Common Errors in Calculating Final Grades

by Richard W. Francis

A. A. Milne's timeless classic Winnie-the-Pooh begins with the sentence "Here is Edward Bear, coming downstairs now, bump, bump, bump, on the back of his head, behind Christopher Robin. It is, as far as he knows, the only way of coming downstairs, but sometimes he feels that there really is another way, if only he could stop bumping for a moment and think of it." After teaching measurement and evaluation for many years, I can't help but note a similar predicament in how we determine student grades. In particular, I've discovered that errors often occur when scores are combined for final marks. These errors are not related to the grading of individual assignments. Rather, they occur when teachers at all grade levels bring individual test and assignment scores together for the students' final grades. Unfortunately, professors of mathematics and psychometrics have stood silently by and not alerted the teaching profession to the mathematical and measurement errors that are common when bringing scores together,1 making it difficult for teachers to find the information that would help them avoid these errors. This article will explain the four errors and how to correct them.

In an attempt to learn more about this problem, I conducted research comparing the effects of four methods of bringing scores together on class rank.[2] I asked professors in various subject areas to submit raw scores for each test and class assignment from one of their classes and received responses for 37 classes. The scores for

Richard W. Francis, *Professor Emeritus, Department of Kinesiology, California State University, Fresno, taught measurement and evaluation for 36 years and developed and taught the graduate course in statistics for 32 years. He developed and directed the Institute for Fitness Research, which became the California State Data Center for physical fitness testing in 1987. He also introduced the statewide electronic data analysis of physical fitness test scores. He earned an Ed.D. in Advanced Kinesiology from University of Northern Colorado in 1975.*

each class were accumulated and ranked using the following categories: (a) total raw score points, (b) percent correct, (c) letter grades, and (d) T-scores (see glossary).

Remarkably, it was common for students to move up or down three to six class ranks because of mathematical and measurement errors. Many students moved 10 or more ranks. In one class of 68 students, one individual moved 32 ranks. As a result of mathematical and measurement errors, the average class rank change per student ranged from a low of 1.8 to a high of 4.1 in the classes studied.[3]

This class rank manipulation suggests that students have been, and will continue to be, denied the academic status they have rightfully achieved. Conversely,

It was common for students to move up or down three to six class ranks because of mathematical and measurement errors.

other students are receiving grades and class rank status above their academic performance. Class standing should depend solely on a student's academic performance, free from mathematical and measurement errors. But because of the unintended class rank manipulation resulting from mathematical and measurement errors, class rank is not always a valid measure of academic achievement.[4]

There are four common errors teachers make when computing the students' final grades: (a) the Average Speed Error, (b) the Weight Problem, (c) the Natural Variation Violation, and (d) the Mars Climate Orbiter Miscalculation. Any one of these errors in the accumulation of scores will result in the unintentional manipulation of class rank and several invalid final grades independent of student performance.

If you travel 120 miles at 60 mph and then travel 120 miles at 30 mph, what is your average speed? The common mathematical mistake is to assume that if you average 60 mph and 30 mph, you will get 45 mph. The quotient derived from a division problem involving two independent variables cannot be combined or averaged with other like quotients unless the denominators or the quotients are equal. The average quotient for problems involving non-equal denominators can be obtained by totaling the numerators and totaling the non-equal denominators and then doing the division."[5]

Miles per hour is a quotient derived from the fraction makeup of the independent variable *time* and the independent variable *distance*. The correct method is to determine the total time and distance traveled—2 hours and 120 miles (at 60 mph) and 4 hours and 120 miles (at 30 mph)—which results in a total of 6 hours and 240 miles. Divide the total 240 miles traveled by 6 hours to obtain the mathematically correct average speed of 40 miles per hour. In the division problem: (240 / 6 = 40), 240 is the total of the numerators, 6 is the total of the non-equal denominators, and 40 is the average quotient.

The combining and averaging of quotients from fractions made up of two independent variables with non-equal denominators is a common mathematical error in percent correct grading. The total points possible make up one independent variable and the points earned make up a second independent variable. If one test is worth 100 points and another assignment is worth 50 points (non-equal denominators), the percentages (quotients) cannot be combined or averaged without making a mathematical error similar to the Average Speed Error. Some teachers compound the Average Speed Error by using various multiplication methods

The combining and averaging of quotients from fractions made up of independent variables with non-equal denominators is a common error.

in an attempt to weight or make the numbers equal.[6]

If you combine all of the points for correct scores (numerators) into one score and all of the points possible (denominators) into a second score, you can divide the total points scored by the total points possible, and the result will be a mathematically correct score for the semester. This percent correct score, however, cannot be averaged with other semesters or years without doing the required math.

