



Computerized Gradebooks and the Myth of Objectivity

Author(s): Thomas R. Guskey

Source: *The Phi Delta Kappan*, Vol. 83, No. 10 (Jun., 2002), pp. 775-780

Published by: [Phi Delta Kappa International](#)

Stable URL: <http://www.jstor.org/stable/20440252>

Accessed: 17/05/2011 13:25

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=pdki>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Phi Delta Kappa International is collaborating with JSTOR to digitize, preserve and extend access to *The Phi Delta Kappan*.

<http://www.jstor.org>

Computerized Gradebooks And the Myth Of Objectivity



Computerized grading programs and electronic gradebooks can be useful tools. But in the end, Mr. Guskey reminds us, teachers must still decide what grade offers the most accurate and fairest description of each student's achievement and level of performance.

BY THOMAS R. GUSKEY

IF YOU ASK middle school or high school teachers today how they determine their students' grades, the first thing most of them will do is open a computerized grading program. They'll show you the vast array of data they keep on each student and explain how they weigh the different pieces of information. At the end of the marking period, they combine these various measures and, with the help of the computer, calculate a summary score to the one-hundred-thousandth of a decimal point. The computer then converts this summary score into the letter grade that is printed on a report card and sent home to parents. Many teachers will also go on to describe the fairness and objectivity of this process, pointing out how the mathematical precision of the computer makes it easy for them to explain and to defend their grading policies to students, to parents, and to administrators.

But do computerized gradebooks really make grading fairer and more objective? Or have the technical capabilities of these programs seduced teachers and school leaders into a false sense of confidence in the

accuracy and validity of the grades they assign?

COMPUTERIZED GRADEBOOKS

Computerized grading programs and electronic gradebooks rank among the best-selling computer software available to educators today. They appeal to teachers primarily because they simplify record-keeping. The spreadsheet formats and database management systems

*THOMAS R. GUSKEY is a professor in the College of Education, University of Kentucky, Lexington. This article is based on material drawn from *Developing Grading and Reporting Systems for Student Learning*, by Thomas R. Guskey and Jane M. Bailey (Corwin Press, 2001). © 2002, Thomas R. Guskey.*

TABLE 1.
Summary Grades Tallied by Three Different Methods

Student	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Average Score	Grade	Median Score	Grade	Deleting Lowest	Grade
1	59	69	79	89	99	79	C	79	C	84	B
2	99	89	79	69	59	79	C	79	C	84	B
3	77	80	80	78	80	79	C	80	B	79.5	C
4	49	49	98	99	100	79	C	98	A	86.5	B
5	100	99	98	49	49	79	C	98	A	86.5	B
6	0	98	98	99	100	79	C	98	A	98.8	A
7	100	99	98	98	0	79	C	98	A	98.8	A

Grading Scale: 90%-100%=A, 80%-89%=B, 70%-79%=C, 60%-69%=D, 59% or lower=F.

included in these programs make it easy for teachers to enter and tally precisely large amounts of numerical information.¹ Thus they are suited particularly well to the point-based grading systems of middle school and high school teachers, who often record numerical data on the performance of more than 100 students each week.

Most computerized grading programs also present educators with a wide range of options. Some simply help teachers to keep more detailed records on students' learning progress.² Others allow teachers to present summaries of their students' achievement and performance in a variety of different formats, including computer displays, online reports, and even digital portfolios. Still other programs actually perform grading tasks. The simplest of these scan, mark, and analyze assessments composed of true/false, matching, and multiple-choice items. More recently, however, exciting advances have been made in the use of computers to evaluate and grade students' essays, compositions, and other writing samples.³

For all their advantages, however, computerized grading programs also have their shortcomings. Perhaps the most serious is that they lead the educators who use them to believe that mathematical precision necessarily brings greater objectivity and enhanced fairness to grading. Many teachers assume that, so long as the mathematical calculations are correct and all students are treated the same, then the grades assigned are accurate and just. But numerical precision is not the same as evaluative fairness, honesty, or truth. While computerized grading programs and electronic gradebooks may greatly simplify record-keeping, they do not lessen the challenge involved in assigning grades that accurately and honestly reflect students' level of performance.

