THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

DISTRIBUTION OF GRADES

J. L. ZERBE

University of Pittsburgh

In April, 1915, a few weeks before the grading of the work of the night students and a month or so before the final grading of the day students of the School of Applied Industries, Carnegie Institute of Technology, the author, with the approval of the Dean, presented a report to the faculty showing the distribution of 28,257 grades, covering the two semesters, 1912-13, 1913-14, the first semester 1914-15 and the grades of all who had graduated from this school. In the letter of transmission to the Dean, which accompanied the report to the faculty, it was stated: "It is thought that this information in the hands of the faculty at this time will insure a more uniform distribution of grades for the final records of this term." The present paper is to indicate the results of this report on the final grading of the work of the second semester.

Space does not permit the publication of the report in its entirety but a summary of facts and the emphasis placed on the matter will be given. In the report the following tabulations were illustrated by graphs.

Total Distribution in Per Cents

Grade	1912-13	1913-14	1st Sem 1914-15	Graduates	Total
H ¹	5%	4%	4%	12%	7%
M	16	17	17	14	16
C	42	39	44	55	43
P	14	14	16	16	15
F	9	9	8	2	8
D	5	4	5	1	4
R	6	5	5		4
Fa	3	8	1		4

The unsatisfactory grades are bracketed and indicate that a large percentage fail to do satisfactory work and that a great waste of teaching energy must result when so large a percentage of failures is due to absences. The survival of the fittest is shown in the grades assigned the graduates.

A more detailed study suggests a separation of the grades assigned for the day and night courses and is here presented.

	1912-13		1913-14		1st Sem. 1914-15	
	Day	Night	Day	Night	Day	Night
H	3%	8%	3%	6%	4%	4%
M	15	14	17	17	16	19
C	46	35	45	33	46	37
P	15	12	15	12	17	15
F	10	9	10	9	9	6
D	3	7	3	5	4	6
R	4	9	5	4	3	8
Fa	4	7	2	14	1	5
	21		20		17	
	32		32		25	

The close proximity of the distribution for the day and night courses respectively from year to year indicates that the grading for the day and the night students ought to be based on different standards.

Different standards for the type of work is indicated by a study of the grades assigned for academic work and those for the shop work. Because of the uniformity of distribution from year to year a detailed study was made only of the term 1913-14. The following tabulation shows difference of grading used by academic teachers and shop instructors.

	H	M	CO	P	F	D	R	Fa
Practical or shop course.	5	19	49	12	5	6	2	1
Summarizing total		25	49			26		
Theoretical or academic.	3	16	30	16	12	5	7	13
Summarizing total.		19	30			51		

If fifty-one per cent. of all grades assigned is valued as passing or below, as is indicated for the academic subjects, is it not to be inferred that the standard is too high for the class of students taking the work? The combined judgment of the fifty-five instructors considered in 1913-14, is that twenty-six per cent. of all grades assigned is below passing, while the distribution for the passing grades by the individual instructors ranges from fifty-five to two per cent. of all grades granted by the respective instructors.

From the difference in the distribution of the grades for the practical work from that of the academic subjects, it might be inferred that the students are a specialized group and emphasise

¹ The manual of instructions concerning grades and ratings issued by the Carnegie Institute of Technology presents the following valuation for the letters used: Honor (H) 95 to 100 per cent.; Merit (M) 85 to 94 per cent.; Credit (C) 70 to 84 per cent.; Passed (P) 60 to 69 per cent.; Failure (F) 40 to 59 per cent.; Repeat (R) 0 to 39 per cent.; Deferred (D) Work unfinished for reasons satisfactory to the Dean. Failure due to Absence (Fa) is used by some instructors and is not found in the manual.

the motor rather than the mental side of their training; or that they, perhaps, do not represent the average in mental ability. Since the academic curve is skewed toward the failure point it might be further argued that they have not the mentality nor the training to properly pursue the academic subjects. Plate 1 shows a distribution of mental ability as determined by a series of psychological tests given to one hundred eighty-three first year students, whose scholastic grades or class standing has been considered in the above figures. The graph is constructed from the distribution of ability for distracted² and undistracted³ attention, ability to interpret written⁴ and oral⁵ instructions, immediate memory under distraction,⁶ logical memory,⁷ association,⁸ vocabulary,⁹ learning capacity,¹⁰ constructive imagination,¹¹ judgment,¹² and mental reaction time.¹³ These one hundred eighty-

² Cancellation test, No. 77261—A (Stoelting's catalogue) Whipple's Manual, while listening to reading.

³ Cancellation test, No. 77261—B.

⁴ Woodworth and Wells Test, No. 88072.

⁵ Ability to follow instructions in performing the tests.

⁶ Reproduce from memory the passage listened to during Test 1.

⁷ Logical memory beginning "Fire at Boston"

⁸ Free association, number of words written in three minutes.