While the above mathematical correction will result in a seemingly valid percent correct score, there remains a measurement error when teachers use percent correct or total points to determine the final grade. The vast majority of teachers, including those in higher education, are often unaware of the Weight Problem. The source of this error is the variation in student scores on different tests. When student scores vary more on one test than on another, the test with the higher variation in scores will have more effect—or weight—on the final grade than the instructor intended.[7]

The standard deviation is a measure of how scores vary: The greater the variation between scores, the larger the standard deviation, and vice versa. The larger the standard deviation, the more weight an assignment or test is given in the final grade. The smaller standard deviation is given less weight. If one assignment has twice as large a standard deviation as another assignment, it will carry twice as much weight toward the final grade.[8] This is true even if the assignment with the lesser standard deviation is worth more points. An extreme example, borrowed from Ory and Ryan[9] and adapted for this discussion, is found in Table 1. Column 1 lists students designated by letter; column 2 the midterm test scores with 100 points possible; column 3 the class rank for the midterm test; column 4 the final exam scores which had 200 possible points; column 5 the class rank for the final exam; column 6 the total points earned out of a possible 300 points; column 7 the final class rank; column 8 the combined percent correct (computed correctly); and column 9 the class rank for the percent correct.

The variation between each of the midterm scores is 10 points, while the variation between the final exam scores is only five points. Therefore, the midterm test had twice as large a standard deviation as the final exam. As the ranks are analyzed, it is clear that the midterm results established the final ranking even though it was only worth half the value of the final test. The ranking for the percent correct was also established by the midterm test.

Table 1
The Effect of Score Variation on Final Class Rank

Student	Mid-term	Rank	Final Exam	Rank	Total Points Correct	Total Percent Correct	Final Rank
Points	100		200		300		
A	100	1	155	10	255	85.00	1
B	90	2	160	9	250	83.33	2
C	80	3	165	8	245	81.67	3
D	70	4	170	7	240	80.00	4
E	60	5	175	6	235	78.33	5
F	50	6	180	5	230	76.67	6
G	40	7	185	4	225	75.00	7
H	30	8	190	3	220	73.33	8
I	20	9	195	2	215	71.67	9
J	10	10	200	1	210	70.00	10

Source: Adopted from Ory, J.C. and Ryan, K.E. *Tips for Improving Testing and Grading*. Newbury Park: Sage Publications.

This illustrates how the standard deviation weights the test independent of the teacher's intentions. While the instructor intended that the final exam would be worth twice the weight of the midterm test, in reality it was worth only half of the weight of the midterm. In the section detailing the Mars Climate Orbiter Miscalculation, I will discuss how to account for the standard deviation.

When the natural variation of a variable is reduced, important information is lost and another measurement error is committed. For example, if each test grade is converted to a letter grade, a score of 90 percent counts the same as a score of 100 percent, since both percentages receive the same grade. An extreme example occurs when a zero grade counts the same as a 59 percent score, since they both receive the same grade of "F".

Earlier, I cited an example of a student who moved 32 ranks when proper techniques were used. This student was at the bottom of each point spread for a grade, whether the grade was an A, B, C, D, or F. However, on each test or assignment, the low points counted the same as the highest points in that grade.

The student did not take one test and his zero score was maximized to equal 59 percent. Letter grades artificially reduce the variation between students. Often the greatest errors in class rank calculation occur with the reduced variation model.[10]

It was a $125 million mathematical mistake! Lockheed Martin Astronautics calculated the navigational data for the Mars Climate Orbiter in English units, while NASA Mission Control calculated the distance to Mars in metric units; consequently, the orbiter burned up in the Martian atmosphere.[11]

Most people understand that there is a difference between metric and English

When numerical grades are converted to letter grades, 90 percent is maximized to equal an A, while 100 percent is minimized to the same grade.

units. But, when the units of measurement are not specified, as in grading assignments, people often have a difficult time realizing that there is a difference. When we don't use a standard unit of measurement to combine the scores in grading, we make the Mars Climate Orbiter Miscalculation. Many studies have shown that unless tests have the same mean and standard deviation, they are on different measurement scales and cannot be combined.[12]

