The Scales of Judgment

Materials scientists are generally well-versed in physics, and physics, above all, is a science of measurements. The first instinct of a physicist is to parse a problem in terms of its measurables in the dimensions of mass, length, and time, and it is the shifting of attention down the scale of length that particularly characterizes our present times as the Nano Age.

In materials laboratories, there has traditionally been an understanding that complete "quantitativity" is not always achievable, but it is better to have a qualitative assessment than none at all. This presents us with a challenge, because our training in physics endows us with a particular value system. If it can't be measured, it's not respectable, so we strive to make our qualitative assessments as quantitative as possible, or at least to seem that way. Long in the past, before the evolutionary lines of materials and geology separated, we could comfortably use comparative measurements like Moh's scale of hardness, which places materials in order by determining which of them will scratch others. Once all of the materials are ranked, ranging from talc to diamond, an arbitrary numerical scale is imposed, assigning a number from 1 to 10 to each material in the list, and providing an illusion of quantitativity. After the great evolutionary divide, however, materials scientists have tended to assess "hardness" in terms of indentation tests rather than anything so arbitrary as scratching one rock against another. At first, indentation tests produced results in terms of standardized "hardness numbers" according to the scales provided by the manufacturers of the testing machines: Rockwell or Vickers. Later, with a wider range of micro- and nanoindenters becoming available, hardnesses began to be reported in numbers that have actual SI units. Clearly, materials science is evolving toward physics, while geology is still in a primitive state, characterized by the striking together of rocks. Just don't get any ideas about hardness—even if it is measured in newtons per square meter—having any physical meaning. It is an example of illusory quantitativity. The number you get will still depend on how you do the test: doubling the load does not double the area of the indentation.

Our many tools for chemical analysis have made us more comfortable with different kinds of analysis, including "qualitative," "comparative," and "quantitative." Interestingly, some manufacturers impose a patina of quantitativity even on this scale, replacing its midpoint with "semi-quantitative": literally, "half-quantitative." I wonder how it was determined to be half-, and not, say, 65%-quantitative. Once when I asked a student to describe the nature of her analytical results, she wavered for a moment and then said "Quanlitative!" Lewis Carroll would have been proud. A new example of what he called a portmanteau word: a merger of two equally appropriate (or inappropriate) terms to give an ideal compromise. Whatever terminology you use, the message is clear: quantitative science is the best, which is by itself a value judgment, too.

Outside the laboratory, we are cast into a world of confusion where it is common to take even simple quantitative measures and obscure them with meaningless names, rather than numbers or simple descriptors. Who, on entering one of the emporia of Starbucks' coffee empire for the first time can place in their correct order the terms "venti," "tall," and "grande?" And when it comes to qualitative measures, the agenda seems even more clearly to confuse the user. Which of the following is more desirable: "Standard" or "Commercial"? (Make your guess before you read on.) How about "Choice" or "Select"? Still trying to work it out? These are four of the eight grades of meat that are approved by the U.S. Department of Agriculture, and the full list in descending order of price (which, after all, is the only truly quantitative measure of quality) would be Prime, Choice, Select, Standard, Commercial, Utility, Cutter, and Canner. How did you do? Could you pick the better grade from each of the pairs that I offered you? Whatever happened to simple designators like "small," "medium," and "large," or "good," "better," and "best"? Even if the latter triad is neither quantitative nor objective, at least it is clear.

All of those who apply to the National Science Foundation (NSF) to support their research are familiar with this simple scale of quality: Excellent, Very Good, Good, Fair, Poor. This would seem to be a reasonable standard scale for subjective assessments. It is pleasingly unambiguous, although I still wince at the grade inflation that it embodies. Once upon a time (when I more frequently heard the phrase "once upon a time"), it made me feel highly accomplished if a teacher wrote "good" on my work, and getting a "very good" was often a cause for a parental reward of some kind. These days, "good" is a major disappointment and "very good" is usually no cause for celebration, either. Despite our familiarity with the NSF scale, many others are still in play. A recent request from a publisher asked me to use a three-point scale of "Innovative/Outstanding," "Good," or "Worthy/Acceptable." Now, in the NSF world, "Good" is not really worthy or acceptable, so how is one to use a scale like this? Is the middle of the list truly the middle of the scale or was it somehow misplaced in transcription? What do those slashes mean? Do they stand in place of "and" or "or?" Is no manuscript ever "Unworthy/Unacceptable"?

Given that all assessments of quality depend upon personal taste and are therefore subjective, unquantitative, and often irreproducible, a simple, clear, and consistent scale should still be attached to all such measurements: a sort of SI of subjective assessment, if you like. It needs to have a sufficiency of levels so people do not feel the need to check two adjacent boxes (I confess, I have always wanted to check two non-adjacent boxes on an NSF review, just to see if the system rejects the notion that a proposal could be, say, excellent/fair.) It needs to be instantly clear to any user—and it should not abuse the English language by making "Good" into the median if the median is unacceptable.

In a recent (unscientific, unquantitative, highly subjective) office poll, I established that the following 8-point scale is correctly ordered from best to worst, and readily understood by most people, at least when it is read aloud:

Bad! – **not bad** – good – not **bad** – not bad – **not** bad – bad – not good.

There are also informal extensions of the scale in both directions, but all of them require the inclusion of some kind of profanity. This Unified Scale of Quality could be applied equally well in the review of proposals or the meat department at the supermarket, at least so long as you never get engaged in a conversation with the butcher: "Do you have any **Bad!** steaks?" might not give sufficient distinction between the first and seventh points of the scale, and could be difficult for those of us with strange accents but, hey, nobody talks to the butcher any more—especially if he has an iPod. And if you could get his attention, you could just say, "I want the **Baddest!** steaks you have," which would be instantly understood. Now, if you want to save a little money and go one step down the scale, then I'm afraid things are going to get complicated.

Maybe others can come up with better offerings than my new Unified Scale of Quality, but in order to compare them and make a sound selection of which is the best one for universal application, we need a means of comparing them. Since superiority evidently comes from quantitativity, we need a scale of quantitativity against which to assess this kind of subjective measurement or other "figures of merit" that might arise in the laboratory. Or we could just accept that quantitativeness is not always an intrinsically valuable quality, although judgment is.

ALEX KING

https://doi.org/10.1557/mrs2006.184 Published online by Cambridge University Press