Whole No. 2 Series I, No. 2

HARVARD MONOGRAPHS IN EDUCATION

# THE MARKING SYSTEM OF THE COLLEGE ENTRANCE EXAMINATION BOARD

BY
L. THOMAS HOPKINS
Graduate School of Education
Harvard University



Series I No. 2
STUDIES IN EDUCATIONAL PSYCHOLOGY
AND
EDUCATIONAL MEASUREMENT
Edited by
WALTER F. DEARBORN

OCTOBER, 1921

Published by
THE GRADUATE SCHOOL OF EDUCATION
HARVARD UNIVERSITY, CAMBRIDGE 38, MASS.

# HARVARD MONOGRAPHS IN EDUCATION

## SERIES I

### Studies in Educational Psychology and Educational Measurement.

Manuscripts for Series I should be addressed to Professor WALTER F. DEARBORN, Psycho-Educational Clinic, Palfrey House, Oxford Street, Cambridge 38, Mass.

Remittances should be made by check or money order to The Graduate School of Education, Harvard University, Cambridge 38, Mass.

Series I of the Harvard Monographs in Education has been established for publishing the results of statistical and experimental studies and of educational tests in the general fields of educational psychology and educational measurements.

The numbers are as follows:

1. A Comparison of the Intelligence and Training of School Children in a Massachusetts Town. E. A. SHAW and E. A. LINCOLN. IN PRESS.

2. The Marking System of the College Entrance Examination Board. L. THOMAS HOPKINS. Postage prepaid, 40 cents.

Whole No. 2 Series 1, No. 2

HARVARD MONOGRAPHS IN EDUCATION

# THE MARKING SYSTEM OF THE COLLEGE ENTRANCE EXAMINATION BOARD

BY
L. THOMAS HOPKINS
Graduate School of Education
Harvard University

Series 1 No. 2

STUDIES IN EDUCATIONAL PSYCHOLOGY
AND
EDUCATIONAL MEASUREMENT
Edited by
WALTER F. DEARBORN



OCTOBER, 1921

Published by
THE GRADUATE SCHOOL OF EDUCATION
HARVARD UNIVERSITY, CAMBRIDGE 38, MASS.

This study was undertaken at the suggestion of Professor Walter F. Dearborn of the Harvard Graduate School of Education. The writer is greatly indebted to him for assistance and counsel during the progress of the investigation.

# The Marking System of the College Entrance Examination Board

This study represents an investigation into the distribution of the marks of the College Entrance Examination Board for the years 1902 to 1920 inclusive. It was made in order to discover if there were any grounds for the strong criticism of the college entrance examinations by New England educators, more especially secondary school principals and teachers. It is published at this time because the Board in its Twentieth Annual Report recognized the existence of sudden and violent fluctuations, from year to year, in the results of the examinations, in many subjects, and voted to employ expert assistance to aid in determining the specific causes.

### SCOPE OF THE STUDY.

The subjects selected were English Readings, Elementary French, Elementary Algebra and Plane Geometry for the reason that they were offered by nearly all candidates, thus involving a relatively large number of cases. The arrangement of marks has been altered somewhat. A sample distribution as published by the board is as follows:

| | 90-100 | 75-89 | 60-74 | 50-59 | 40-49 | 0-39 |
|---|---|---|---|---|---|---|
| Solid Geometry | | | | | | |
| 1916/1152* | 1.8% | 6.1% | 18.2% | 12.8% | 14.1% | 47% |

Most of the larger colleges and universities admit on a mark of 60 or above while some of the smaller institutions will accept as low as 50. Assuming that the distribution ought to approximate the normal, for reasons which will be established later, and that anyone rated below 50 has failed to pass, the data in each case have been corrected from the above to read as follows:

| | 90-100 | 75-89 | 60-74 | 50-59 | 0-49 |
|---|---|---|---|---|---|
| Solid Geometry 1916/1152 | 1.8% | 6.1% | 18.2% | 12.8% | 61.1% |

The highest number of cases involved in any distribution was Elementary Algebra 1920/5249 and the lowest Elementary French 1902/509 with only 13 out of the 76 instances when the number fell below 1000.

### FACTS BROUGHT TO LIGHT.

The following significant facts were discovered:

(*a*) Out of 76 distributions graphed every one is bimodal with the exceptions of:

English Readings 1902/800, 1906/1380, 1907/1661, 1908/1698, 1912/1731.

---

* In this and all similar cases the numerator of the fraction represents the year and the denominator the number of persons taking the examination.

In every instance the second mode in the distribution occurs in the assignment of the lowest marks and very often contains a greater percentage of cases than the one in the middle.

(*b*) Every distribution is skewed negatively or toward the lower end of the distribution of marks except:

Elementary Algebra 1906/1180, 1913/1916, 1918/3826.
Elementary French 1909/1196, 1916/2872.
English Readings 1903/996.

(*c*) The order in which the subjects approximate the normal distribution is as follows: English Readings, Elementary French, Elementary Algebra, Plane Geometry. In Figs. I and II are reproduced twenty selected graphs, five for each of the above subjects respectively.