⁹ Number and quality of words written in Test 7.

¹⁰ Woodworth and Wells substitution test, No. 88042.

¹¹ Original test of author, to fill in missing lines in geometrical figures.

¹² Estimation of number of dots on card.

¹³ Easy opposites No. 77341—C-B.

three students represent, no doubt, the average ability of the entire student body and the results as shown in this curve indicate that their mental ability, as tested, approximates very nearly the curve for the surface of frequency and consequently represents the average distribution of mental activity. If this be so, and if the 183 students are a fair representation of the entire student body, it is evident that the grades as distributed for the shop work are based on a much lower standard than are the grades assigned for the theoretical subjects. Since the academic curve is skewed toward the failure point might it not be inferred that the standard is too high for the average student taking these courses?

A further detailed study leads to a consideration of the assignment for the separate courses. Plate II shows the distribution in per cents for the different courses. It does not however, indicate grades for the separate branches of the courses. Comparing the individual judgments for each subject with the combined judgment of the entire group, it will be noted that few if

any approximate the distribution of the entire faculty. In two very fundamental subjects forty per cent. of all grades given are below passing, while in another subject forty-five per cent. of all grades are below passing. In one academic subject thirty-five per cent. are below passing, while in one of the shop subjects only two per cent. fall below that grade. In another academic subject thirty-six per cent. are failures, while in another shop subject only three per cent. do unsatisfactory work.

Plate III shows the distribution of the grades by the individual instructors. Here the lack of conformity to a standard is especially marked. At the bottom of the plate the distribution of the 9954 grades represents the curve of distribution for the entire faculty. Of the 388 grades given by one instructor not one grade is accorded the value of "H," while another instructor in another subject gave the "H" grade to fifty-one per cent. of the 165 grades he assigned, and the "M" grade to thirty-four. In a different subject another instructor, with no doubt the same students, grants 351 grades, none of which is "H" and only nine per cent. are as high as "M." Of 136 grades given by another instructor fifty-seven per cent. are below passing, while of 165 grades granted by another instructor less than one per cent. is below passing, not one as high as "H" and eighty per cent. are given the grade "C." For 1914-15, this instructor assigned no "H" grades, ninety-three per cent. are "C's" and none is below passing except three per cent. of "D's."

Can such grading be justified on the assumption that the quality of work is different, that some instructors command ability, that the class is more interesting or better prepared, that the subject is more important or fundamental, that the instructor is superior? Or what can be truthfully assumed as the cause of this wide variation? Without some standard, foot-rule, or gauge one reason is just as plausible as another, and just as faulty as any other. The student who receives an "H" from one instructor cannot possibly receive that grade from another who holds his subject so important and his standard so high that no one can be considered perfect enough to merit the highest grade of the school—"Honor." Two of the instructors whose figures are quoted above did not give an "H" grade for the entire term 1912-13 and one of them none for the first term 1914-15. In other words, this instructor has not given an "H" grade in two and one-half years covered by this report. Is it possible that in two and one-half years not one student (who was graded by the large number of grades assigned by this instructor although many of them were honor students) could do work sufficiently well to merit an honor grade in this subject?

President Foster, of Reed Collège, who has made an exhaustive study of the distribution of grades at Harvard University and other institutions of learning has this to say: "Not only are there

extreme variations among different courses, but there are variations in the same courses from year to year that cannot be accounted for apparently by any of our scientific studies in the distribution of abilities among human beings. . . . From Maine to California the administration of college credits, though alike in no other particular, agree in this . . . that its basis is personal rather than scientific."¹⁴ The instructor who considers that out of 388 grades assigned there are fewer than one per cent. that are equal to the distinction of "H" has a very different conception of the requirements necessary to obtain that distinction from the instructor who places fifty-one per cent. of his 165 grades in the honor list. It is unjust to the student, to the instructor, or to the institution at large to distribute scholarships, class honors, or other meritorious "distinctions" on the basis of grades resulting from the personal opinions of instructors.

There is evident need of standardizing judgments in all school procedure and there must be definite scales, absolute units of value for school work, just as the mechanical field of endeavor has the foot-rule, galvanometer, thermometer, etc. Binet and Simon with their graded tests for general intelligence of children have led the way for the scientific study of mental attainments. The Russell Sage Foundation and school surveys have suggested means for measuring general school progress. Ability in school subjects is already being reduced to scientific measurement.¹⁵ Until a measuring scale is established for each subject a more scientific method of grading than the personal judgment standard is necessary in order that all concerned may receive equal and just recognition for school work.

There are several approaches to a scientific method of grading and the one herein described is so nearly scientific that, until the intellectual yard-stick is standardized which will be constant at all times and places and with all trained individuals, it may be used to good advantage. It is a system that was adopted to overcome the persistency of instructors to distribute their grades according to personal opinion. It was when a professor of the University of Missouri "flunked" his entire class, and the Board over-ruled him by passing the entire class, that some one of the faculty urged the adoption of a scientific system of grading.