When using standard scores, the raw scores are all converted to the same mean and standard deviation. This approach is used in all standardized tests, such as the SAT, GRE, or GMAT. Their scores are normalized because population distributions tend to be normal. Because most class scores tend to be skewed, normalizing standard scores for class grades would add an unnatural manipulation of the distance from the mean. Therefore, a linear transformation is justified in class grading because the only interest is in the average distance from the mean. I recommend linearly transformed T-scores, which have a mean of 50 and a standard deviation of 10. Linearly transformed T-scores are computed directly using the individual scores and the mean and standard deviation for each graded assignment (test 1, test 2, etc.). If test 1 is worth 25 percent of the grade, you would multiply the T-score by .25. When different measurement scales are converted to standard scores, they then have a common mean and standard deviation and can be weighted and combined.

Table 2 is simply Table 1 with the addition of correctly weighted and combined T-scores in column 10 and the final ranking in column 11. The obvious change is in the final ranking. The T-score ranking is exactly the opposite of the original ranking because the final exam is now truly worth twice the weight of the midterm test. In this contrived example, the class rank error was extreme to demonstrate the potential problems caused by adding total points or percent correct scores. The resulting manipulative effect on students' final grades, independent of their academic performance, can be considerable. Knowledge is expanding in every facet of our lives. Sadly, the combining of scores into a final grade has not kept pace with other areas of education.

Table 2
Correcting the "Weight Problem" With T-Scores

Student	Mid-term	Rank	Final Exam	Rank	Total Points Correct	Total Percent Correct	Final Rank	T-Scores	Final Rank
Points	100		200		300				
A	100	1	155	10	255	85.00	1	67.57	10
B	90	2	160	9	250	83.33	2	69.22	9
C	80	3	165	8	245	81.67	3	70.78	8
D	70	4	170	7	240	80.00	4	72.52	7
E	60	5	175	6	235	78.33	5	74.17	6
F	50	6	180	5	230	76.67	6	75.83	5
G	40	7	185	4	225	75.00	7	77.48	4
H	30	8	190	3	220	73.33	8	79.13	3
I	20	9	195	2	215	71.67	9	80.78	2
J	10	10	200	1	210	70.00	10	82.43	1

Source: Adopted from Ory, J.C. and Ryan, K.E. *Tips for Improving Testing and Grading*. Newbury Park: Sage Publications.

The best way to compute a linearly transformed T-score is to use a spreadsheet. I use Excel, which is found in Microsoft Office; therefore, the symbols used below apply to Excel. The T-score equals the score (X) minus the mean (Average) divided by the standard deviation (Stdev) times 10 plus 50. You then multiply the T-score by the weight of the assignment. If you want an assignment to be worth one-quarter of the final grade, multiply the T-score by .25: T=(((X-Mean)/Stdev)*10+50)*.25. You can use a spreadsheet to make the T-score computation and keep a complete, up-to-date class record. In Excel, the = sign is required at the beginning of the formula to indicate that you are performing a mathematical calculation. The parentheses establish the desired order of the mathematical process. The $ in front of the row number makes the row permanent, which is necessary when the formula is copied down through a series of rows. A block address includes a beginning cell address, a colon, and an ending cell address - C6:C30. As an example, when calculating T-scores in Excel for data we have placed in column C6 to C30; compute the mean in C31: =AVERAGE(C6:C30), and the standard deviation in C32: =STDEV(C6:C30). The weight for this assignment would be placed in D5. Calculate the first T-score in D6: T is equal to the score minus the mean: =(C6-C$31), divided by the standard deviation: =(C6-C$31)/C$32 times 10 plus 50: =((C6-C$31)/C$32)*10+50 times the weight for that assignment: =(((C6-C$31)/C$32)*10+50)*D$5. Then copy the formula down to D30 by pulling the lower right corner of D6 down to D30. This formula assumes the high score is best, however, if the low score is best as in many timed events; just switch the score and the mean: =(((C$31-C6)/C$32)*10+50)*D$5.

