Reforming Higher Education Through Statewide Examinations

A Proposal by:
Daniel Bonevac, Professor & Chair, Department of Philosophy, University of Texas and
Robert C. Koons, Professor, Department of Philosophy, University of Texas

PUTTING TEATHE IN HIGHER EDUCATION

Several University of Texas faculty members, including Robert C. Koons (Professor, Department of Philosophy), Joseph Horn (Professor, Department of Psychology), and Daniel Bonevac (Professor & Chair, Department of Philosophy), have proposed a new system of measuring excellence in State Universities, colleges, and departments within these institutions. This proposal, *Texas Excellence through Assessment of Teaching in Higher Education* (TEATHE), calls for students earning a bachelor's degree from a state-affiliated college or university in Texas to take an appropriate standardized test— typically, one of the tests comprising the Graduate Record Examination (GRE). Students could use the scores to demonstrate mastery of their fields. Although students' scores would be recorded on their transcripts, no specific score would be required for graduation. Test data would be used to measure not how smart a student is, but rather the *change* in student learning. Just as importantly, others could use them to evaluate programs of higher education.

THE NEED FOR MEASURING STUDENT LEARNING

Trends in higher education in Texas pose serious challenges for Texas policymakers, administrators, students, employers, and taxpayers. Due to changing demographics, a larger student population is heading toward the State's colleges and universities. In addition, access to higher education needs improving across the board, especially for minorities, in order to prepare a workforce for an information-based economy. Colleges and universities are already developing programs in distance learning and web-based instruction, and will surely emphasize them more in the future. All these changes put stress on our system of higher education.

As higher education adapts to increased demands, the question of how to make rational decisions about programs, directions, and resources will intensify. This is true at several levels.

It will be vital for legislators and University System officials to reward campuses and programs that successfully educate their students. However, currently there are too few ways of measuring this. We count inputs in the form of dollars, student/faculty ratios, and percentages of tenured and tenure-track faculty. We measure outputs in the form of degrees awarded or credit hours earned.

We have no way of measuring what really matters - <u>learning.</u> Without a rational basis for decision-making, the competition for resources becomes largely political.

EVALUATING TEACHING

College and university administrators (Presidents, Provosts, Deans, and Chairs) recognize that learning is their goal, and that it is important to reward good teaching and effective programs. But we currently have no rational way to identify such accomplishments. Administrators make serious attempts to evaluate teaching. But even the best methods tell us little about how much students learn.

Administrators evaluate faculty by considering a number of factors: 1) research, 2) student surveys, 3) peer reviews, 4) exit surveys, and 5) post course evaluations.

1) Research, a major factor in faculty evaluations, is relevant to teaching primarily at the graduate level.

2) Student surveys given during the class have little relation to effective undergraduate teaching. Student surveys do not measure learning directly, but instead measure entertainment value, the ease of the course, the likeability of the professor, and other factors. Most surveys do not even ask students how much they learned.

3) Peer reviews, conducted by other teachers, usually stop after promotion to full Professor. In any case, the faculty who conduct peer reviews, having much greater acquaintance with the field, are often not well suited to judge how effectively a teacher addresses an audience of novices.

4) Student exit surveys tend to measure only the very best teachers and are heavily influenced by a student's most recent and short term attitudes. Graduating students are much more likely to mention professors in their major field they have encountered as seniors than professors in other fields they encountered as freshmen.

5) Later evaluations by students and peers often differ substantially from those conducted at the end of the course. It is hard to say which evaluation is more accurate or trustworthy.

Administrators make serious attempts to evaluate teaching. But even the best methods tell us little about how much students learn.

Most importantly, all these measures are subjective. They measure what people <u>like</u>, not what people <u>learn</u>.

EVALUATING PROGRAMS

These problems are multiplied at the level of programs. No one has objective information about how well programs educate students. Often, administrators reward innovation per se. So, programs spring up and are rarely pruned back. Decisions are, in the absence of any useful information, political.

Increased demands on higher education, coupled with the lack of benchmarks to measure learning, produce downward pressures on educational institutions. Using the number of students, credit hours, or degrees as measures of success gives institutions incentives to lower standards. Using student/faculty ratios or percentages of tenured and tenure-track faculty as measures of quality gives institutions incentives to consume more resources. Nothing gives institutions any incentive to increase productivity or student learning.