MATHEMATICAL PRECISION VERSUS VALID GRADES

Consider, for example, the data in Table 1. The scores on the left side of the table reflect the performance of seven students over five instructional units. The scores on the right represent summary scores for these students calculated by three different methods. The first method is the simple arithmetic average of the unit scores, with all units receiving equal weight. The second is the median or middle score from the five units.⁴ Because the median is positional rather than proportional, it's not influenced by extreme scores, as is an average. The third method is also an arithmetic average, but with the lowest unit score in the group deleted. This method is based on the assumption that no one, including students, performs at a peak level all the time.⁵ These are the three tallying methods most frequently used by teachers and most commonly employed in computerized grading programs and electronic gradebooks.

Consider, too, the following explanations for these score patterns:

- Student 1 struggled in the early part of the marking period but continued to work hard, improved in each unit, and performed excellently in unit 5.
- Student 2 began with excellent performance in unit 1 but then lost motivation, declined steadily during the marking period, and received a failing mark for unit 5.
- Student 3 performed steadily throughout the marking period, receiving three B's and two C's, both near the cutoff between B and C.
- Student 4 began the marking period poorly and failed the first two units but, with newfound interest, performed excellently in units 3, 4, and 5.

- Student 5 began the marking period excellently but then lost interest and failed the last two units.

- Student 6 skipped school (an unexcused absence) during the first unit but performed excellently in every other unit.

- Student 7 performed excellently in the first four units but was caught cheating on the assessment for unit 5 and received a score of zero for that unit.

As is evident from Table 1, all three of these tallying methods are mathematically precise. Yet each one yields a very different pattern of grades for these seven students. If you use the simple arithmetic average, all seven students would receive the same grade of C. If you use the median, there would be just two C's, one B, and four A's. And if you use an arithmetic average with the lowest score deleted, there would be just one C, four B's, and two A's. Note, too, that the one student who would receive a grade of C using this third method had unit grades of just two C's and three B's. More important, not one student would receive the same grade across all three methods. In fact, two students (Student 4 and Student 5) could receive a grade of A, B, or C, depending on the tallying method you use.

The teacher responsible for assigning grades to the performance of these seven students has to answer a number of difficult questions. For example, which of these three methods is fairest? Which method provides the most accurate summary of each student's achievement and level of performance? Do all seven students deserve the same grade, as using the arithmetic average suggests, or are there defensible reasons to justify different grades for certain students? And if there are reasons to justify different grades, can these reasons be clearly specified? Can they be fairly and equitably applied to the performance of all students? Can these reasons be clearly communicated to students before instruction begins? Would it be fair to apply them if they were not communicated to students?

The nature of the assessment information from which these scores are derived could make matters even more tangled. It might make a difference, for example, if the content of each unit assessment was cumulative. In other words, the assessment for unit 2 contained material from units 1 and 2, and the unit 5 assessment included material from all five previous units. And if it did, would this make these grading decisions any easier, or would it further complicate summary calculations?

What should be evident in this example is that the use of computerized grading programs won't solve these

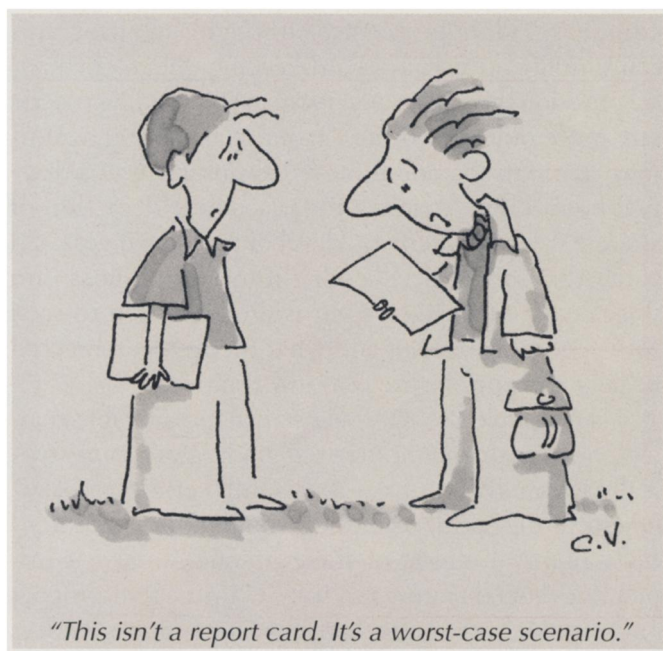
complex grading problems. Although such programs can simplify numerical record-keeping, the mathematical precision they offer does not make the grading process any more objective or any fairer. Calculating a summary score to the one-hundred-thousandth of a decimal point doesn't yield a more accurate depiction of students' achievement and level of performance. Each teacher still must decide what information goes into the calculation, what weight will be attached to each source of information, and what method will be used to tally and summarize that information.

