Grade inflation: causes and cures[1]

R. R. Weil and W. Kroontje[2]

ABSTRACT

The grade point averages of college students have increased annually since the early 1960's, while scores on standard aptitude tests have dropped. Most educators believe the escalation of grades is not due to increased learning or better students. Hence, the phenomenon is often referred to as "grade inflation." The five causes of grade inflation discussed are 1) innovations like Pass/Fail options in grading systems, 2) a rise in anti-elitist philosophy, 3) increased use of mastery learning and competency-based evaluation without appropriate accommodations in grading systems, 4) student evaluation of instructors, and 5) competition for student enrollments both within and between institutions. The authors propose a package of four reforms to eliminate grade inflation and allow grades to effectively perform their positive functions of feedback, record-keeping, and motivation. The first reform calls for appropriate use of and clear separation of norm-reference and criterion-reference grading. Second, numerical scores would not be lumped into letter grade categories. Third, readily interpreted, standardized scores would be used as grades. Finally, transcripts would provide some information to describe the populations with which students are compared when standardized scores are used.

Additional index words: Standardized scores, Grades, Transcripts, Norm-referenced grading.

ACROSS the nation, student grade point averages have risen steadily since the early 1960's. Suspicions that this was occurring were confirmed by a study of 134 institutions of higher learning in 1973 (Juola, 1974). Since then concern over the phenomenon has become widespread, and the latest figures for the 1975-76 academic year show the first decline in over a decade (Fig. 1). It is too early to say for sure that this heralds a reversal or even a leveling off of the rise in grades, but there are indications that administrators and faculty at most universities are taking steps to "tighten up" their grading systems (Juola, 1976; Scully, 1975).

Looking at the phenomenon from a different angle we see that not only have grade point averages increased, but so has the proportion of high grades awarded to students (Fig. 2). Virginia Tech is a typical example. Most Land Grant universities experienced similar trends. In 1966, the majority of grades were C through F. By the 1974-75 academic year, the situation had undergone a dramatic

[1]Contribution from the Agronomy Dep., Virginia Polytechnic Institute, Blacksburg, VA 24061.

[2]Instructor and professor of agronomy, respectively.

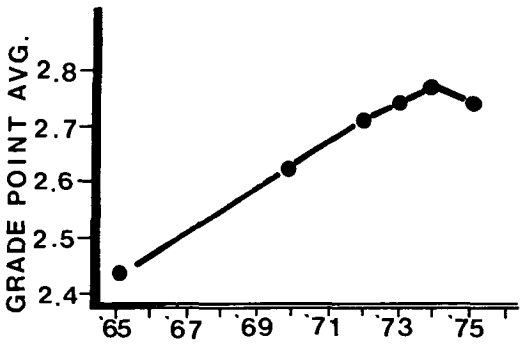

Fig. 1. Mean grade point average for 134 institutions, based on a 4.0 scale (from Juola, 1976).

reversal so that A's and B's accounted for 58% of the grades awarded (Harper, 1976). This is by no means an extreme example. At Amherst College that year fully 85% of the grades in the Spring term were A's and B's (Pressley, 1976). *Time* magazine (22), reported that A's alone accounted for 42% of the grades awarded to undergraduate at Yale University in the Spring of 1974.

Now all this would be a source of pride, rather than dismay, if we truely believed that the rising grades reflected a trend toward better students or greater learning. Unfortunately, educators do not put much stock in such an hypothesis (Moulds, 1974; Juola, 1974; Etzioni, 1975). In fact, indications are that, if anything, today's college student is less prepared than his predecessors. Under pressure to keep up enrollments and to serve a wider constituency, many colleges have allowed their entrance standards to slip and college students have become a less select group of people.

During the same years as grades were rising, scores on standard tests were falling (Magarrel, 1976). Scholastic Aptitude Test (SAT) scores, for example, showed declining aptitude for both math and verbal skills (Fig. 3). Similarly, American College Testing (ACT) scores have dropped about one-fifth of a standard deviation between 1964-65 and 1973-74 (Wilhelms, 1975). It was the feeling that the higher grades being given were not justified; hence the term "grade inflation."