### EFFECT OF YEARLY INCREASE.

Various reasons suggested themselves as to why the results are so far from those expected. Bimodal distributions usually indicate a poor selection of cases. As the second mode in every instance is in the lower end or failure group, this might be caused by the influx of a large number of unprepared persons in the hope of slipping by. This explanation is discarded, however, for (*a*) the data show that this does not occur at intervals but appears regularly in all subjects, (*b*) the yearly increase in the number of candidates, with the exception of 1916, has been relatively constant as is shown in Table I.

### RECOMMENDED CANDIDATES.

If all candidates of doubtful preparation could be eliminated a different result might be obtained. Consequently graphs were made for the years 1912-1916 inclusive for "only those candidates who were recommended for examinations on the ground of full and satisfactory preparation."*

It was found, however, that

(*a*) In Elementary Algebra and Plane Geometry, every distribution is bimodal, seven out of every ten are skewed negatively or toward the lowest grades, while the other three are skewed positively or toward the highest grades.

(*b*) Of the five in Elementary French, four are bimodal and three are skewed positively.

(*c*) In English Readings only one, 1916/2431, is bimodal, all the others tending roughly toward the normal.

---

* Further study of the group could not be made, as only these limited data are published by the Board.

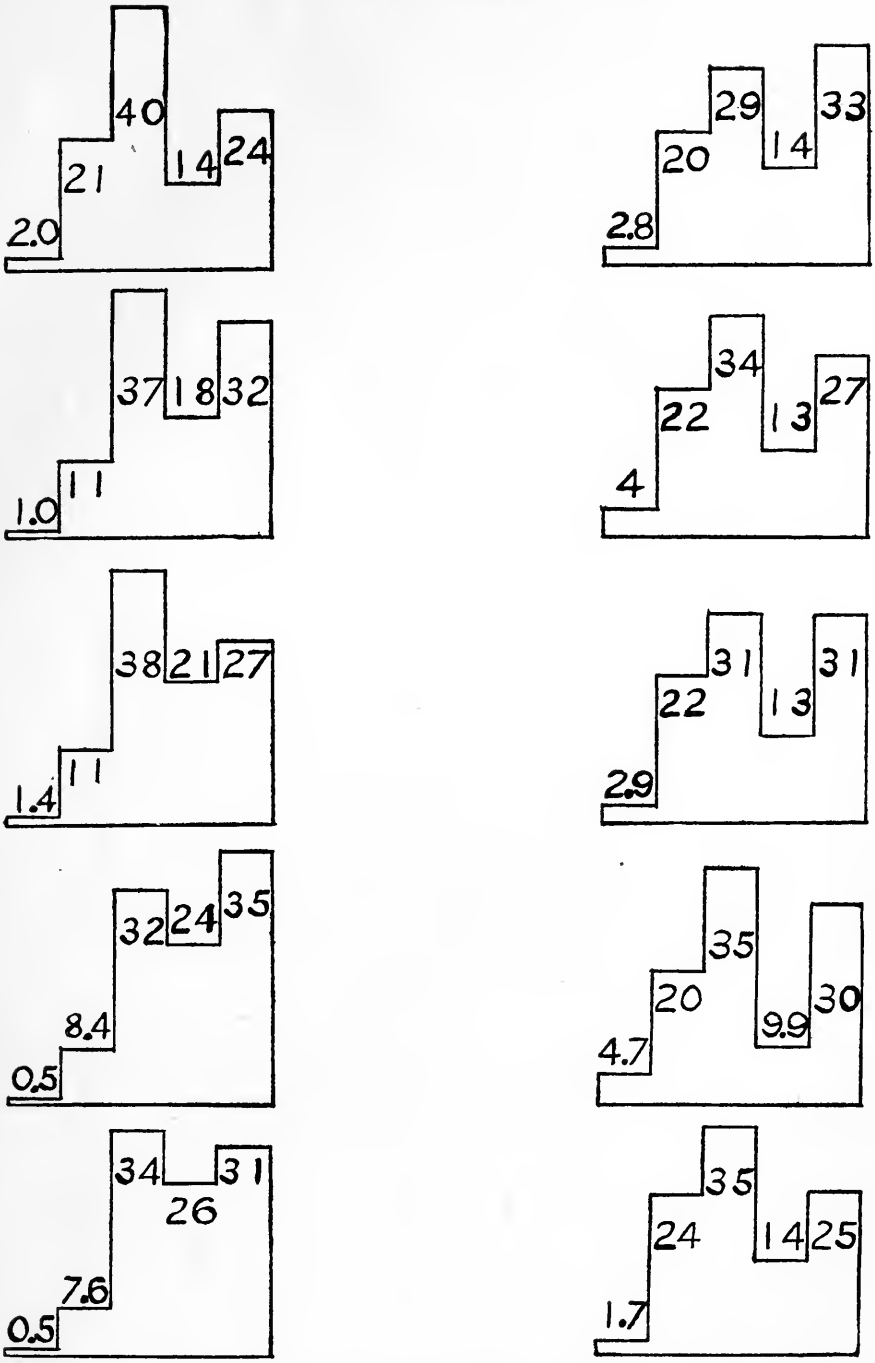


Fig. I—Graphs in the first column represent English Readings, the second Elementary French. The different divisions are as follows: 90-100, 75-89, 60-74, 50-59, 0-49. The figures show the percentage of cases.