¹⁴ W. T. FOSTER. *Administration of the College Curriculum*, pp. 262, 266.

¹⁵ DANIEL STARCH, *Educational Measurements*.

Since 1908 when the University of Missouri adopted a method and defined definitely the means of establishing various grades, many institutions have adopted similar plans which have for the most part satisfied the demands of the executives, the faculty, and the student body. The most recent adoption of a plan is by the George Washington University, and the following is taken from their announcement as published in the JOURNAL OF EDUCATIONAL PSYCHOLOGY.

The Committee on Standardizing Grades appointed in June submitted the following report March, 1914, on the following two problems: First, How can the amount of work for each unit of credit be approximately equalized in the various courses? Second, What common standard of grading can the various members of the faculty observe so that they will grade approximately the same standard?

From this report the following is quoted:

"In submitting principles and standards for the solution of these problems, the committee wishes, first of all, to be understood that it does not wish to dictate, or even to suggest, how any member of the faculty should do his work. It not only has no intention of curtailing the legitimate rights and freedom of any teacher, but it desires especially to emphasize that these rights and freedom are sacred; that they are an indispensable condition for the best type of university work. But in schools, colleges and universities the personal side is not the only side of teaching. There is present also a social side, which grows out of the fact that a school is in some fundamental aspects a social unit. The various members of the faculty are all working to contribute piecemeal to the same end. . . . Considered from the social standpoint, the college in common with other schools, performs two interrelated, and although distinguishable, fundamental functions. It (1) educates and it (2) selects. . . .

"The giving of grades to students is only one of a number of means that the school uses in discharging the selective function of education, but it is one of the most important. Like other educational functions it must be done carefully, intelligently and uniformly in order to avoid injustice to the student. The desideratum of uniformity requires not only that each teacher always uses approximately the same standard with all of his students, but that all teachers use the same standard with all of

the students. When this is not done the educational equilibrium of the school is disturbed and injustice is done to the earnest and conscientious student. The less serious the students are, the more they tend to gravitate toward teachers that give the highest grades, and the injustice that this tends to work upon the conscientious student when it comes to the awarding of honors and the recommending for positions is obvious. The giving of many high grades, furthermore, gives many students a false and exaggerated notion of their ability. The grade "A" should be reserved for the very exceptional ability which in the nature of the case is rare.

"The principles underlying a uniform standard of grading are found in the distribution of mental ability as revealed by psychological investigations. These investigations have shown, when sufficiently large numbers are considered, that ability in general or in any particular line is distributed in the form of a bell-shaped curve, technically known as the probability curve or the normal curve of frequency. Letting the base line represent the degrees of ability from the poorest to the best and the vertical lines the number of persons possessing each degree of ability, there is, at best, but a small number of students with excellent ability, a large number with good ability, a relatively larger number with medium or average ability, a smaller number with sub-medium but passing ability and a smaller number with distinctly unsatisfactory ability.

"There are, of course, no sharp dividing lines between these groups. The grade of A or excellent should be assigned to about 4 per cent. of the students, B or good to about 24 per cent., C or medium to about 44 per cent., D or sub-medium to about 24 per cent. and E or failure to about 4 per cent. It is quite likely that the percentage of failure in the lower classes may properly be somewhat higher than that in the upper, with corresponding changes in the other percentages, and failures may perhaps also properly be more frequent in professional schools than in liberal culture schools. Because of its immediate social responsibility, it is the duty of the professional school to apply the principle of selection rigidly."

THE EFFECT OF THIS REPORT

As stated in the letter of transmission it was thought that the instructors would recognize the value of a uniform distribution

and that they would endorse and encourage the adoption of some system by their own willingness to approximate a suggested standard. This assumption is not justified by the distribution of grades for the night or day students for the second semester. While there seems to be a tendency to approximate nearer the normal distribution when the curve for the entire distribution for the second semester is plotted; yet the variation of distribution of the day grades with that of the night grades by the individual instructors is marked and seems to indicate that no conscientious effort was made to systematize the distribution. In fairness to all it might be said that during the second semester at the time when this report was circulated and for the remainder of the term the Dean was on sabbatical leave, so that the report received no official sanction other than it was sent out from the Dean's office at his request. Many of the faculty, however, discussed the report with the author and all were familiar with the psychological testing of the first year students, and many availed themselves at faculty meetings and elsewhere of learning more definitely the type of mentality of their poorer students as determined by these tests.

It will be observed from the following figures that the combined distribution for the second semester gives "H" plus "M" an equal value with "P" plus "F" with fifty per cent. as the central distribution when grades "D," "R," and "Fa," representing conditions over which the instructor has little or no control, are left out of account.