If grades are a form of academic currency, grade inflation results in devaluation of that currency. Not only is the A cheaply earned and cheaply valued, but grade inflation has, by eroding confidence in the whole system of academic evaluation, devalued all grades and even the degrees to which they lead. Veterinary schools, graduate schools, employers, and the students themselves, have become increasingly skeptical about grades (Juola, 1976; Pressley, 1976).

Perhaps even more important is the concern that

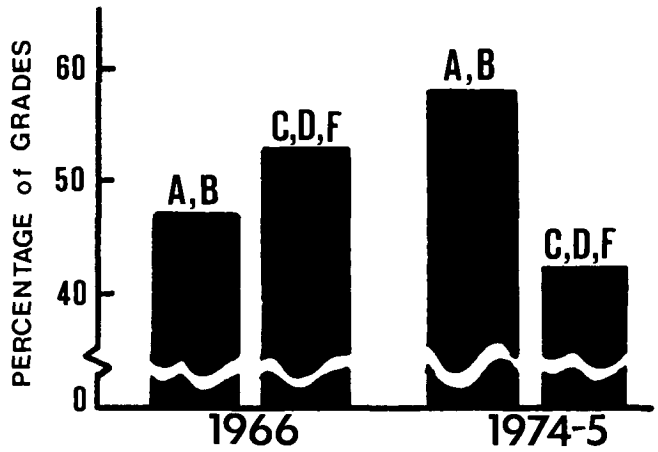

Fig. 2. High and low grades awarded at Virginia Polytechnic Institute and State University (source: Harper, 1976).

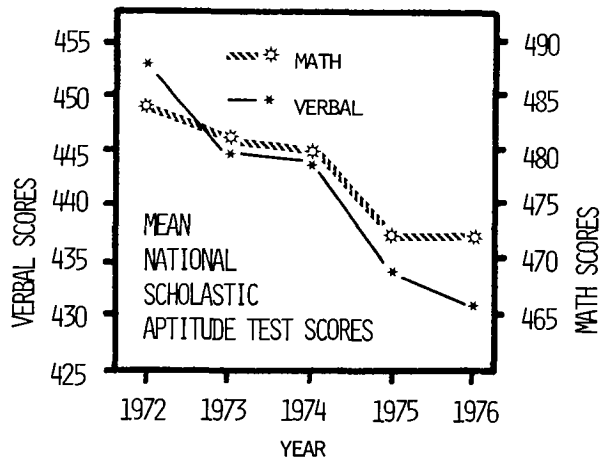

Fig. 3. National decline of Scholastic Aptitude Test scores. (source: Magarrel, 1976).

grade inflation is but a symptom of a sick grading system. Grades seem to have lost what little meaning they once had. They can hardly perform their positive functions of providing accurate feedback, predicting future performance, providing a basis for administrative decisions, recording achievement, and motivating both students and instructors (for a discussion of these functions, see Feldmesser, 1972). Under conditions of grade inflation, grades cannot be effective tools for communication. Does the A on a student's transcript signify outstanding work, average work, or the achievement of some minimum standard? It is around this broader concern with the health of the grading system that the remainder of this paper is centered.

SOME CAUSES

It is important to understand the major causes of grade inflation. For the most part, they fall into five categories (Butler, 1975; Collins and Nickel,

1975; Juola, 1974; Lippincot, 1973; Pressley, 1976; Walker, 1974):

1) innovations in grading systems
2) a rise in "anti-elitist" philosophy
3) mastery learning and performance-based evaluation
4) student evaluation of instructors
5) competition for student enrollments

A recent study of 544 institutions revealed that 77% of them had adopted some sort of innovation in their grading systems (Collins and Nickel, 1975). Nearly all of these changes tended to raise grade point averages. The most common innovation was the partial use of Pass/Fail or Credit/No Credit systems. The use of these grading schemes was usually at the option of the student, thus allowing him/her to insulate his/her grade point average from the effects of certain courses. Some schools stopped giving F's or ceased to calculate them into the grade point average. Some even used a system known as "ABC Disappear", by which no record is kept of grades below C.