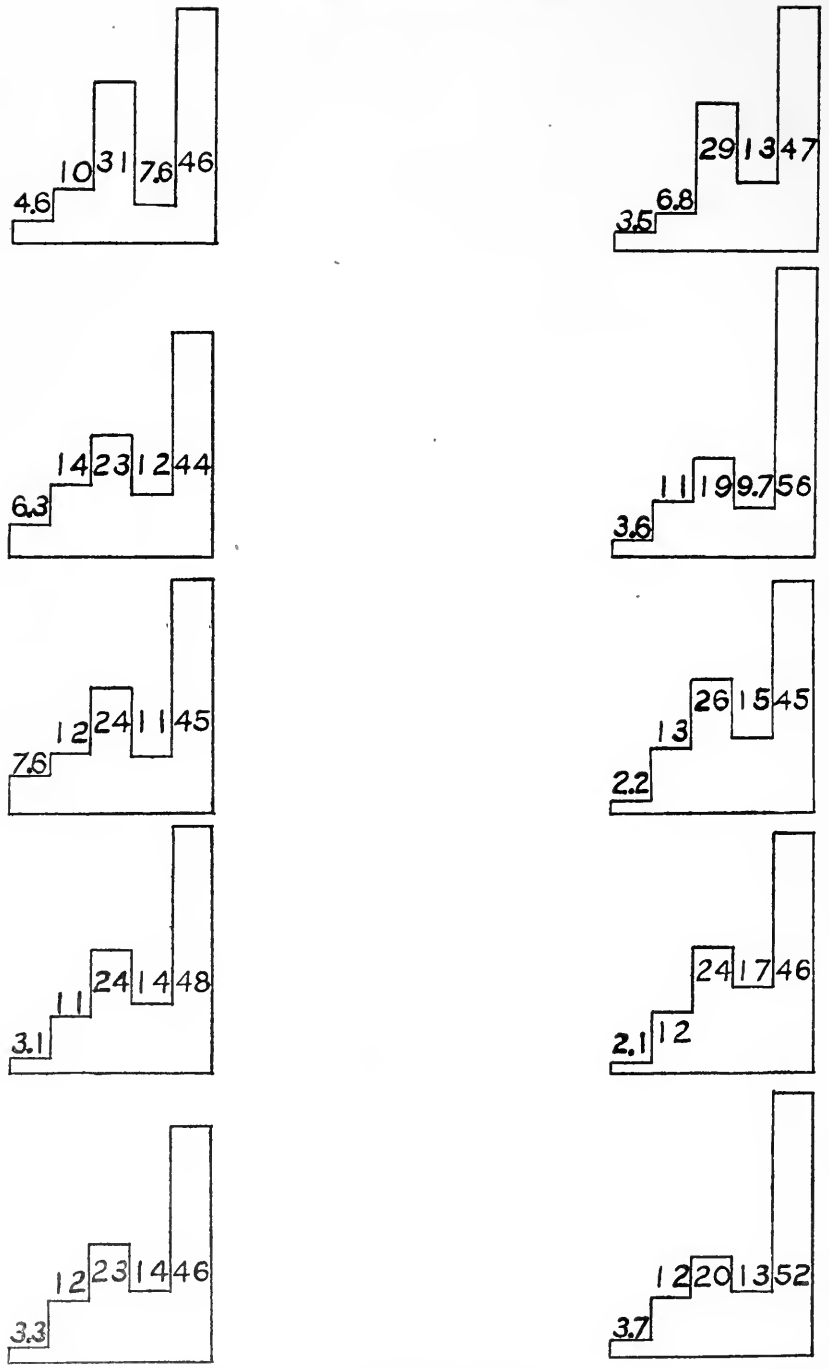


Fig. II—Graphs in the first column represent Elementary Algebra, the second Plane Geometry. Divisions as in Fig. I.

