



Bringing Back Teachers' Professional Judgment in the Era of Computerized Grading Systems

Written by Ken O'Connor

Edited by Sally I'Anson

November 2016

Traditionally, teachers have used the mean of a student's scores because determining grades was simply a mechanical, numerical exercise in which they recorded scores for everything that students did and calculated grades as the average of those scores. Computer grading programs, introduced in the mid-1980s, caused teachers to rejoice because these programs made their job easier, as the software now did the calculations and released them from the tedious task of using a calculator to determine the mean. Teachers also appreciated those programs because they were able to justify their grades to students and parents by saying that the computer had done the calculation and it was always right. With considerable regret, I admit that is what I did when I was a classroom teacher!

Of course, it wasn't as simple as that, because calculating the mean frequently resulted in inaccurate grades. Way back when, probably in fifth grade, I learned that there are three measures of central tendency—mean, median and mode. Each measure has advantages and disadvantages and is appropriately used with some distributions and not with others. What I learned then about a mean score is it's susceptible to being skewed by outlier scores. When outlier scores are present, one of the other measures of central tendency should be used. Students frequently have outlier scores, and more often than not those outlier scores are low. The calculation of the mean inappropriately penalizes students with low outlier scores.

One of the dirty little secrets about schools of education is that pre-service teachers receive minimal training on assessment and almost none on grading. As a result, teachers lack confidence in their ability to determine grades and have graded the way their teachers determined grades or the way their mentors trained them to grade early in their career. The result of this is that teachers have most commonly hidden behind the calculator and computer grading programs and so grading was a mechanical, numerical exercise in the calculation of the mean. Many teachers were

dissatisfied with being “mean” teachers, but they didn't know how to do grading differently, and so they just went along.

An experience I had in 2004 illustrates this. In late April of that year, I was asked to do a presentation at a very traditional high school in a wealthy enclave near New York City. The principal wasn't expecting teachers to make any immediate changes as a result of the presentation, but she wanted them to be aware that there were ideas about different ways to determine grades. What caused the most heated discussion was my suggestion that teachers consider the median (the score that falls in the middle of a number line of all the scores) and the mode (the most frequently occurring score) in addition to, or as alternatives to, the mean (the average).

One of the dirty little secrets about schools of education is that pre-service teachers receive minimal training on assessment and almost none on grading. As a result, teachers lack confidence in their ability to determine grades.

These grading ideas were new to them and caused great consternation. But, one teacher communicated with me by email in late June. She told me that she calculated the mean from all of the scores for each of her 105 students and then looked holistically at the scores for all her students to see if the mean score gave an accurate representation of their achievement. Most of her students achieved consistently at a high level, so, with a relatively quick look, she was able to determine that the mean was appropriate for 100 of her students. However, for five students it wasn't, and for those five students she used the median or the mode. This change was small but significant, and the last few words of her email were fascinating because she said, “I feel liberated”; she was grading using her professional

judgment and felt released from the constraints of established practice.

As the use of computer grading programs became more common throughout the 1990s and early 2000s, many teachers began to understand that there were serious shortcomings in traditional grading and in the computer grading programs they were using. Tom Guskey, a professor at the University of Kentucky, and a leading expert on grading, is particularly critical of computer grading programs' reliance on the use of percentages. In his article "The Case Against Percentage Grades," published in *Educational Leadership* in September 2013, he states:

Percentage grades continued to be relatively rare in U.S. schools until the early 1990s when grading software and online grade books began to gain popularity among educators. Today, schools can choose from more than 50 electronic grading software programs. Because these programs are developed primarily by computer technicians and software engineers rather than educators, they incorporate scales that appeal to technicians—specifically, percentages. (69)

In another Guskey article that appeared in the June 2002 issue of *Kappan* titled "Computerized Gradebooks and the Myth of Objectivity," Guskey wrote:

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 as to which components they will include in determining students' grades, how those components will be combined and summarized, and the format that will be used to report the summaries. While automated grading programs and electronic grade books 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. (780)

At the time, Guskey's suggestion that the use of computer programs didn't "relieve teachers of the professional responsibilities involved in making these critical decisions (about grades)" was not well received. Teachers still viewed grading as a numerical, mechanical exercise and preferred to assign this responsibility—and blame when students and parents complained—to the computer. Gradually, however, teachers began to realize that there was a lot wrong with traditional grading and that they needed to be able to control the computer grading program rather than the program controlling them.

The introduction of standards into the teaching and learning process greatly influenced the moves made by many teachers and schools to standards-based grading. Increasingly, there were complaints that computer grading programs were inflexible and were controlling what teachers could do to determine grades. The developers of these programs responded to these concerns by making their programs more flexible with changes, such as:

1. providing additional calculations;
2. making it possible to record scores for all assessments but not have all scores included in the calculation of grades;
3. adding symbols for identifying late and missing assessments;
4. including a variety of grading scales; and
5. recording of evidence of student achievement tied to standards in place of assessments.