This example also illustrates several questionable grading practices that computerized grading programs typically ignore. Although not new and certainly not inherent in the use of technology in grading, the potentially harmful effects of these practices make it imperative that educators carefully examine their impact and consider other alternatives. Three such practices include 1) averaging scores to determine a grade, 2) the use of zeroes, and 3) taking credit away from students or lowering their grade because of behavioral infractions.

AVERAGING SCORES TO DETERMINE A GRADE

If a mark or grade is supposed to represent an accurate description of how well students have learned, as most experts on grading agree it should,⁶ then the practice of averaging generally falls far short. For example, how often have you heard students lament, "I have to get an A on the final exam in order to pass this course"? But does this situation really make sense, or does it illustrate the inappropriateness of averaging? If a final examination or summative performance truly represents a comprehensive assessment of what students have learned, how can an A level of performance there translate to a C or D for the course grade? Similarly, if a final grade is to reflect what students have learned and can do at the end of the course, can averaging scores from past assessments with measures of current performance be considered appropriate?

Educators generally recognize learning as a progressive and incremental process. Most also agree that students should have multiple opportunities to demonstrate their learning. But is it fair to consider all these learning trials in determining students' grades? If at any time in the instructional process students demonstrate that they've learned the concepts well and mastered the intended learning goals, doesn't that make all previous information on their learning of those con-



cepts inaccurate and invalid? Why then should such information be “averaged in” when determining students’ grades?

Because any single measure of learning can be unreliable, most researchers recommend using several indicators to determine students’ marks or grades.⁷ Nevertheless, teachers must continually ask themselves, “What information provides the most accurate depiction of students’ learning at this time?” In nearly all cases, the answer is “the most current information.” If students demonstrate that past assessment results no longer accurately reflect their learning, that information must be discarded and replaced by the new information. Continuing to rely on past assessment data miscommunicates students’ achievement. Can you imagine, for example, the karate teacher suggesting that a student who starts with a white belt but then progresses to earn a black belt actually deserves a gray belt?

Averaging can also have detrimental effects on student motivation. Suppose, for example, that a student does poorly on one or two major assessments administered early in the marking period, as was the case with Student 4 and Student 6 in Table 1. Knowing that those scores will be “averaged in” as part of the final grade, what motivation do these students have to do well on other assessments? Even if they perform at the highest level from that time on, the practice of averaging gives them virtually no chance of attaining a high grade.

And consider this extreme but true occurrence. A high school student I know experienced the death of

a beloved family member during the first marking period of his senior year. The trauma of that experience proved exceptionally difficult for this young man. As a result, he neglected his schoolwork completely and received failing grades in all his courses. But then, with help from counselors, family and community members, and his teachers, he recovered emotionally, re-dedicated himself to his schooling, and with diligent effort attained A’s in all his courses during the remaining three marking periods of the school year. Because of his school’s policy of averaging, however, his final course grades were all C’s. Did those C’s accurately reflect what he had learned? Did they represent what he had accomplished? Did they adequately describe his achievement or level of performance? Was this fair?

Recognizing that single measures of student learning can be flawed or unreliable, most teachers use multiple sources of information when assigning marks or grades. But simply combining all such measures and calculating an average is rarely appropriate or fair. Some educators argue that the median or middle score provides a more appropriate measure,⁸ but that practice, too, can be problematic.

To provide an accurate summary of students’ performance, teachers must begin by looking for consistency in the evidence gathered. If that evidence is consistent across several indicators, then deciding what grade to assign is relatively straightforward. This would be the case, for example, for students who obtained very similar scores on a class project, on two summative examinations, and on an oral report. But even these cases get complicated when scores consistently fall near the cutoff between two grades. Note, for example, the scores of Student 3 in Table 1.

If the evidence of student achievement is inconsistent, then teachers must look deeper and search for the reasons why.⁹ They also have to face the difficult challenge of deciding what evidence or combination of evidence represents the truest and most appropriate summary of students’ achievement and performance. In such cases, three general guidelines can be recommended.¹⁰

First, the most recent evidence should always be given priority or greater weight. Because grades are usually meant to represent students’ current achievement status or level of performance, the most accurate evidence is generally the evidence collected most recently. Therefore, scores from assessments at the end of the marking period are typically more representa-

tive of what students have learned than those collected at the beginning.