# Table 1

## Increase in the Number Taking Examinations

| Year | English Reading | | Elementary French | | Elementary Algebra | | Plane Geometry | | In All Subjects | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Total | Increase | Total | Increase | Total | Increase | Total | Increase | Total | Increase |
| 1902 | 800 | | 509 | | 810 | | 782 | | 11744 | |
| 1903 | 996 | 196 | 625 | 116 | 973 | 163 | 927 | 145 | 14263 | 2519 |
| 1904 | 1033 | 37 | 661 | 36 | 1060 | 87 | 994 | 67 | 15275 | 1012 |
| 1905 | 1244 | 211 | 742 | 81 | 1079 | 17 | 940 | 54* | 16189 | 914 |
| 1906 | 1380 | 136 | 854 | 112 | 1180 | 101 | 1069 | 129 | 17467 | 1278 |
| 1907 | 1661 | 281 | 1044 | 190 | 1291 | 111 | 1206 | 137 | 20034 | 2567 |
| 1908 | 1698 | 37 | 1143 | 99 | 1324 | 33 | 1171 | 35* | 20607 | 573 |
| 1909 | 1706 | 8 | 1196 | 53 | 1445 | 121 | 1425 | 254 | 22208 | 1601 |
| 1910 | 1748 | 42 | 1166 | 30 | 1482 | 37 | 1340 | 85* | 22189 | 19* |
| 1911 | 1814 | 66 | 1317 | 151 | 1655 | 173 | 1586 | 246 | 22932 | 743 |
| 1912 | 1731 | 83* | 1153 | 164* | 1476 | 179* | 1473 | 113* | 20568 | 2364* |
| 1913 | 1795 | 64 | 1299 | 299 | 1960 | 484 | 1743 | 270 | 22975 | 2407 |
| 1914 | 1963 | 168 | 1424 | 125 | 1233 | 727* | 1833 | 90 | 23350 | 375 |
| 1915 | 1734 | 229* | 1441 | 17 | 1380 | 147 | 1936 | 103 | 23990 | 640 |
| 1916† | 4163 | 2429 | 2872 | 1431 | 3179 | 1799 | 3775 | 1739 | 47842 | 23852 |
| 1917 | 3327 | 836* | 2284 | 588 | 2851 | 328* | 3179 | 596* | 37992 | 9850* |
| 1918 | 3399 | 72 | 3211 | 927 | 3826 | 975 | 3832 | 653 | 41621 | 3629 |
| 1919 | 3582 | 183 | 3983 | 772 | 4181 | 355 | 4442 | 610 | 44406 | 2785 |
| 1920 | 2733 | 749* | 4883 | 900 | 5249 | 1068 | 5227 | 785 | 48449 | 4043 |

NOTE:—*Decrease.

†The large increase in all subjects for 1916, as given in the Sixteenth Annual Report, is due chiefly to the joint action of Harvard, Yale and Princeton Universities, who agreed to discontinue their own June examinations and accept the results of the comprehensive papers prepared by the Board.

It is very evident from this that there is slight improvement in the ratings of the recommended candidates in English Readings and Elementary French but none in Elementary Algebra and Plane Geometry. The difference, however, is not marked enough to conclude that it is due to better preparation.

### TOTAL YEARLY RANKS.

Theoretically, as the number of cases increases the nearer the distribution should correspond to the normal. Graphs were prepared showing the distribution of the total number of marks given for all subjects from 1902 to 1920 inclusive for all candidates, and from 1912 to 1916 for recommended candidates only. These show that in every case, (*a*) the distribution is bimodal, (*b*) it is skewed toward the lower end. Fig. III gives a selected list of graphical representations for totals of different years.

If all of the marks assigned in all subjects from 1902 to 1920 inclusive were combined into one grand total average distribution it would be as follows:

| Grand Total | 90-100 | 75-89 | 60-74 | 50-59 | 0-49 |
|---|---|---|---|---|---|
| 445,620 | 4.78% | 18.34% | 31.14% | 13.78% | 31.96% |

In other words out of 445,620 cases only 4.78% received the highest grade while 31.96% failed. How many of the latter tried over again and succeeded there are no data to show.

A grand total average distribution for only those candidates recommended on the ground of full and satisfactory preparation as published for 1912 to 1916 inclusive is

| Grand Total | 90-100 | 75-89 | 60-74 | 50-59 | 0-49 |
|---|---|---|---|---|---|
| 87,642 | 6.35% | 22.32% | 32.28% | 13.69% | 25.36% |

This is slightly better than the one given above, but considering the fact that the individuals involved here were highly selected, a failure of one-fourth, or 21,910 cases out of 87,642, places upon the Board the responsibility for a condition which is far reaching in its social and economic effects.

### SELECTED DISTRIBUTIONS.

That the reader may have some samplings of extreme variations as a basis of comparison a selected list of graphs is given in Fig. IV. These are taken from different subjects and different years. The lowest number of cases involved is 641 while the highest is 2063.

### WHAT WAS EXPECTED.

As was said at the beginning of this article, it was expected that the results would approximate the normal distribution. Briefly the evi-