I have encountered several complaints related to using standard scores. The first is that, when using standard scores, grading is on a curve, therefore as many "F" grades must be given as "A" grades. This is a misconception. Teachers have an obligation to use good judgment and the option to draw the cutoff point for each grade level, as they deem appropriate. The second complaint is that a student's grade is dependent on how the other students perform. This statement is also without merit because the teacher establishes the cutoff points. If the teacher has a class of 30 students and 20 are working at the A level, the teacher may give 20 A grades. But, if in a class of 30, only two students are working at the A level, the teacher may give only two A grades. If the entire class is doing poorly, there simply will not be any A grades with the use of standard scores. This method places students in the correct class rank, which enhances the teacher's ability to use good judgment in setting the cutoff points. Another complaint is that students do not understand standard scores, which is true. If all teachers used standard scores, though, students would quickly learn to understand them.

When using standard scores, the instructor converts all scores to a common metric and the class ranks are not artificially manipulated. The teacher now has the ability to make an informed judgment regarding grade cutoff points based on valid information. In addition, there will be improved scholarship applied to grading as a result of the accurate application of mathematics.

GLOSSARY

Class Rank—the position of a student when scores or grades are arranged from highest to lowest.

Denominator—the number below the line in a fraction showing the number of units in which a whole is divided.

Dependent Variable—a quantity in a logical or mathematical expression whose value depends on an independent variable.

Independent Variable—a variable whose values are independent of changes in the values of other variables.

Mean—the sum of a list of numbers, divided by the number of numbers, commonly called the "average."

Metric—a standard measurement.

Numerator—the number above the line in a fraction showing the share of the whole.

Quotient—the amount derived from dividing one quantity by another.

Raw Scores—data as reported (i.e., with no weighting assigned to them).

Standard Deviation—a measure of how spread out the data are from the mean.

T-Scores—a standard score that assigns the mean a value of 50 with a standard deviation of 10.

Weight—the relative importance of a particular statistic. nea

Source: University of California at Berkeley, **www.stat.berkeley.edu/~stark/SticiGui/Text/gloss.htm.**

ENDNOTES

[1] Modern textbook writers on the topic of grading tend to present only the current grading methods and never discuss the errors that result when combining individual scores into a final grade. Ory and Ryan are one exception to the above and provide an excellent discussion of errors in grading (1993), 118-120.

[2] The author conducted research in grading as a sabbatical leave project, 1993.

[3] Results from the Sabbatical Leave Research Report, 1994.

[4] One of the conclusions of the Sabbatical Leave Research Report, 1994.

[5] The Postulate for Independent Variables is a new mathematical postulate developed by the author. In mathematics, there is a tendency to think of all fractions as the same even though the quotients derived from fractions made up of two independent variables cannot be averaged or combined like normal fractions when the denominators are not equal.

[6] When you multiply or divide test scores to change their weight or value, you artificially increase (when you multiple) or decrease (when you divide) the size of the standard deviation and its influence in weighting the test or grading assignment independent of the instructor's intended weight.

[7] Edwards, (1954), 104-106; Ory and Ryan, (1993), 118-119.

[8] Ory and Ryan, (1993), 118-119

[9] Ibid., 119

[10] One of the conclusions of the Sabbatical Leave Research Report, 1994.

[11] Fordahl, Report in the *Fresno Bee* (March 30, 2000), A4.

[12] Edwards, (1954), 104-105; Guilford(1965), 510-516; Ory & Ryan (1993), 118-119; Roscoe, (1975), 74-78; Toothaker, (1986) 136-137.

WORKS CITED

Edwards, A. L. *Statistical Methods for the Behavioral Sciences*. New York: Holt, Rinehart, and Winston, 1954.

Fordahl, M. "NASA Chief Admits His Cost-Saving Measures Went Too Far." *Fresno Bee*, March 30, 2000, A4.

Francis, Richard. *The Manipulative Effect of Four Methods of Grading on Class Rank*. (Sabbatical Leave Research Report), 1994.

Garrett, H. E. *Statistics in Psychology and Education*. (5th ed.). New York: David McKay Company, Inc., 1958.

Guilford, J. P. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill Book Company, 1965.

Milne, A.A. *Winnie-the-Pooh*. London: Penguin Books, 1954.

Ory, J. C. & Ryan K. E. *Tips for Improving Testing and Grading*. Newbury Park: Sage Publications, 1993.

Roscoe, J. T. *Fundamental Research Statistics for the Behavioral Sciences*. (2nd ed.) New York: Holt, Rinehart and Winston, Inc., 1975.

Toothaker, L. E. *Introductory Statistics for the Behavioral Sciences.* New York: McGraw-Hill Book Company, 1986.