

Who has the lowest prices?

Author: Lloyd Bond, Senior Scholar
The Carnegie Foundation for the Advancement of Teaching

March 2005

Abstract: Bond calls our attention to the many traps associated with one of the most frequent uses of assessment: the technical difficulties of measuring changes in learning over time.

Essay:

If one wished to know what knowledge or skill Johnny has acquired over the course of a semester, it would seem a straightforward matter to assess what Johnny knew at the beginning of the semester and reassess him with the same or equivalent instrument at the end of the semester. It may come as a surprise to many that measurement specialists have long advised against this eminently sensible idea. Psychometricians don't like "change" or "difference" scores in statistical analyses because, among other things, they tend to have lower reliability than the original measures themselves. Their objection to change scores is embodied in the very title of a famous paper by Cronbach and Furby, "How should we measure change, or should we?"

Fortunately, many educators have chosen to take this advice with a grain of salt. And well they should. The logic underlying the difference between what Johnny knew before instruction and what he knows after instruction is simply too compelling to be trumped by statistical niceties.

The power of change scores to reveal important aspects of teaching and learning is best illustrated by example. I was fortunate to have had a remarkable teacher for my first statistics class. He began the class by assigning us to teams of three each. We were given a week to answer a simple question, and we had to describe and justify the things we did to arrive at an answer. The question was, "Among the three local grocery stores, Kroger, A & P, and Hi-Lo, who has the lowest prices?" How do novice statistics students go

about answering such a question? My team was typical. We began in a haphazard and utterly frustrating way, proposing one inefficient strategy, then another. One member should take canned goods, another fresh fruits and vegetables.... Maybe one could take aisles 1, 2, and 3, another aisles 4, 5, and 6.... We quickly realized that no matter how we broke the task down, a census of every item in each store would take weeks. One of us had a vague notion about sampling, but we had no idea of how to conduct a scientific sample, let alone one weighted by purchasing patterns. We thought only of the average (arithmetic mean) price, never considering other measures of central tendency, and standard errors were not a part of our vocabulary. Nor did we consider the sampling implications of price changes from one day to the next.

Although we were initially enthusiastic about answering this simple question, by the end of the second day, we complained bitterly about the impracticality of it all. The instructor's response was always the same, "Well, just do the best you can."

For those of us who survived the course, the same question was assigned again toward the end of the semester. The instructor gave us minimal feedback on our responses the first time around. He simply put our reports away and they were never mentioned. Both sets of responses, along with detailed comments, were returned at the end of the semester. The difference in quality between the two sets of responses to the same question was stunning.

Our responses to the question did not figure in our final grade. Grades were awarded on the basis of other quizzes, examinations, and projects. The instructor used the results to grade his own teaching. (In retrospect, one would think that such effort expended on an activity that had no immediate grade payoff would have been resented. To my knowledge there was not a single complaint.) The instructor told me years later that that simple question, "who has the lowest prices?", was in the back of his mind during the entire portion of the course devoted to inferential statistics. It brought a certain coherence to the way he sequenced successive ideas central to the novice student's understanding of what is required when making statistical inferences.

In passing, it is noted that the grocery store question is a powerful example of what cognitive psychologists call "ill-structured" problems: problems that can rarely be solved quickly, that may have more than one defensible solution, that may have multiple routes to a single solution, and that may have many sub-problems that must be solved before arriving at an answer. By contrast, "well-structured" problems (e.g., solve for x in the equation $3x + 2 = 17$) have a unique answer, can usually be solved quickly, and have a very limited number of ways to a solution. The grocery store question also has enormous "pulling power." It evokes a variety of different answers and different approaches to the answers, and it provides deep insight into students' thinking, into how they organize what they know into a coherent argument.

For years I have argued that measurement and assessment should have a more prominent place in teacher education curricula. I still believe that. But beyond a good knowledge of the essentials, teachers need not be assessment experts. Nor need they fret over

measurement specialists' admonitions about measuring change. Rather, teachers could spend their time more productively by concentrating on what they want their students to know and be able to do at the end of the year. Often this implies something as simple as asking the right question of their own teaching.

 THE CARNEGIE FOUNDATION
for the ADVANCEMENT *of* TEACHING
51 Vista Lane
Stanford, CA 94305
<http://www.carnegiefoundation.org/perspectives/>

© 2005