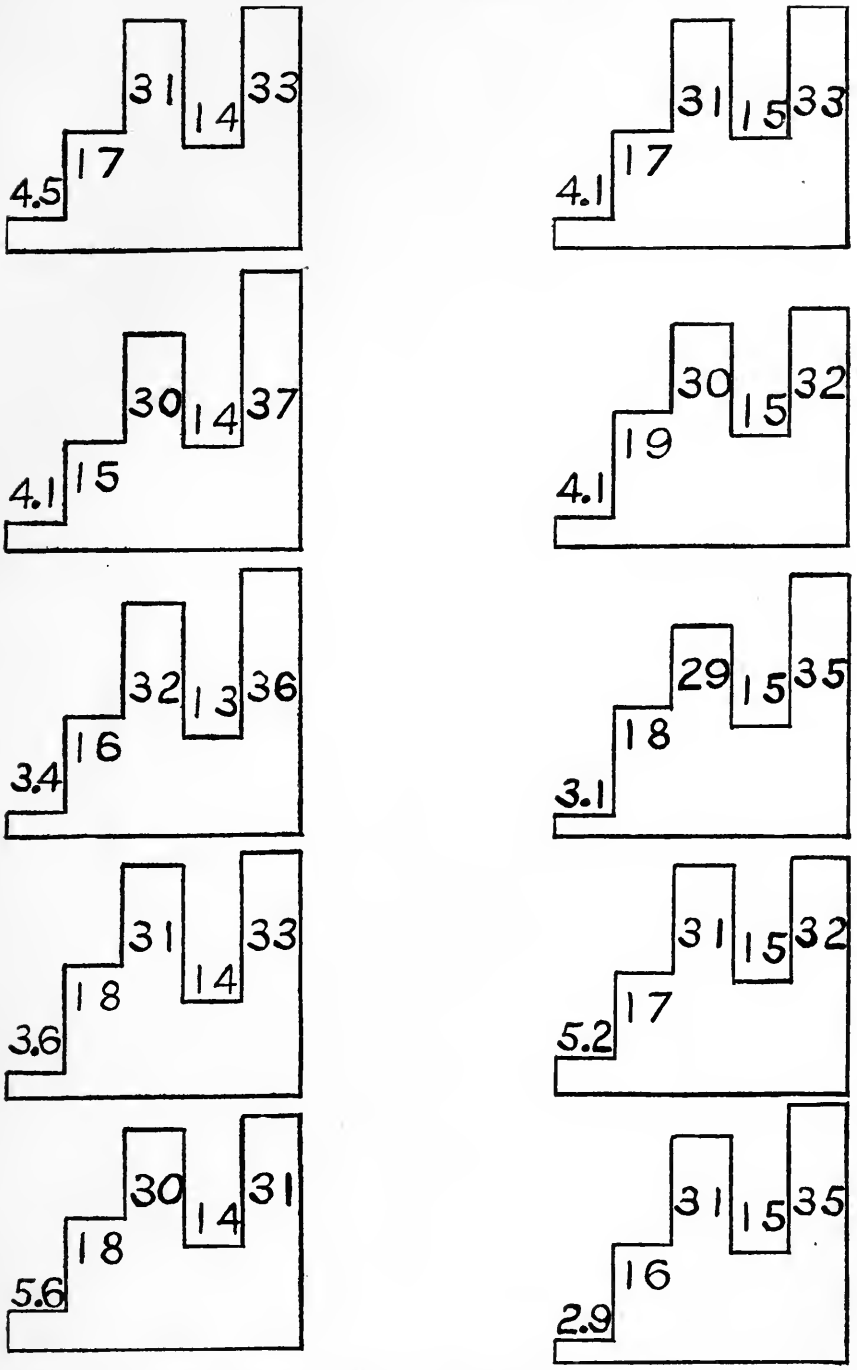


Fig. III—Totals for different years. Number of marks assigned will be found in Table I. Divisions as in Fig. I.

