

Numbers That Bedazzle, Numbers That Benumb

April 21, 2011

By **Bernard Fryshman**

Numbers fascinate and inform. Numbers add precision and authority to an observation (although not necessarily as much as often perceived). The physical sciences revolve around the careful measurement of precise and repeatable observations, usually in carefully controlled experiments.

The social sciences, on the other hand, face a much more challenging task, dealing with the behavior of people who have an unfortunate tendency to think for themselves, and who refuse to behave in a manner predicted by elegant theories.

Under the circumstances, it's really quite remarkable that statistical predictions are as useful as they are. Advertisers ignore, at their peril, conclusions based on data gathered on large numbers of people acting alike. Supermarket shoppers or football fans behave in much the same way, no matter the infinite number of ways each member of the population differs in other respects. In their interaction with the location of shelved foods -- or forward passes caught -- few of these variations make a difference.

Population samples comprised of large numbers of uniform members can be defined, observations made, statistical calculations made, and policy deduced with astonishing accuracy.

Efforts have been made to extend this methodology to the classroom, and trillions of data elements have been gathered over the past 30 years describing K-12 activities, students, inputs, and outcomes. But judging from the state of K-12 education, little in the way of useful policy or teaching strategy has emerged. The reason is not immediately clear, but one surmises that while the curriculum path for K-12 children is similar, the natural variation among children, in teachers, in social circumstances and in school environment makes it impossible to create a uniform population out of which samples can be drawn.

At the postsecondary level, the problem facing the number gatherer is greatly exacerbated. Every student is different, almost intentionally so. A college might

have 25 different majors each with three or four concentrations. Students take different core courses in different order, from different teachers. They mature differently, experience life differently and approach their studies differently. When all the variables which relate to college learning are taken into account, there is no broad student population. Put another way, the maximum size of the population to be examined is one!

This reality informed traditional accreditation. Experts in a field spoke to numbers of students, interviewed faculty, observed classroom lectures, and, using their own experience and expertise as backdrop, arrived at a holistic conclusion. There was nothing "scientific" about the process, but it proved remarkably successful. This is the accreditation that is universally acknowledged to have enabled American colleges and universities to remain independent, diverse, and the envy of the world.

In 1985, or thereabout, voices were heard offering a captivating proposal. Manufacturers, they said, are able to produce vast numbers of items successfully, with ever-decreasing numbers of defects, using counting and predictive strategies. Could not similar approaches enhance higher education, provided there were sufficient outcome data available? Some people, including then-Secretary of Education William Bennett, swallowed the argument whole. Others resisted, and the controversy played itself out (and was recorded!) in the proceedings of the National Advisory Committee on Accreditation and Institutional Eligibility (predecessor of the current National Advisory Committee on Institutional Quality and Integrity) between 1986 and 1990.

Advocates persisted, and states, one by one, were convinced of the necessity to measure student learning. And measure they did! Immense amounts of money, staff time, and energy went into gathering and storing numbers. Numbers that had no relevance to higher education, to effectiveness, to teaching or to learning. "Experts" claimed that inputs didn't count, and those who objected were derided as the accreditors who, clipboard in hand, wandered around "counting books in the library."

At one point, the U.S. Department of Education also adopted the quantitative "student outcomes" mantra, and accrediting agencies seeking recognition by the education secretary were told to "assess." "Measure student learning outcomes," the department ordered, "and base decisions on the results of these

measurements."

Under duress, accreditors complied and subsequently imposed so-called accountability measures on defenseless colleges and universities. In essence, the recognition function was used as a club to force accreditation to serve as a conduit, instead of barrier, to government intrusion into the affairs of independent postsecondary institutions.

Today, virtually all those who headed accreditation agencies in the 1990s are gone, and the new group of accreditors arrived with measured student learning outcomes and assessment requirements firmly in place. Similarly, college administrators hired in the last decade must profess fealty to the data theology. Both in schools and in accrediting agencies, a culture of assessment for its own sake has settled in.

But cautionary voices remain, arguing that the focus on quantitative measures and the use of rubrics which have never been substantiated for reliability and validity, are costly to the goals of teaching and learning.

Numbers displace. Accreditors have been forced to rely on irrelevant numerical measures, rather than on the intense direct interaction that is one of the essentials of peer review. If there are failings to accreditation, they are at least partially due to decisions made on the basis of "data," rather than the intensely human interaction between site visitors and students, faculty, alumni, and staff.

Numbers mislead. Poor schools are able to provide satisfactory numbers, because the proxies proposed as establishing institutional success are, at best, remotely connected to quality and are therefore easily gamed. Bad schools can almost invariably produce good numbers.

Numbers distort. Participants at a national conference sponsored a few years ago by the U.S. Department of Education were astonished to learn that colleges had paid students to take the Collegiate Learning Assessment. Other researchers pointed out that seniors attributed no importance to the CLA and performed indifferently. Under the circumstances, it is impossible to use CLA results as a basis for a value added conclusion. Can we legitimately have a national conversation about the "lack of evidence of growth of critical thinking" in college, based on such data?

Numbers distract. The focus on assessment has captured the center stage of

national educational groups for almost two decades. A quick review of annual meeting agendas of major national education conferences reveals that pervasive assessment topics moved educators from their proper concentration on learning and teaching. Seemingly, many people believe that effective assessment will result in improved teaching and learning. One observer compared this leap in logic to improving the health of a deathly ill person by taking his temperature. The current emphasis on "better" measures, then, would correspond to using an improved thermometer.

Numbers divert. Faculty members spend an untold number of hours outside of classroom time on useless assessment exercises. At least some of this time would otherwise have been available for engagement with students. Numbers divert our focus in other ways as well. Instead of conversations about deep thinking, lifelong learning, and carefully structured small experiments to address achievement gaps, faculty must focus on assessment and measurement!

Assessment has become a recognizable cost center at some institutions, still without any policy outcomes or improvements to teaching and learning, in spite of almost thirty years of effort.

This is not to be taken as a blanket attack on numbers. There are fields, particularly those with an occupational component, for which useful correlations between numerical outcomes and quality can be made. There are accrediting agencies which are instituting numerical measures in a carefully controlled, modest fashion, establishing correlations and realities, and building from there. Finally, there are fields with discrete, denumerable outcomes for which numbers can contribute to an understanding and a measure of effectiveness. But many other accreditors have been forced to impose measuring protocols, which speak to the flaws noted above.

It's time to restore balance. Government must begin to realize that while it is bigger than anyone else, it is not wiser. And those who triggered this thirty-year, devastatingly costly experiment should have the decency to admit they were wrong (as did one internationally known proponent at the February 4th NACIQI meeting, stating "with respect to measuring student learning outcomes, we are not there yet").

The past should serve as an object lesson for the future, particularly in view of the recently released Degree Qualifications Profile (DQP) bearing all the signs of

another "proxy" approach to the judgment of quality.

Our costly "numbers" experience tells us that nothing should be done to implement this DQP until after a multi-year series of small experiments and pilot programs has been in place and preliminary conclusions drawn. Should benefits emerge, an iterative process with ever more relevant features can be presented to the postsecondary community. If not, not.

But no more should a social experiment be imposed on the American people, without the slightest indication of reliability, validity or even relevance to reality.

BIO

Bernard Fryshman is an accreditor and a professor of physics.