A second factor behind the trend toward higher grades is the anti-elitist philosophy, which became fashionable during the 1960's (Hobbs, 1974; Muller, 1975; Milton and Edgerly, 1976). This egalitarian way of thinking abhors making distinctions between human beings and attempts to extend the idea of equality of opportunity to equality of achievement. It rejects the idea of striving for, and recognizing, excellence. Its emphasis is often on mass mediocrity. This attitude holds that to grade a person is to degrade him. There are also concerns that strict grading standards hold back already disadvantaged minorities. It is really an argument against grading itself. While it might make just as much sense for the anti-grading advocates to give everyone C's, the easiest, most generous response is to hand out A's and B's to all comers. Of course, handing out easy A's will no more resolve social inequalities than printing money would alleviate poverty.

The advent and widespread acceptance of "mastery" or competency-based learning is a third factor which has put upward pressure on grades. This approach to instructional development is probably best exemplified by the Keller Plan or Personalized System of Instruction (PSI). Evidence in the literature concerning the effectiveness of PSI is inconclusive (Kulik et al., 1974; Johnson, Gnagey, and Chesebro, 1974), but it would appear that this type of instruction is popular with students and may have real advantages for those with diverse backgrounds and for normally slow learners. Mastery learning operates on the assumption that just about anyone can master the course objectives if given enough chances and enough help. Theoretically, everyone should get an A. In practice, more than half of the class usually does.

The evaluation of instructors by students is another practice which became widespread during the period of rapid grade inflation. The temptation, conscious or not, is to be generous in grading students hoping that they will return the favor. And besides, there is the humanly natural desire to be liked and not to offend. Research on the subject, however, shows no certain relationship between the grade expected by the student and his opinion of the instructor (Walker, 1974; Frey, 1973; Remmers, 1966; and Weaver, 1960).

Last, but by no means least, grades have become tools in the increasingly fierce competition for enrollments (Butler, 1975). This competition has been both *between* and *within* institutions and has often provided students with educationally counterproductive incentives for enrolling in particular schools or courses. This is one of the most insidious aspects of grade inflation, and it is a vicious cycle. Why should a student risk a C or D in my elective course, when he is assured of a A or B by my colleague down the hall?

SOME CURES

Before discussing ways in which the grading system might be improved it should be pointed out that grading is only one of several processes in a typical educational model (Fig. 4). The grading process is usually incompatible with the teaching-learning process (Hobbs, 1974) and therefore, should be clearly delineated. Cardwell (1976) has likened the plight of a teacher whose students hide their ignorance for fear of grades to that of a physician whose patients are afraid to reveal their symptoms.

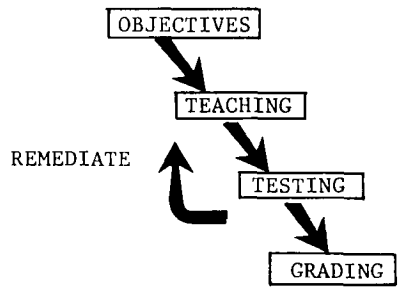

Fig. 4. A simple instructional model.

For the purpose of this discussion we shall make the assumption that grades should reflect objective evaluations of student achievement, not subjective impressions of student behavior unrelated to the stated objectives of the course. We will also assume that the setting of clear course objectives, the use of instructional techniques, and the tests used to measure achievement are all educationally valid and not at issue here. This paper focuses on the grading process. It deals with the evaluation and recording of raw data on student achievement once they are obtained.

How then can we cure grade inflation in particular and the current grading system in general? The moves being taken presently toward stricter grading may help to stem the rise of grade point averages but will do little to make grades more meaningful. What is required is a thorough, rational reform of the grading system. No instructor can change the meaning of grades unilaterally. The definitions must be clearly communicated both to the instructor, who summarizes student behavior in symbols, and to the user of grades, who interprets what behavior is symbolized. This paper presents four reforms which, taken together, would eliminate the possibility of grade inflation, restore integrity to the grading system, and vest grades with clearly interpretable meaning so that they can perform their useful functions.