The lack of a measure of student learning also poses serious problems for students and parents. It is important for students pursuing a degree not just to gain a credential but to learn. Otherwise, we could simply mail students degrees on their twenty-first birthdays! But students now have little real information about how effective a course, teacher, program, or institution will be in educating them. The effects are everywhere. Frequent mismatches between students and programs lead to high dropout rates, high rates of transfers between programs, and increased time-to-degree. Students rarely graduate in four years these days; five is typical, and six is common. Meanwhile, without meaningful information, students and parents base decisions on prestige, leading to a rush to certain institutions that enjoy good reputations but which may or may not teach students with certain backgrounds and goals effectively.

No one has objective information about how well programs educate students. Decisions are, in the absence of any useful information, political.

Finally, the lack of a measure of learning places a growing burden on employers, who need to know how much people know, and who are forced to spend vast amounts on testing and training of new employees. They too rely on prestige rather than concrete information about learning. Grade inflation is rampant in Texas higher education and employers know it. Thirty years ago, a 3.5 GPA was an above average score that reflected real academic achievement. That is no longer true. How much does a graduate of a Texas college or university - even one with a high GPA - actually know? Employers must either guess or administer their own tests.

OUR SOLUTION

These problems can be solved. How can one evaluate how much students with different backgrounds learn at different institutions with different grading standards? Graduate programs face just that problem every year. To solve it, they rely on standardized tests— the GREs, subject tests, and major field exams. GREs evaluate higher-level critical thinking skills as well as mastery of information.

By combining academic records with standardized test results, graduate schools find ways of evaluating how much students have learned in a given area.

We are proposing to do the same thing. To evaluate higher education programs, we suggest using the *value added* by the program – the contribution made by that program – by considering the difference between students' scores on standardized tests at or near the end of their program and their entering Scholastic Aptitude Test (SAT) or other test scores (such as TASP), together with high school GPA. Thus by controlling for entering achievement levels, we remove any disadvantage that might otherwise be perpetuated to lower skilled students.

The value-added measure can be used at various levels, to help answer questions such as: Which institutions and programs use resources most effectively? Which courses are central to progress in a given discipline? Which teachers teach most successfully? Which programs best educate students of certain kinds?

We are not proposing that our metric be the *only* tool used for answering these questions. Just as graduate schools consider standardized

test scores along with other information for enrollment, so administrators, students, employers, and the public should combine test results with other considerations to showcase how well certain programs meet certain goals. But without our metric, no picture can be complete, for it omits objective information about what matters most— how much students learn.

Some essential features of the TEATHE exams:

1. The TEATHE exams would consist of the GRE Subject Exams, administered by the Education Testing Service (ETS) in Princeton, plus, in subjects for which no GRE Subject Exam is available, the ETS Major Field Tests. At present, the GRE Subject exams cover fourteen major subjects, with the Major Field Tests covering an additional three subjects. In collaboration with ETS, the State of Texas will oversee the gradual expansion of this battery of exams to cover the majors of the overwhelming majority of college students in Texas by the year 2006.

2. The exams would be open to everyone. Registered students would be able to take the exams free of charge. For other students, or for repeat test-takers, a schedule of fees would be set by ETS.

3. Taking the GRE Subject or Major Field exams would be made a requirement for receiving the B.A./B.S. degrees in all covered subjects, at all state colleges and universities. Students would **not** be required to pass any of these exams in order to graduate. The score would be included on the student's official transcript. Individual test scores would be confidential: only aggregate results for whole departments would be made public.

4. Each major program at each state college or university would be evaluated in terms of the **value added** by instruction: that is, how well did students in the program do on the TEATHE exams, as compared to how well they would be expected to do, given their entering SAT scores and high school grades? Programs that excel could then be identified and rewarded with additional funding.

GREs evaluate higher-level critical thinking skills as well as mastery of information.

5. Each program in state-supported schools would evaluate each course, each instructor, and each course/instructor combination by means of calculating the differential contribution each makes to the value added by instruction. It would be possible to determine with some precision what effect each course has had on the exam results of the students who have taken it. It would also be possible to determine the instructional value added by each course, by controlling for the SAT scores and high school grades of the students in question.

Objections

- *GPAs and courses of study, as documented by official transcripts, provide students with ample opportunity to document their education achievement, and these provide prospective employers with all the information they need to assess the student's accomplishments.*

  If this were true, graduate programs at state colleges and universities would rely exclusively on GPAs and college transcripts in

making admission decisions. Instead, they uniformly insist on the addition of standardized, nationally-normed GRE test results. If grades alone are insufficient for graduate admission decisions, how can they be sufficient for all other purposes? Furthermore, many employers retest job applicants rather than rely on transcripts.