A second strategy is to give priority or greater weight to the most comprehensive forms of evidence. If certain sources of evidence represent cumulative summaries of the knowledge and skills students have acquired, then these should hold the greatest weight in determining students' grades. Exceptions to this approach might be necessary, however, for students who suffer inordinate test or performance anxiety. Such students typically do remarkably well on assignments, quizzes, and class discussions, but then "freeze" during larger assessments or performances. In these cases, teachers may have to consider other means of gathering evidence, such as orally questioning those students or providing some other means for them to demonstrate their learning, in order to get a more valid representation of what they can do.

A third approach would be to "rank order" the evidence gathered in terms of its importance to the learning goals or standards of the course. Those sources of evidence that relate to the most important goals or standards should then be given priority. For example, teachers might attach greater importance to students' scores on a project that required them to synthesize and apply what they had learned than they might give to the scores students attained on assessments designed to tap basic knowledge and comprehension of course content.

Whatever strategy teachers choose, they must be sure to apply that strategy consistently. Although exceptions to accommodate unusual or extenuating circumstances are always permissible, fairness in grading dictates that teachers inform students about their grading policies and practices in advance and then faithfully and consistently apply those policies.

THE USE OF ZEROES

Few teachers believe that grades should be used to punish students for their lack of effort or for demonstrating inadequate responsibility. At the same time, however, many teachers assign zeroes to student work that is missed, neglected, or turned in late.¹¹ Obviously, if grades are to represent how well students have learned, then the practice of assigning zeroes for "administrative or behavioral" reasons clearly misses the mark.

Zeroes have an even more profound effect if combined with the practice of averaging. Students who receive a single zero have little chance of success because

such an extreme score so drastically skews the average. (Note, for example, the scores of Student 6 and Student 7 in Table 1.) For this reason, in scoring Olympic events like gymnastics and diving, the highest and lowest judges' scores are always eliminated before the averaging takes place. If they were not, a single judge could control the results of an entire competition simply by giving extreme scores.

Some teachers defend the practice of assigning zeroes by arguing that they cannot give students credit for work that is incomplete or not turned in — and that's certainly true. But there are far better ways to motivate and encourage students to complete assignments than by assigning them zeroes, especially considering the overwhelmingly negative effects.

One alternative approach is to assign an "incomplete" and then require students to do additional work to bring their performance up to an acceptable level. Students who miss an assignment or neglect a project deadline, for example, might be required to attend after-school study sessions or special Saturday school programs in order to complete their work. In other words, these students are not "let off the hook" with a zero. Instead, students learn that they have responsibilities in school and that their actions have specific consequences. In addition, it helps to make the grade a more accurate reflection of what the students have actually learned.

LOWERING GRADES BECAUSE OF BEHAVIOR

Another typical grading practice with detrimental effects is lowering students' grades because of behavioral infractions. Some teachers lower students' grades for classroom disruptions and similar forms of misconduct. Other teachers consider tardiness or class attendance in determining students' grades and often reduce the grades of students who are late or who miss class sessions. Teachers also vary widely in how they handle such offenses as plagiarism, copying another student's work, and other forms of "cheating." But most teachers weigh such transgressions heavily when determining students' grades.

Student 6 and Student 7 in Table 1 offer excellent examples. Although Student 6 performed exceptionally well throughout most of the marking period, a zero due to an unexcused absence could severely affect his or her course grade. Student 7 performed excellently in four units but was then caught cheating on the assessment for unit 5 and received a zero. Most teachers

would undoubtedly consider this a fair response to Student 7's infraction. But when it comes to determining this student's course grade, the issues become thornier. Some teachers would look at the achievement history over the marking period, conclude that this incident was an exception, and assign the student a high grade. Others would reason that the high marks in earlier units could well have been attained through cheating as well, although the student didn't get caught. Hence, they would feel justified in assigning a lower grade.

The essential question the teacher must address in each of these cases is, "What is the purpose of grading?" If the purpose of grading is to present a summary judgment of students' achievement and level of performance, then to count these behavioral infractions in determining the grade clearly miscommunicates. Although such infractions cannot be ignored, it's clear that they are not part of the evidence that shows what these students have learned and are able to do.