First we must recognize that there are basically two types of grading systems, and that there is no rational way they can be mixed. These are the *norm-reference* and the *criterion-reference* systems. The norm-reference system, essentially competitive and democratic, usually compares students with their peers. The criterion reference system is essentially absolute and authoritarian, and compares students to a standard set by the instructor. Defining a C as average is normative. Defining it as, say, 70 to 80% correct is criterion-referenced. It cannot be stressed too strongly that these two systems are based on completely different assumptions, and that it is meaningless to average the resulting grades together. Yet nearly every college student's grade point average now includes both types of grades. Criterion-based evaluation should not be tolerated if the criteria are arbitrary and cannot be rationally defended.

Criterion-based evaluation is best suited to mastery learning, but even there the system should only be bivariate. One either masters the material or one does not. Transcripts should show a distinctive grade for Mastery courses—capital "M" for Mastery would be a logical symbol. Perhaps the difficulty and number of the objectives mastered could be reflected by awarding more or fewer credit hours. Thus, a student might be awarded anywhere from 1 to 5 hours of credit in a crop production course, depending on the number and sophistication of the objectives he had mastered.

For those courses not structured for mastery learning, normative grades should be used. It should be made clear, however, that the use of norm-reference grading in no way relieves the instructor of the responsibility to insure that the achievement being measured relates directly to clearly stated behavioral objectives. What normative grading does do is rescue the student from being at the mercy of the instructor's idiosyncrasies.

The second reform proposed is to stop trying to convert scores from tests and assignments into letter grades. Are we so confident of the difference between an 89% average and a 90% average that we can give one student a B and the other an A for the course? Many universities have computerized test scoring services which provide information on an exam's standard error of measurement. We try to use this data in making grading decisions, but it is frustrating to realize that the 90% confidence interval for a student's score is usually at least 5 to 10%, and we are asked to draw a line between students whose scores are only 1 or 2% apart.

Even our best tests and assignments provide only imprecise measurements of student achievement. Yet we subject these measurements to further imprecisions. Typically we begin with a test graded on a 100 point scale from 0 to 100%. We then lump these scores into five classes: A through F. Finally we proceed to multiply and divide by credit hours and carry the resulting grade point average out to two or three decimal places—a 400 to 4,000 point scale. How many of us would think of treating our research data in this manner?!

But perhaps the most telling argument against converting to letter grades is that the decision of where to draw the line is usually left to the whim of the instructor. One instructor may decide that 90% should be the cutoff for an A. Another may choose 95% or 85%. If a normative system is used, one might decide the top 10% should get A's. While another might decide the top 20 to 30% should get A's. The temptation to be an easy grader and the invitation for grade inflation are obvious. So is the absurdity of attempting to average together grades awarded according to different standards.

One institution that has taken steps in this direction is Vanier College in Quebec, Canada. Tran-

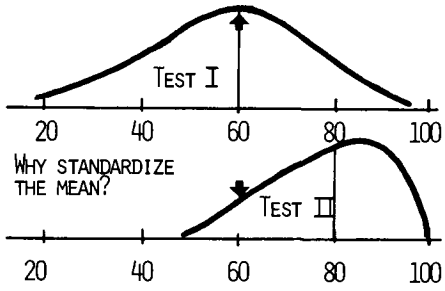

Fig. 5. Hypothetical distributions of student test scores for two tests differing in difficulty. The means are indicated by vertical lines.

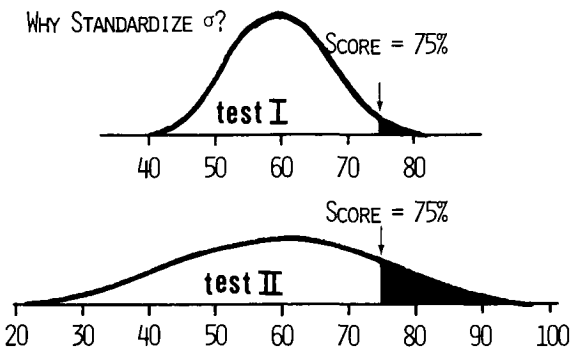

Fig. 6. Hypothetical distributions of student scores for two tests with a common class mean but differing standard deviations. The shaded area represents the number of students exceeding 75% correct.

scripts from Vanier show student performance in each course as a percent score. An adjacent column on the transcript lists the mean score for the class, information which greatly enhances the meaning of scores as norm-reference grades.

