Student evaluations of teaching: brought to you by computer

Joseph S Fulda

Address for correspondence: 701 West 177th Street, #21 New York, NY 10033, USA
Email: Fulda@acm.org

The Chronicle of Higher Education, the academy’s trade paper, recently reported on the results of a little experiment with student evaluations undertaken by an inventive and skeptical professor. He found that teaching otherwise identical courses differing only in teaching style had quite different evaluations for such unrelated substantive items as the quality of the textbook! I was amused but not surprised at his little experiment yet the reaction by the Chronicle’s published respondents conceded only that individual items on such surveys are neither specific nor sensitive.

I believe that this attitude is probably unreflective of the attitude of the professoriate as a whole. Indeed, it is widely believed that these instruments are not a valid measure of teaching effectiveness (ie, whether students learn, not whether they are made happy or, at least, comfortable) or a reliable measure of any variable (everyone has their own method of picking the day on which they are handed out and plans that class session with particular attention to student comfort levels). Nevertheless, because they are quantitative, lending themselves to aggregation, summary, and statistical analysis by computer—and in a matter of minutes, it has been the lot of numerous assistant and associate deans to devote the better part of their scholarship to teasing out the exact meaning of contradictory—or more often appositional and puzzling—responses to essentially meaningless aggregates of unreliable and invalid indicia.

No one claims that students should not have a voice, perhaps even a determining voice, in the determination of teaching effectiveness. But there is a method—albeit one which would require of those assistant and associate deans the real scholarly tasks of synthesis and integration—which grants students such a voice, and is likely to be fairly reliable and more-or-less valid. That is, and we draw a legal analogy here because few would dispute that the professor has a proprietary interest, that is, a legal interest, in his job, to simply ask of students who are graduating seniors to serve as bona fide witnesses and to write up a signed one- or two-page letter of evaluation for each professor they have had—without questions to lead (or mislead) the witness—to which would then be appended their transcripts of record. The latter is, of course, essential since evidence of quality of teaching is opinion evidence and that is admissible only when those giving the opinions are experts. The transcript serves as the informal equivalent of a voir dire, and we would expect of the assistant and associate deans that they would weight the
composite picture the student draws in his letter by the student’s academic standing (perhaps excluding the courses of the professor under evaluation). Since witnesses—and students are hardly an exception—often have agendas beyond mere truth-telling as they see it, cross-examination is normally allowed. I suggest as an informal equivalent that before the file is complete, the professor be given an opportunity to add clarifications, circumstances, and situations that perhaps the student overlooked (intentionally or otherwise) in his letter of evaluation. The picture that would emerge from all this would very likely be an honest one.

The present system, in contrast, causes both grade inflation and the lessening of serious intellectual content in coursework, not so much directly—that is a factor only with the minority of the professoriate without tenure, but indirectly. Professors, after all, are people first and professors second and it is only natural that they should want those they interact with to like them, and, in the case of Generation X—at least at most colleges—that means both inflated grading and courses which do not require the sophistication and discrimination such courses required a mere 25 years ago. They know better, of course, but pressure from the Chairman (to avoid grading complaints) and the Dean’s office (to maintain enrollment) gives them a ready excuse to relax their standards. If, however, they were actually called to account on the issue of teaching effectiveness they might not slip as often or as far. Now, however, they have an even stronger excuse for going with the flow—the need to maintain high evaluations—and an ex post justification that they are “successful” teachers. But if the students don’t learn very much—as countless surveys show—then the teachers probably aren’t teaching very much, and an initial restorative would be to strip such teachers of the illusion of success that numerical comfort-based ratings provide and to hold them to account for what they were hired to do: strive to bring forth the best in their students with all their considerable energies and, then, evaluate the results with genuine integrity.

It is particularly important for those who are computer ethicists to make these points, because without computer technology—and, indeed, before computer technology—such systems would (did) not exist. This is one instance where the computer has had a wholly negative impact on education, serving as the indispensable aid to the watering down of curricula and the wholesale inflation of transcripts of record. Eventually, AI may make possible the composition of an aggregate picture drawn from extensive written evaluations, but, for the nonce, the repetitive, numerical, and calculational tasks at which computers excel have been allowed to replace the carefully considered judgment of individuals and the collective wisdom that emerges from many such judgments. Computers are being taught in AI laboratories across the world to make judgments and show wisdom, but such projects are in their infancy—at the application level, it is Lotus 1-2-3 or Excel (perhaps with SAS or SPSS) which is being used to take the best system of higher education in the world down a few notches at a time.