

Never Scale Grades

Copyright Pat Griffin, November 3, 2022

Suppose you are a professor, and you devise a test such that you believe a score of 90-100 indicates full mastery of the subject, 80-89 indicates good mastery, 70-79 indicates some mastery and a score less than 70 indicates no more useful mastery of the subject than someone who never took the class. Further, suppose that the test results indicate the highest score is 80. What should you do?

One far too common option is to scale the grades. This might mean the person who scored 80 gets an A and someone who scored 70 gets a B, etc. By doing this, you have relinquished your role as professor to the students. That is, you have let the students decide which scores indicate mastery of the subject. Why would you do this?

A common argument is if you don't scale, you don't understand the nature of numbers or psychometric scaling. In other words, if you stick to your preconceived scoring system, you are assuming an interval scale (the difference between 80 and 90 is the same as the difference between 60 and 70 in terms of knowledge of the subject) or maybe even a ratio scale (a score of 100 indicates twice the knowledge of the subject as a score of 50). At best, you only have an ordinal scale (a score of 90 indicates more mastery of the subject than a score of 80). So, you should look at the results and cobble together a scale that reflects that distribution.

Following this recommendation results in a test based on what the students do know, **not**, what they should know. Remember, you designed the test to reflect what a student needs to know to demonstrate mastery of the subject. A second consequence may result from scaling the test. You may be tempted to remove items that few if any students got correct. So, now you have a test that may be easier than the original. If you then scale this test and remove items, your test gets even easier.

The results of the scenario are inflated test scores and students with the impression they have mastered a subject when they have not. If these students are required to take a standardized test on the subject, they may be surprised by their low performance. The latter may lead to an abrupt change in their academic plans. This surprise may be particularly harmful if it comes so late in the students' academic progress that it cannot be remedied without extraordinary effort.

Now consider an alternative approach. Assume the same results as above and assume you believe the nature of numbers and psychometric scaling argument that all you have is an ordinal scale. But, an ordinal scale is all you need. You designed the test such that a score of 90-100 indicates full mastery. It is not necessary to assume anything other than scores in this range indicate greater mastery than scores in the 80-89 range. Further, you know what is required of the students, i.e., what they should know, and the test results show what they do know. The results are used only for this purpose and not to validate the test ¹.

What are the consequences of this approach? First, students which receive an A in the course earned it and will likely show mastery of the subject on a standardized test. Secondly, the items that few if any students got correct can be used to tweak the instructional material and/or lectures. They are not simply eliminated as some sort of outliers.

By designing a test with mastery in mind and abiding by the original scale, the product will be students whose score indicates mastery. The students will not be subject to the disservice of scaling and can confidently continue on their academic journey.

¹ Test validation is a difficult and sometimes complex as well as controversial subject and may involve testing, retesting, and scaling. These procedures are appropriate as are procedures to determine the extent to which the professor's definition of mastery of the material reflects reality. I am assuming here the professor's judgement is sound.