A third reform of the grading system would be a refinement of this procedure: the use of standardized scores to grade all norm-reference (not mastery) courses. A standardized score transforms all distributions of raw scores into distributions with a common standard deviation (SD) and a common mean. The T-score is a popular standardized score which is based on a standard deviation of 10 (σ_T) and a mean of 50 ($\bar{X}_T$) (Angoff, 1971). Thus, a student achieving a score 1.5 SD above the class mean would have a T-score of 65. A T-score (X_T) can be readily calculated for a raw score (X_o) such as the number correct on a test. To do so, one must calculate the mean ($\bar{X}_o$) and the standard deviation (σ_o) of the raw scores and apply equation [1].

$$X_T = (\sigma_T/\sigma_o)\,(X_o - \bar{X}_o) + 50 \qquad [1]$$

Table 1. A portion of a hypothetical transcript showing the use of standardized T-scores for norm-referenced grading

Course	Mastery	Credits	T-score	T-score pts.
Intro. Soils		4	70	280
Crop Mgt.		3	60	180
Gen. Chem.	M	3	--	--
Farm Planning		3	62	186
Calculus (honors)		4	60	200
Syst. Analysis		3	71	213
	Total credits	20		
	normative	17	T-score point average:	62.29
	mastery	3		

Table 2. The same transcript as in Table 1 modified to show the use of the mean S. A. T. scores for each class to aid in interpretating the T-scores

Course	Mastery	Credits	T-score	T-score pts.	Class avg. SAT score
Intro. Soils		4	70	280	1100
Crop Mgt.		3	60	180	1250
Gen. Chem.	M	3	--	--	--
Farm Planning		3	62	186	1025
Calculus (honors)		4	50	200	1400
Syst. Analysis		3	71	213	1200
	Total credits	20			
	normative	17	T-score point average:		62.29
	mastery	3			

Why standardize the mean? The example in Fig. 5 shows that achieving 60% correct on one test may be average, while 60% correct on another test may be poor. T-scores thus allow one to average together tests (or courses) of varying difficulty.

Why standardize sigma? Figure 6 illustrates that even when the class average is the same for two tests, the relative standing of a student is dependent on the spread of the scores. A score of 75% correct is a more outstanding achievement on test I than on test II.

If we combined these three reforms, a student transcript might look like Table 1. Only norm-reference grades are entered into the grade point calculations. Credits in mastery learning courses are recorded separately. There is no lumping of scores into letter grades, since standardized T-score are recorded on the transcript. No longer is there any question as to what each grade means. There are no "easy A's" or "gentleman's B's". There is no possibility of grade inflation because there are no arbitrary grades. We can see that this student was average in his honors calculus class, and that he was one of the best students (2 SD's above the mean) in introductory soils. Credit hours could still be used to arrive at T-score points and ultimately at T-score point average analogous to the old grade point average.

One problem still remains, that of defining the populations with which the student is being compared. In a large class—50 or more students—this

should present no great problem since it would very likely be quite representative of the university as a whole. However, for smaller classes and special cases, it would be useful to provide some information describing the population from which the T-score was derived.

Using today's computerized registration systems it would not be too difficult to use a variety of information about students in each class to transform scores standardized with respect to the class into scores standardized with respect to the entire student body of an institution (I. J. Goode, 1976, personal communication). However, it might be difficult to explain to students and other users of grades just what had been done and how the new grades should be interpreted.

A simpler possibility would be to list the mean SAT score for each of the classes. Even the mean grade point average of the class would be helpful. Thus in our example we can see that the calculus class was made up of very gifted students, with a mean combined SAT of 1400. Being average in that class was a respectable achievement. By contrast, the T-score of 70 in soils wasn't terrific. One can imagine that this type of grading system would eliminate the bad incentives for students to take courses which may be irrelevant to their educational goals, but in which they anticipate an easy grade. Not only is it impossible for one instructor to be a much easier grader than another, but there is little advantage for a student to seek out relatively dull classes in which (s)he can readily excel.