- *The TEATHE exams would cause a lowering of quality to a common denominator, by setting minimal proficiency targets (as the TAAS tests have done).*

  Unlike the TAAS tests, the TEATHE exams would provide a *range* of results, including very high scores that represent the caliber of the very best students admitted to graduate school. This would provide not a floor of minimum required achievement, but an inducement to the pursuit of excellence.

- *Instruction in colleges and universities would be distorted as teachers concentrate exclusively on "teaching the test."*

  It is an error to assume that multiple-choice exams are limited to testing factual knowledge. ETS regularly uses such exams in testing analytical and interpretive skills. In addition, it would be possible to require students to take nationally-normed essay tests, such as the GRE Writing Assessment or the ETS Tasks in Critical Thinking. Finally, a statewide collaboration between faculty, education consultants and the ETS will make possible the continuous improvement and refinement of the Subject Exams and Major Field Tests. For example, it should be possible for the exams to incorporate essay questions and problems in critical thinking. We must devise examinations that test for the skills and knowledge that are believed to be most important. Instead of "teaching the test," the faculty would have, over time, the responsibility of helping to create tests that effectively test whatever it is that they believe they ought to be teaching. The quality of instruction could only improve as faculty refine and articulate the objectives of their instruction.

- *Like other standardized exams, the TEATHE exams would incorporate cultural biases that disadvantage minority groups.*

  Programs would be evaluated by their instructional value added, controlling for the SAT scores of their students. This means that any cultural bias built into the SAT tests would be factored out of the evaluation of value added. Programs that enable students from socially disadvantaged groups to succeed and to excel would be recognized and rewarded. In fact, the introduction of the TEATHE exams would reduce reliance on SAT scores as a means of rationing access to higher education. Colleges would be eager to admit students who are disciplined and highly-motivated, even if their academic preparation for college is not all it might have been.

- *The TEATHE exams would favor prestigious campuses (like UT Austin and A & M College Station) over other campuses; and,*

- *The TEATHE exams would favor satellite campuses over the flagship institutions, like UT Austin and A & M College Station.*

  In fact, it is impossible to say at this point which programs on which campuses would prove to be most effective in instructional value added. A value added exam score from UT El Paso will be every bit as valuable as the same score earned after a degree from UT Austin, so the artificial value of merely attending a high-prestige campus would be

dramatically reduced. Programs and campuses that are effective in enabling students to perform better on the GRE Subject exams than expected, in light of those students' SAT scores and high school experience, would be identified and rewarded. Similarly, programs that spur students with high SAT scores to even higher levels of excellence in the TEATHE exams would also receive the recognition they deserve. The TEATHE exams would not arbitrarily favor one segment of the educational enterprise over another.

- *Why not simply use student evaluations or exit surveys to evaluate the quality of instruction?*

Student evaluations do provide valuable information, but are not an adequate measure of student learning. First, student evaluations do not provide a meaningful basis for comparing teaching at one institution with teaching at another. When students rank their teachers, they implicitly compare their teachers with other teachers at the same institution. Furthermore, there is no research demonstrating that high student evaluations correlate with successful instruction. Student evaluations represent how pleased students are with their in-class experiences, but students are typically not qualified at that point to measure and evaluate how much they have actually learned, nor how important or central to their discipline was the content of the course. Student evaluations are heavily influenced by irrelevant factors, such as the personal charm of the instructor or the entertainment value of the lectures. Finally, approximately fifteen percent of the difference between student evaluations of different courses can be explained by variations in grade inflation: students rate courses significantly higher when they receive a grade that is higher than their GPA. (See Max O. Hocutt, "De-Grading Student Evaluations: What's Wrong with Student Polls of Teaching," *Academic Questions* 1(Winter 1987-88):55-64.) The combination of student evaluations with independent and consistent assessments of student learning would provide more comprehensive and accurate information than student evaluations alone.

Endorsements

The following individuals have endorsed the basic framework of reform embodied in the TEATHE exam proposal. Endorsements identified in this list should not be taken as indicating that these scholars endorse every detail of the proposal in its present form.

Daniel Bonevac, Professor and Chair
Department of Philosophy
University of Texas at Austin

Walter Bradley, Professor
Mechanical Engineering
Texas A & M University

Ronald L. Carter, Professor
Department of Electrical Engineering
University of Texas at Arlington

John Cogdell, Associate Professor
Electrical and Computer Engineering
University of Texas at Austin

David B. Cohen, Professor
Department of Psychology
University of Texas at Austin

Donald G. Davis, Jr., Professor
Graduate School of Library and Information Sciences
University of Texas at Austin