A better strategy is to report these behavioral infractions separately and not include them as part of the course grade. For example, in a growing number of schools, reporting forms are designed to include indicators of students' class behaviors and work habits in addition to grades representing their achievement and level of performance.¹² In other words, teachers report "multiple grades" in each course, separating evidence of students' learning from information about their behavior and conduct.

Some educators might feel that reporting multiple grades makes both record-keeping and grading procedures overly complicated. But those who use this approach report that it actually simplifies grading. They collect no additional information from students and have eliminated the final step of having to combine these diverse sources of evidence. By separating the different aspects of students' performance in school, these teachers provide more specific information to parents and to students. In addition, they are able to identify more clearly students' strengths as well as areas in which improvement is needed.

Computerized grading programs and electronic gradebooks greatly simplify the record-keeping tasks teachers face. They allow teachers to collect and efficiently summarize large amounts of data on student learning. But the efficiency and mathematical precision of these programs does not make the grades they generate more accurate, honest, fair, or objective.

Grading requires careful planning, thoughtful judgment, a clear focus on purpose, excellent communication skills, and an overriding concern for the well-being of students — qualities that no computer possesses. Teachers at all levels must make carefully reasoned decisions about which components will be included in determining students' grades, how those components will be combined and summarized, and what format will be used to report the summaries. While computerized grading programs and electronic gradebooks can be useful tools, they do not relieve teachers of the professional responsibilities involved in making these crucial decisions. In the end, teachers must still decide what grade offers the most accurate and fairest description of each student's achievement and level of performance.

1. Joe Huber, "Gradebook Programs: Which Ones Make the Grade?," *Technology Connection*, vol. 4, no. 1, 1997, pp. 21-23; and Edward L. Vockell and Douglas J. Fiore, "Electronic Gradebooks: What Current Programs Can Do for Teachers," *Clearing House*, vol. 66, 1993, pp. 141-45.
2. Kenneth W. Eastwood, "Reporting Student Progress: One District's Attempt with Student Literacy," in Thomas R. Guskey, ed., *Communicating Student Learning: 1996 Yearbook of the Association for Supervision and Curriculum Development* (Alexandria, Va.: Association for Supervision and Curriculum Development, 1996), pp. 65-78.
3. Ellis B. Page and Nancy S. Petersen, "The Computer Moves into Essay Grading: Updating the Ancient Test," *Phi Delta Kappan*, March 1995, pp. 561-65; and William Wresch, "The Imminence of Grading Essays by Computer — 25 Years Later," *Computers and Composition*, vol. 10, no. 2, 1993, pp. 45-58.
4. Russell G. Wright, "Success for All: The Median Is the Key," *Phi Delta Kappan*, May 1994, pp. 723-25.
5. Robert Lynn Canady and Phyllis Riley Hotchkiss, "It's a Good Score! Just a Bad Grade," *Phi Delta Kappan*, September 1989, pp. 68-71.
6. Thomas M. Haladyna, *A Complete Guide to Student Grading* (Boston: Allyn & Bacon, 1999); Robert J. Marzano, *Transforming Classroom Grading* (Alexandria, Va.: Association for Supervision and Curriculum Development, 2000); Ken O'Connor, *How to Grade for Learning* (Arlington Heights, Ill.: Skylight, 1999); and Richard J. Stiggins, *Student-Involvement Classroom Assessment*, 3rd ed. (Upper Saddle River, N.J.: Merrill/Prentice-Hall, 2001), pp. 409-65.
7. Gary Natriello, "The Impact of Evaluation Processes on Students," *Educational Psychologist*, vol. 22, 1987, pp. 155-75.
8. Wright, op. cit.
9. Susan M. Brookhart, "Teaching About Communicating Assessment Results and Grading," *Educational Measurement: Issues and Practice*, vol. 18, 1999, pp. 5-13.
10. Thomas R. Guskey and Jane M. Bailey, *Developing Grading and Reporting Systems for Student Learning* (Thousand Oaks, Calif.: Corwin Press, 2001).
11. Canady and Hotchkiss, op. cit.; Richard J. Stiggins and Daniel L. Duke, "District Grading Policies and Their Potential Impact on At-Risk Students," paper presented at the annual meeting of the American Educational Research Association, Chicago, 1991.
12. Cathy Costello and Barry McKillop, "Dealing with Lates and Absences," *Orbit*, vol. 30, no. 4, 2000, pp. 43-46.

