different schools or localities.

(Test Service Bulletin No. 77, published by World Book Company and available upon request, gives further information covering the Intelligence Quotient.)

A SURVEY

of

OPINIONS CONCERNING

TERM TESTING AND CONTINUAL TESTING

in a

COMPARATIVE STUDY

of

two

Grade IX

EXAMINATION SYSTEMS

November, 1966

Dear Sir (Madam):

I am at present making a comparison between two systems of testing used in Grade IX. The Faculty of Education of the University of Manitoba has approved this piece of research and has appointed A. M. McPherson, assistant professor, as my adviser.

The opinions of the people vitally concerned with testing programs are necessary in order to make full comparisons. Your co-operation in filling out this questionnaire would be most helpful and greatly appreciated.

If you wish a copy of the findings, write to me and such information will be furnished as soon as available.

Sincerely yours,

A. P. Hildebrand
Altona, Manitoba

Please DO NOT PUT YOUR NAME ON THIS PAPER. Respondents are to remain anonymous. You are asked for frank, honest opinions. The information in #1, 2, 3 of Section A is required only for classification of data. Section B poses a series of questions which can be answered by ticking off either "term" or "continual". If you feel that they are equally good or that neither really applies, tick off the answer "no opinion" because I am comparing the two systems and would like to establish their strength and weakness relative to each other.

For the purpose of this survey, <u>term testing</u> is defined as a system of examinations where students are tested after a given period of time, usually a matter of three months. Such a period of time is called a term. <u>Continual testing</u> is taken to denote a system of testing where no such fixed period of time exists between tests, but where tests are given at the discretion of the instructor. A minimum number per month is required in order to report progress to the parents. Every assignment that is marked by the teacher or marked in class under the guidance of the teacher is recorded and used to arrive at the mark for the month. Testing may also consist of objective tests, review tests, timed tests, chapter tests or any other manner of checking devised by the instructor.

Section A - for the classification of data.

1. What is your sex?
   - Male_____
   - Female_____

2. What is your occupation?
   - Student_____
   - Teacher_____
   - Parent_____
   - Inspector_____

3. If parent, do you have children in the following grades?
   - IX_____
   - X_____
   - XI_____
   - XII_____

Section B - your opinions regarding tests.

1. Which testing system do you prefer?
   - Term_____
   - Continual_____
   - No opinion_____

2. Which testing system provides the better motivation to complete assignment?
   - Term_____
   - Continual_____
   - No opinion_____

3. Which testing system provides the better motivation to review continually?
   - Term_____
   - Continual_____
   - No opinion_____

4. Which testing system provides the better motivation to study out of books other than text books?

Term_____

Continual_____

No opinion_____

5. Which testing system causes a student to actually complete assignments more regularly?

Term_____

Continual_____

No opinion_____

6. Which testing system causes a student to actually review continually?

Term_____

Continual_____

No opinion_____

7. Which testing system causes a student to actually study out of books other than text books?

Term_____

Continual_____

No opinion

8. Which testing system produces better study habits?

Term_____

Continual_____

No opinion_____

9. Which testing system causes less cramming?

Term_____

Continual_____

No opinion_____

10. Which testing system produces a better attitude towards examinations?
Term_____
Continual_____
No opinion_____

11. Which testing system puts a student more at ease during exams?
Term_____
Continual_____
No opinion_____

12. Which testing system puts a student more at ease from day to day?
Term_____
Continual_____
No opinion_____

13. Which testing system gives a student a false sense of security?
Term_____
Continual_____
No opinion_____

14. Which testing system causes students to become indifferent to study habits?
Term_____
Continual_____
No opinion_____

15. Which testing system causes students to become indifferent to examinations?
Term_____
Continual_____
No opinion_____

16. Which testing system tends to tempt a student to cheat?

Term_____

Continual_____

No opinion_____

17. Which testing system tends to put too much pressure on a student?

Term_____

Continual_____

No opinion_____

18. Which testing system tends to assist the weak student?

Term_____

Continual_____

No opinion_____

19. Which testing system requires too much homework?

Term_____

Continual_____

No opinion_____

20. Which testing system is too time consuming?

Term_____

Continual_____

No opinion_____

21. Which testing system places too much emphasis on testing?

Term_____

Continual_____

No opinion_____

22. Which testing system has too much emphasis placed on marks?

Term_____

Continual_____

No opinion_____

23. Which testing system has too much time spent on checking and taking up results with students?

Term_____

Continual_____

No opinion_____

24. Which testing system produces better term results?

Term_____

Continual_____

No opinion_____

25. Which testing system produces meaningless term results?

Term_____

Continual_____

No opinion_____

26. Which testing system produces better June departmental examination results?

Term_____

Continual_____

No opinion_____

27. Which testing system produces meaningless yearly averages?

Term_____

Continual_____

No opinion_____

28. Which testing system gives parents a better idea of how the student is getting along? Term_____

Continual_____

No opinion_____

29. Which testing system gives teachers a better idea of how the student is getting along? Term_____

Continual_____

No opinion_____

30. Which testing system gives the student a better idea of how he himself is getting along? Term_____

Continual_____

No opinion_____