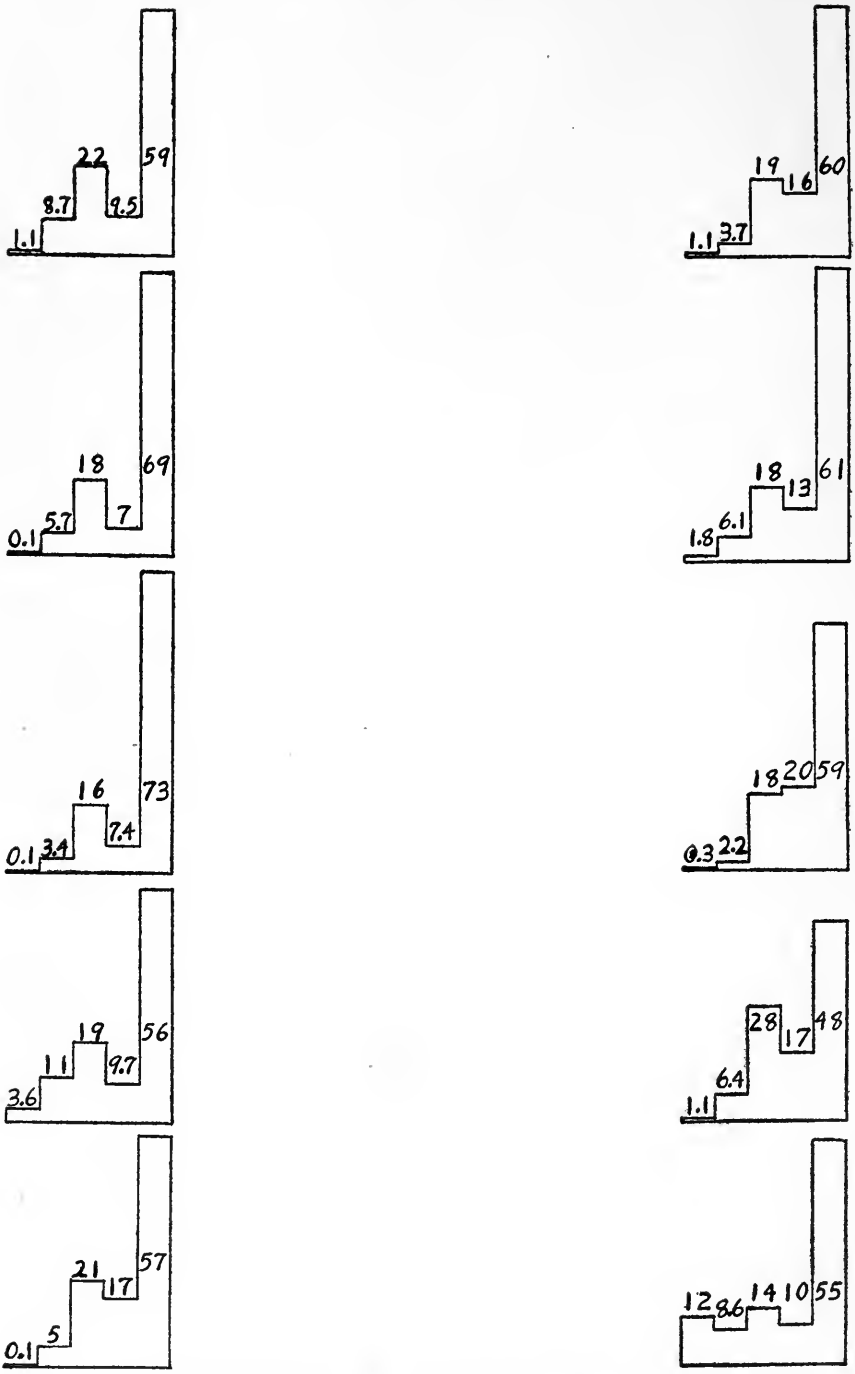


Fig. IV—Selected distribution in different subjects and years. The range of cases involved is from 641 to 2063. Divisions as in Fig. I.

dence supporting this is as follows: (a) Physical differences approximate the normal curve* as do mental characteristics,† (b) Marks, representing, as they do, estimates of mental abilities, are themselves distributed according to the same frequencies as the abilities they are designed to represent,‡ (c) The normal distribution of marks is the one usually found when a fairly large number of students are graded.§

Concluding then that the assignment of any relatively large number of grades ought to approximate the normal distribution and steadily so as the number increases over 500, this further question remains: What is the best method of dividing this distribution into groups for translating standing into a scale of marks? After a careful examination of all possible schemes we have concluded that the five division one is best. This is based on the orientation of a large number of cases around a central group whose accomplishment is considered median or average. Above and below lie groups of smaller size containing superior and inferior students in relation to the average and above and below these the still smaller groups of exceptions or failures.

The method of dividing our theoretical distributions into the five divisions which we will represent by the letters A, B, C, D, E, would be as follows: Find the median of the distribution and lay off on the base, on either side, the distance of 1 P. E. Within the area embraced by this $\pm$P. E. there will fall 50% of the total number of cases. This would represent the center or average or C group. Now lay off on either side of $\pm$ P. E. a distance equal to 2 P. E. Each one of the areas thus designated will contain 23% of these cases,‖ and would be represented by the letters B and D respectively. Again laying off the distance of 2 P. E. on either side we will reach the limits of the normal curve as for all practical purposes the ordinate may be taken as zero when the abscissa is 5 P. E. The last two divisions just made would each contain 2% of the total number of cases and would be represented by the letters A and E. The relationship between the cases represented by the five divisions of our normal probability integral and our marking system would now be as follows:¶

| A | B | C | D | E |
|---|---|---|---|---|
| 2% | 23% | 50% | 23% | 2% |

* Brooks: The Foundation of Zoölogy, pp. 156-157, and Yule: An Introduction to the Theory of Statistics, p. 84.

† See the distribution of the IQ's of 905 unselected children 5-14 years of age in Terman: The Measurement of Intelligence, p. 66.

‡ Dearborn: School and University Grades. University of Wisconsin Bulletin No. 368.

§ Dearborn, Ibid, also Foster: The Administration of the College Curriculum, pp. 250-300.

‖ A table of the values of P. E. of the normal probability integral will be found in Rugg: Statistical Methods Applied to Education, p. 391.

¶ This was the division used by Buckingham in the standardization of the Buckingham Spelling Scale.