These were all positive changes. In my opinion, one of the market leaders in computerized grading software, PowerSchool, blazed a trail with the addition of median, mode, weighted mean,

highest and most recent scores to the calculations for any set of (standards) scores. These grading advancements clearly showed to teachers that, unless a student performed remarkably consistently, the same set of scores could result in a very different summary metric and thus very different grades. It also made it clear to teachers that grading isn't merely a mechanical, numerical exercise and that professional judgment has to be brought to bear on the determination of grades.


There are in fact no "right" grades, only justifiable grades.

These electronic grade book enhancements provide significant improvements over existing grading programs. However, teachers, who may have upwards of 150 students, still had to visually scan all student scores to search for grades that needed additional consideration.

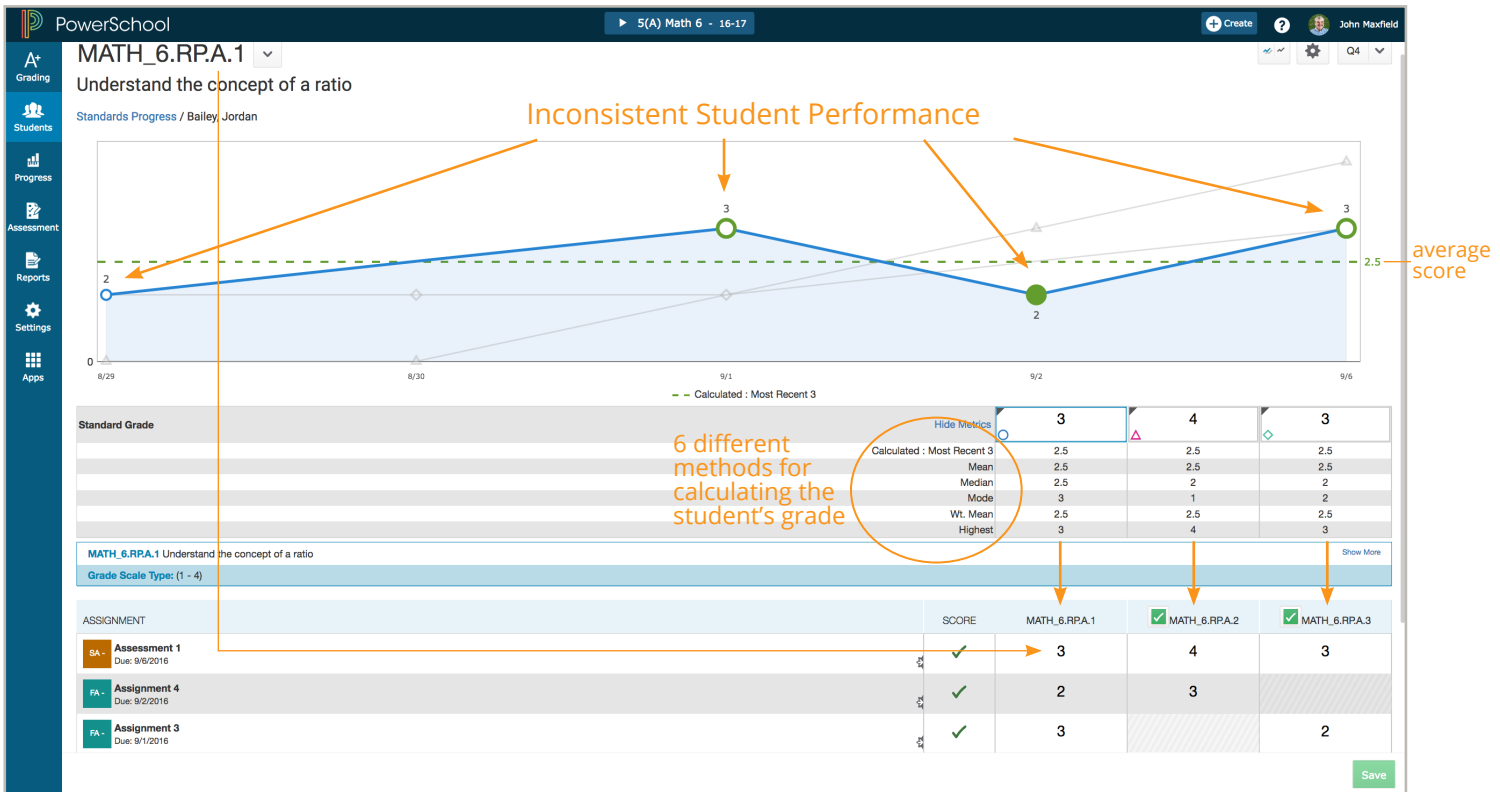
Here is an example of the metrics available to teachers when assigning grades:

Student	Standards Scores	Total	Mean	Most Recent 3	Grade
A	0 1 2 3 3 3	= 12	2.0	3.0	3
B	1 2 2 3 4 3	= 15	2.5	3.3	?
C	2 3 4 4 4 4	= 21	3.5	4.0	4
D	1 1 2 3 4 4	= 15	2.5	3.3	?
E	1 2 3 2 2 2	= 12	2.0	2.0	2

The grade assignment decisions are easy for students A, C and E, but more difficult for students B and D. If these sample scores are representative of the scores for 150 students, there would be 90 students for whom the grade assignment decision is easy and 60 for whom a professional judgment should be considered. Scanning the scores for 150 students to see which students' grades need special consideration would take a lot of time.

To make this task easier for teachers for students such as B and D, PowerSchool has recently created a feature that highlights those students by placing a Professional Judgment Indicator icon next to their names  so the teacher can immediately

see which student grades require their attention. Please see examples highlighting the grading advancements recently created by PowerSchool on the next page. The scores for one standard are shown graphically with the three most recent scores highlighted, and below the graph, the six metrics are displayed. As the student has not been performing consistently on assessments, the Professional Judgment Indicator feature alerts the teacher that he or she should consider whether the default, the average of the three most recent scores, is the appropriate summary grade for this standard or whether it should be changed based on his or her professional judgment.



BARNES, Heyes - Standards Progress

PS.HIST HISTORY

Grade Scale Type: (1 - 4)

STANDARDS	Q1	Q2	S1	ASSIGNMENTS	Q3	Q4	S2	Y1
1(A-B) Earth Science								
PS.HIST - HISTORY	3	3	3	16	3	3	3	3
PS.HIST.A - Concepts	3	2+	2+	17	3	3	3	3
PS.HIST.A.1 - Vocabulary	3+	2+	2+	15	3	2+	2+	2+
PS.HIST.A.2 - Process Skills	3+	2+	2+	15	3	2+	2+	2+
PS.HIST.A.3 - SCIENCE	3	3	3	14	3+	2+	2+	2+
PS.SCI - Science	3	3	3	12	3	2+	2+	2+
PS.SCI.A - SCIENCE	2+	3	3	13	3	2+	2+	2+
PS.SCI.A.1 - Understands Concepts	2+	3	3	14	2+	2+	2+	2+

Icon alerts the teacher that they may want to consider using a metric other than the average for this student's grade

One of the hallmarks of being professional is that members of a profession make decisions based on evidence and their specialized knowledge and experience. In general, when a person challenges a professional on one of his or her decisions, the professional defends it by citing the evidence and reasoning that led him or her to the decision. I believe that teachers are professionals and as such require electronic grade books to provide them with the freedom to use their judgment when assigning grades. A professional judgment indicator feature allows teachers to make, in Tom Guskey's words, "carefully reasoned decisions" and to use "thoughtful judgments" to determine grades. Teachers can now utilize new advancements in grading technology to sort, organize and present student achievement data and to use their professional judgment to ensure students are assigned grades that accurately describe their level of academic performance and achievement.

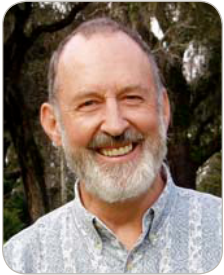
REFERENCES

- Guskey, T. R. (2013, September). The Case Against Percentage Grades. *Educational Leadership*, 71(1), 68-72.
- Guskey, T. R. (2002). Computerized Gradebooks and the Myth of Objectivity. *Kappan*, 83(10), 775-780.





Ken O'Connor a.k.a. The The Grade Doctor



Ken O'Connor is widely recognized as an industry expert in grading practices in K-12 education, and specializes in issues related to the communication of student achievement,

especially grading and reporting. Ken enjoyed a 23 year teaching career and was the Curriculum Coordinator responsible for Student Assessment and Evaluation and Geography for the Scarborough (Ontario) Board of Education. Through books, articles, presentations and working with small groups, Ken helps individuals, schools and school districts to improve communication about student achievement. Ken has a M.Ed. from the University of Toronto as well as a B.A. (Hon) and a Diploma of Education from the University of Melbourne.

Sally l'Anson, Ed.D.



Sally l'Anson has trained educators in curriculum, instruction, and assessment as well as in data-driven educational practices. She has taught at many levels, served as an administrator in private, public, and

charter school settings, and consulted with state departments of education. Dr. l'Anson earned a doctorate in curriculum and instruction from Teachers College, Columbia University, a master's in educational leadership from the University of Maine, and a bachelor's in economics and French from the College of William and Mary.