	H	M	C	P	F	D	R	Fa	H	M	C	P	F
1st Sem.	4	16	44	17	9	4	3	1	3	17	50	17	11
2d Sem.	6	18	47	14	9	1	2	3	7	18	50	15	10

There has been a gradual shifting toward the upper end of the curve as indicated by the following tabulation which represents the distribution of day grades for both semesters of each year:

	H	M	C	P	F	D	R	Fa
1912-13	3	15	46	15	10	3	4	4
1913-14	3	17	45	15	10	3	5	2
1914-15	4	18	47	14	9	1	2	3

And while the grades for the second semester 1914-15 tend to show progress in this direction it can hardly be ascribed to the results of this report. Plate IV presents the distribution of grades for the first and second semesters, day courses for the

HONOR

MERIT

CREDIT

PASSED FAILURE

years 1912-13, 1913-14, and 1914-15 and the total distribution for both semesters of the night courses for the years 1912-13 and 1913-14, but only the second semester for 1914-15. The grades for the second semester, night are given instead of the total distribution for the night courses for 1914-15 for the reason that it is desired to show a second attempt at grading after the report on the advisability of some uniform grading system was made known to these instructors. The distribution for the night grades is shown in black. Of the eighty-three instructors that assigned the 31,391 grades—the total number considered in this report—only twenty-six have been with the institution long enough to assign grades for each semester of the three years. Of these one taught only night courses, another only day courses. The marks of four others, whose grading shows no peculiarities compared with the others recorded in Plate IV, are omitted for the sake of brevity. In an institution where the faculty is constantly changing it is obvious that many difficulties arise in any attempt to enforce uniformity; but at the same time there is a greater need for uniformity in such an institution in order that the ever changing instructors may readily know the quality and the amount of work to be exacted. From a careful study of Plate IV, I am led to believe that there has been no conscious attempt made to assume a normal curve of distribution for the two final gradings after this report was made known. It is evident that the strong appeal made by the Committee of the George Washington University had little or no weight on the grading by the faculty of the Carnegie Institute of Technology. Personal opinion seems to be the only standard they were willing to use.

From the foregoing study the following points are summarized:

1. Grading for the vast majority of teachers is based on personal opinion.
2. Personal opinions are influenced more by temperamental attitude than by the nature and type of subject matter taught.
3. Elimination due to failure is unjust, especially when failure is determined by the teacher's personal opinion.
4. The granting of honors, etc., is debasing when the curve for graduation deviates so markedly from the distribution curve of all students as at Carnegie.

5. Night students ought to have a different standard from that of day students.
6. Shop instructors as a class are less exacting than academic instructors.
7. Instructors in shop subjects vary their grades as much as do academic instructors.
8. Distribution of ability as shown by grades assigned is not in conformity with ability as indicated by psychological tests.
9. Instructors are not willing to burden themselves with standardization of their grades so long as the institution is willing to accept haphazard personal judgments.
10. The distribution of grades by the entire faculty can hardly be considered a true curve for that institution since some instructors may distribute their grades so widely that they destroy the educational equilibrium of the school.
11. A curve of distribution can be based on the mental ability of the students as shown in some well devised psychological tests; or on a curve as established by a distribution of grades within that institution covering a number of years.
12. Such curves of distribution are fairer to the students and the institution than some arbitrary standard as suggested at Missouri, George Washington and other institutions which base their curve on the normal surface of frequency. If the entrance requirements have been selective of a group then the normal distribution curve is not an adequate distribution.
13. A curve based on psychological tests limits the standard of the institution to the type of general mental ability of the students. This may be very beneficial to the student but what would eventually be the standard of the institution?
14. A curve based on the former grades of an institution upholds the traditions of the institution but is not just in the long run to many of the students.
15. A curve which expresses the correlation of mental tests of the students within the institution with the distribution of grades as assigned by many instructors through many years ought to give a distribution which would maintain

the traditional standards of that institution and at the same time give the individual student full justice. Once this curve is established the entrance requirements can be measured by it in order to eliminate some of the wasteful teaching in the earlier years. "About one-third of all elimination is caused by the fact that the scholarship rating of the freshmen at the university is only one-third of the rating of the same student in the last year of high school, when based upon the actual amount of time spent in study. The serious break in a student's standing when he enters the university tends to discourage him and to drive him out of school. The university should raise its scholarship rating of freshmen students about 10 per cent., or one letter, according to the system of marking in use, and the high schools should lower that of their senior classes about the same amount. This would close the unnatural and harmful breach between them, and would reduce the large percentage of elimination."¹⁶

¹⁶ E. D. JENNINGS. *Elimination of Freshmen from the University of Texas*. JOURNAL OF EDUCATIONAL PSYCHOLOGY, April, 1915, p. 221.