In like manner if we should lay off on either side of the mean the distance of A. D. we would find the following distribution:

| A | B | C | D | E |
|---|---|---|---|---|
| 2% | 20% | 56% | 20% | 2% |

or if we should take for our unit .5$\sigma$ and then lay off 1$\sigma$ on either side our relationship would be as follows:*

| A | B | C | D | E |
|---|---|---|---|---|
| 7% | 24% | 38% | 24% | 7% |

What is more commonly used by writers than either of the two preceding is to lay off the distance Q on each side of the mean. We would then have:†

| A | B | C | D | E |
|---|---|---|---|---|
| 3% | 22% | 50% | 22% | 3% |

One of the first thoro treatments of variation in the marking of examinations was published by an English economist, Professor F. Y. Edgeworth, in the Journal of the Royal Statistical Society, September, 1888. This paper showed that there is a probable error of 3% and a possible error of 9%, in assigning a mark as representative of a student's real proficiency. Professor Edgeworth argued as a remedy that marks should be distributed according to the normal probability curve, but offered no suggestions as to its division. Many of the later writers, however, made definite divisions as given below:

| | | A | B | C | D | E |
|---|---|---|---|---|---|---|
| Cattell | (1905)‡ | 10% | 20% | 40% | 20% | 10% |
| Meyer | (1908) | 3 | 22 | 50 | 22 | 3 |
| Dearborn | (1910) | 2 | 23 | 50 | 23 | 2 |
| Foster | (1911) | 3 | 22 | 50 | 22 | 3 |
| Slosson | (1911) | 3 | 22 | 50 | 22 | 3 |
| Smith | (1911) | 10 | 15 | 50 | 15 | 10 |
| Ruediger | (1912) | 4 | 24 | 44 | 24 | 4 |
| Gray | (1913) | 7 | 22 | 42 | 22 | 7 |
| Cajori | (1914) | 7 | 24 | 38 | 24 | 7 |
| Starch | (1917) | 7 | 24 | 38 | 24 | 7 |

* This was the division used by Ayres in the construction of the Ayres Spelling Scale.

† Tables of the values of AD, $\sigma$ and Q of the normal probability integral will be found in Thorndike: Mental and Social Measurements, pp. 219, 220.

‡ Professor Cattell recognized the P. E. distribution of cases. He altered the percentages to more nearly meet the needs of classroom teachers who deal with small numbers, usually not exceeding 40.

A study of Figures 1 to IV inclusive will show no such relationship between the percentage of cases in the five divisions as is brought out here. Indeed one is amazed at the remarkable extent of divergence.

**EFFECT OF READING METHODS ON THE DISTRIBUTION.**

A number of examiners and readers have been consulted, from whom the following facts have been ascertained:

(*a*) Any paper marked between 50 and 60 by a reader is re-read by one or more before a permanent rating is given. This is due to the fact that the passing mark for some of the larger universities is 60 while that of many smaller colleges is 50. The re-examination of the paper is to determine whether the writer shows sufficient actual knowledge of subject matter and indicates enough potential possibilities of development to profit by the work offered in that department of a large university. If in the opinion of the examiner he does not, then the mark is below 60 which will admit only to the smaller colleges.

(*b*) Any paper marked over 90 by a reader is re-read by one or more readers before it is given its final mark. This is due to the fact that many prizes depend upon the highest awards.

(*c*) Any paper originally marked between 60 and 90 is never re-read except in rare instances when the rating is only a few points above 60.

(*d*) At the beginning the examiners agree on a value to be assigned to each question. There are two different methods of determining this. In some cases it is arrived at as follows: (1) Accepting 100 as the highest possible score, when there are ten questions each is given a value of 10. If there are eight questions each is given a value of 12½. When there are two or more parts to any question each part is given a proportion of the value assigned to the question as a whole, i. e. if there were ten questions the value of each would be 10. If one were divided into two parts, 5 would be given to each part. (2) In other instances the rating assigned is arrived at by taking the composite evaluation of each question by the readers. A clear exposition of this method as applied to French will be found in an article by Professor Donald C. Stuart of Princeton in the Bulletin of the New England Modern Language Association, September 1917.

That this method of reading the papers is a contributing cause of the poor distribution of marks is evident for, (a) no conferences are held between the examiners and readers to agree on the interpretation and value to be assigned to questions, (*b*) no attempt is made to standardize values of questions by considering the percentage of answers correct or incorrect, (c) the principle is not recognized that the assignment of

marks aggregating 1000 to 5000 in a subject, or 11,000 to 44,000 for a yearly total, ought to conform to the curve of error and hence no attempt is made to check up or correct results on the basis of the normal distribution.

## CONCLUSION.

The facts seem to show clearly that, (a) only in rare instances, in the subjects studied, does the assignment of marks nearly approximate the normal, (b) the same condition holds true for the annual total for all subjects, (c) the results in cases where the pupils taking the examination are recommended by their school authorities on the ground of full and satisfactory preparation are only slightly improved, (d) this cannot be due to an influx of unprepared candidates as the increase in numbers each year is relatively constant and the poor distribution is found annually from 1902 to date, (e) the method of reading and scoring the papers, especially the lack of standardization of values and corrections in conformity with the curve of error, is a very natural factor in causing the existing conditions, (f) the suggestion is made that some approximation to the normal curve offers the best basis for solving present irregularities. This need not affect the passing marks as they may still be determined by such principles as govern them at the present time, altho a reconsideration of these might well be made by the Board.

Finally, in view of the large number of cases, no sufficient justification exists for the wide difference in the relative percentages assigned in the different subjects. Whether the distribution approximates the curve of error, or some other form, a certain uniformity in the different subjects may reasonably be expected. To accomplish this there must be co-operation between examiners and readers in the different subjects.