IN CONCLUSION

A package of four basic reforms in the grading systems would aleviate grade inflation and other associated problems caused by the present lack of a clearly defined grading policy. We hope that this paper will stimulate discussion and, ultimately, reformulation of grading policies. If grades are to serve useful functions, grading policies should 1) recognize the basic incompatibility between norm and criterion referenced systems, 2) do away with the necessity for drawing arbitrary lines that lump students into letter grade categories, 3) use a standardized scoring system that is readily interpretable; and 4) provide some information to describe the populations with which students are compared in norm-reference graded courses.

Changes such as these will not come easily, but if the result will be a grading system we can live with, the effort will have been worthwhile.

LITERATURE CITED

1. Angoff, W. H. 1971. Scales, norms, and equivalent scores. *In* R. L. Thorndike (ed.) Educational measurements. 2nd Ed. American Council on Education. Washington, D. C.
2. Butler, R. C. 1975. The grading game. Coll. Univ. 50(40)733-739.
3. Cardwell, C. E. 1976. Some reflections on grade inflation. Context 10(4)8-11.
4. Collins, J. R., and K. N. Nickel. 1975. Grading policies in higher education: the Kansas Study/the national survey. University Studies No. 103, Wichita, Kan.
5. Etzioni, Amitai. 1975. Grade inflation: neither freedom nor discipline. Hum. Behav. 4(10):11-12.
6. Feldmesser, R. A. 1972. The positive functions of grades. Educ. Rec. 53:66-72.
7. Frey, P. W. 1973. Student rating of teaching: validity of several rating factors. Science 182:83-85.
8. Harper, L. J. 1976. Grade escalation at the college and university level: it is a fact but what are the factors. Presented to the Southern Association of Deans of Agriculture, Starkville, Mississippi, March 30, 1976.
9. Hobbs, John. 1974. Grading versus teaching. Improv. Coll. Univ. Teach. 22(4):239-243.
10. Johnson, J. U., W. J. Gnagey, and P. A. Chesebro. 1974. Does "mastery" really improve student achievement? Improv. Coll. Univ. Teach. 22(1):59-60.
11. Juola, A. E. 1974. Grade inflation (1960-1973) A preliminary report. Mimeo. Office of Evaluation Services. Michigan State Univ., E. Lansing, Mich.
12. ———. 1976. Grade inflation—1975: is it over? Mimeo. Office of Learning and Evaluation Services. Michigan State Univ., E. Lansing, Mich.
13. Kulik, J. A., Chen-Lin Kulik, and Devin Carmichael. 1974. The Keller Plan in science teaching. Science 183:379-83.
14. Lippincot, W. T. 1973. On abandoning grading and reconsidering standards. J. Chem. Educ. 50(7)449.
15. Magarrel, Jack. 1976. Aptitude-test scores: is the decline slowing? Chronic. Higher Educ. 13(2):3.
16. Milton, Ohmer, and J. W. Edgerly. 1976. The testing and grading of students. Change Magazine and Educational Service. New Rochelle, N. Y. 62 p.
17. Moulds, Henry. 1974. To grade or not to grade: a futile question. Intellect 102:501-4.
18. Muller, Steven. 1975. American education standards are slipping. Today's Educ. 64(4):50-52.
19. Pressley, Milton. 1976. Inflation hits the campus. Wall Street J., 21 Jan. 1976. p. 16.
20. Remmers, H. H. 1966. The college professor as the student sees him. Purdue Stud. Higher Educ. 93:14-18.
21. Scully, Malcolm. 1975. Inflated grades worry more and more colleges. Chronic. Higher Educ. 13(37):1.
22. Too many A's: college grades. Time 104:106-7. 11 Nov. 1974.
23. Walker, B. D. 1974. Anticipated grade and student evaluation of teacher. Improv. Coll. Univ. Teach. 22(4):220-221.
24. Weaver, C. H. 1960. Instructor rating by college students. J. Educ. Psychol. 51:21-25.
25. Wilhelms, F. T. 1975. What about basic standards? Today's Educ. 64(4):46-48.