The writer wishes to emphasize the fact that this article does not claim to present an exhaustive study of the marks given by the College Entrance Examination Board. There are many phases of the subject which have not been touched. Sufficient evidence has been produced, however, to show the existence of an unwarranted condition and it is hoped the movement already inaugurated by the Board will result in a definite, workable plan for improvement.

## SELECTED REFERENCES


Boring, E. G. Marking System in Theory. Pedagogical Seminary 21: 269-77, 1914.

Breed, F. S. Administering the Relative Marking System. School and Society, 5: 474-9, 1917.

Cajori, F. A New Marking System and Means of Measuring Mathematical Abilities. Science, 39: 874-881, 1914.

Carter, R. E. Correlation of Elementary Schools and High Schools. Elementary School Teacher, 12: 109-118.

Cattell, J. McK. Examinations, Grades and Credits. Pop. Sci. Mon., 66: 367-78, 1905.

Dearborn, W. F. Relative Standing of Pupils in the High School and in the University. University of Wis. Bull. No. 312.

——— School and University Grades. University of Wis. Bull. No. 368

Dickson, J. D. H. Percentages in School Marks. Nature 81: 367- 1909.

Edgeworth, F. Y. The Element of Chance in Examinations. Journal of the Royal Statistical Society, 1890, 460-75, 644-73.

Ferry, Dean Grading College Students. Williams College Bull., Series 8, No. 5, 1911.

Finkelstein, I. E. The Marking System in Theory and Practice, Warwick and York, 1913.

Fisk, T. S. Analysis of the Examination of 1911 of the College Entrance Examinations Board. Educational Review, 43: 155-67.

——— Reports of the Secretary of the College Entrance Examination Board, 1902-1920.

Foster, W. T. The Administration of the College Curriculum, Boston, 1911.

Holmes, Henry W. The Teaching of Economics in Harvard University. Harvard Studies in Education, Vol. III.

Inglis, A. J. Variability of Judgments in Equating Values in Grading. Educational Administration and Supervision, January, 1916.

Judd, C. H. A Comparison of Grading Systems in High Schools and Colleges. School Review, 18: 460-70.

Lincoln, E. A. Relative Standing of Pupils in High School, in Early College and in College Entrance Examinations. School and Society 5: 417-20, 1917.

Meyer, Max The Grading of Students. Science, 28: 243-52.

Nickolson, F. W. New Methods of Admission to College. Education, 32: 261-65.

Pettit, W. A. W. Comparative Study of New York High School and Columbia College Grades. Teachers College, Master's Essay, 1912.

Roecker, W. F. An Objective Study of the Rating of Traits in School Achievement. School Review, 22: 406-410, 1915.

Rugg, H. O. Teachers' Marks and Marking Systems. Educational Administration and Supervision, February, 1915.

Russell, J. E. Educational Value of Examination for Admission to College. School Review, 11: 42-54.

Sargent, E. B. Education of Examiners, Nature 70: 63-68.

Sies, Raymond W. Scientific Grading of College Students. University of Pittsburgh Bull., Vol. 8, No. 31, 1912.

Slosson, E. E. A Study of Amherst Grades. Independent, 70: 836-39,

Smith, A. G. A Rational College Marking System. Journal of Ed. Psy., 2: 383-93.

Smith, F. O. A Rational Basis for Determining Fitness for College Entrance. University of Iowa Studies in Education.

Starch, D. Can the Variability of Marks be Reduced? School and Society, 2: 242-43.

Thorndike, E. L. Empirical Study of College Entrance Examinations. Science, 23: 839-45.

Young, W. H. The High Schools of New England as Judged by the Certification Board. School Review, 5:15.

## HARVARD STUDIES IN EDUCATION

Volume I: The Oberlehrer. A Study of the Social and Professional Evolution of the German Schoolmaster. By WILLIAM SETCHEL LEARNED. $1.25.

Volume II: The Appointment of Teachers in Cities. A Descriptive Critical and Constructive Study. By FRANK WASHINGTON BALLOU. $1.50.

Volume III: The Teaching of Economics in Harvard University. A Report Presented by the Division of Education at the Request of the Department of Economics. $2.00.

## HARVARD BULLETINS IN EDUCATION

I. The School System as an Educational Laboratory. By WM. S. LEARNED. 25 cents.

II. Scales for the Measurement of English Composition. By FRANK W. BALLOU. 50 cents.

III. Bridging the Gap: The Transfer Class. By F. W. WRIGHT. 30 cents.

IV. A Selected Critical Bibliography of Vocational Guidance. By JOHN M. BREWER and ROY W. KELLY. 50 cents.

V. A Descriptive Bibliography of Measurement in Elementary Subjects. By HENRY W. HOLMES and others. 50 cents.

VI. Business Practice in Elementary Schools By ROY DAVIS. 50 cents.

VII. Sight-Saving Classes in the Public Schools. By R. B. IRWIN. 35 cents.

**Published by**

THE GRADUATE SCHOOL OF EDUCATION

HARVARD UNIVERSITY, CAMBRIDGE 38, MASS.