April 3, 2013

CATHERINE P. KOSHLAND
Vice Provost for Teaching, Learning, Academic Planning and Facilities

  Subject: Online Evaluation of Courses Initiative, 2011-12 mid-term progress report

Dear Cathy,

Thank you for providing the mid-term progress report for the Online Evaluation of Courses Initiative, and for soliciting feedback from key Senate committees. On February 25, 2013, Divisional Council (DIVCO) discussed the progress report, informed by commentary from the committees on Budget and Interdepartmental Relations (BIR), Courses of Instruction (COCI), Educational Policy (CEP), and Graduate Council (GC). Each of the reporting committees highlights issues that fall within its purview. The comments are appended in their entirety.

While we appreciate the work of the initiative thus far to develop and refine the online evaluation of courses, the discussion in DIVCO underscored a number of issues and concerns.

- We discussed the value of open-ended evaluation questions. Ready access to the resulting written comments is critical as they provide valuable feedback.

- We remain concerned about response rates. These concerns are described well by both BIR and COCI.

- DIVCO members emphasized the need to gather data appropriate to the intended purpose, and to generate reports in useful formats. BIR's comments provide specific examples of the latter. While we recognize that an online system can collect important and useful data, we are also mindful of the limitations of such data to understand deeper, more complex issues, such as learning outcomes and teaching effectiveness.

- The discussion in DIVCO highlighted a number of factors that might skew evaluation data, such as whether the course is being taken pass/fail, whether the evaluation is submitted by a major or non-major, and the influence of the grade the student receives or expects to receive. Correlating evaluation data with these types of factors will provide needed context.

Sincerely,

Christina Maslach

Christina Maslach
Chair, Berkeley Division of the Academic Senate
Professor of Psychology

Encls.

Cc: Shannon Jackson, Chair, Committee on Budget and Interdepartmental Relations
        Daniel Melia, Chair, Committee on Courses of Instruction
        Ronald Cohen, Committee on Educational Policy
        Mark Stacey, Chair, Graduate Council
        Linda Song, Associate Director staffing Graduate Council
        Aimee Larsen, Manager, Committee on Budget and Interdepartmental Relations
        Linda Corley, Senate Analyst, Committee on Educational Policy
        Sumei Quiggle, Senate Analyst, Committee on Courses of Instruction
        Assistant Vice Provost Cynthia Schrager

January 24, 2013

CHAIR CHRISTINA MASLACH
BERKELEY DIVISION OF THE ACADEMIC SENATE

RE: Online Evaluation of Courses Initiative Mid-Term Progress Report

Thank you for the opportunity to comment on this report. We focus our attention on issues that affect this committee's capacity to successfully assess the teaching performance of faculty. We begin by recapping the major results as described in the report and then turn to the key issues for us.

In 2011-12, the Office of the Vice Provost for Teaching, Learning, Academic Planning & Facilities (OVPTLAPF) ran pilot tests of both a new questionnaire and of an online evaluation system. In fall 2011, 42 sections of eight departments distributed new paper evaluation forms. (Sections refer to main classes and the labs or discussion sections associated with the lecture classes.) In spring 2012, 71 sections of eight departments participated in the online evaluation. It is not clear which of the eight departments in the spring overlap with the eight in the fall. Response rates to the paper evaluations were higher (with a median of 81%) than to the online evaluations (with a median of 65%). However, the results were encouraging because 65% is much better than had been expected based on other institutions' experiences and, moreover, it had been achieved without providing students any incentives to fill out the evaluations. Notably, response rates for lectures and lab or discussion sections were as high in the spring as in the fall; response rates for small classes dropped. Students were more likely to fully complete the online evaluations than the paper forms. Staff members examined the open-ended comments and concluded that the length and quality of the online comments were comparable to those in the paper evaluations. (Other institutions have reported improvements in comment quality.) The OVPTLAPF sought student feedback on the experience of filling out the evaluations and faculty feedback on the utility of the evaluations, but received very few replies, so that exercise was not informative.

We have five comments:

1. We are encouraged by the results of the online experiment to date and appreciate the careful explorations conducted by the OVPTLAPF.

2. The matter of response rate is, as all parties understand, critical. The report repeatedly refers to an average 67% response rate as a target; it was nearly achieved in the online pilot. That assessment is, however, based on a misunderstanding. Current policy treats 67% as a *minimum* response rate for making a decent assessment of teaching in any particular class. A *median* response rate of 65% means that about *half* the sections yielded *unsatisfactory* data. We are led

to make two suggestions:

> (a) The project must seek to obtain much higher response rates as a norm such that few courses have yields under 67%;
> (b) In the event that further experimentation fails to achieve this goal, we must understand the nature of the bias introduced by shortfalls in response rates. Should we assume that had the non-respondents participated, they would have raised or would have lowered the observed ratings? There are survey and statistical techniques available to explore that question in the Berkeley experiments, and there may be findings in the student evaluation literature that would be useful.

3. It is critical for this committee to have the tools with which to assess, and to place in context, the scores that emerge from student evaluations. One of the promises of an online system is that generating many useful benchmarks should be automatic once the programming is done. We would want to know, at minimum:

(a) both mean and median effectiveness scores
(b) what the distributional features of the scores in any one course are, whether in the form of a histogram or similar data
(c) what the mean and/or median scores are for comparable courses, where the comparability takes into account the department; the level of the course (upper- or lower-division); the size of the enrollment; and, perhaps, depending on the department, the nature of the course (e.g., gateway introduction, required vs. not, methods, lab, etc.).

4. We request the implementation of an easy way to read the open-ended comments to any question, presumably via a link to the database.

5. There are further aspects of courses that influence course evaluations above and beyond the true quality of the instruction, such as workload. One critical factor is the grades students receive or expect to receive. We request a system for taking grades into account. This could involve presenting the grade-point average for students in each class along with the information requested above. It could also involve some more sophisticated method of calculating statistically adjusted versions of the course evaluations that take into account individual respondents' grades or, preferably, *anticipated* grades or the course's grade-point average. Again, we expect that the existing research literature on evaluations and the experiences of our peers could provide guidance. (We add a request that, as the OVPTLAPF experiments with incentives to raise response rates, it examine whether different incentives affect the numerical or written answers.)

Thank you for the opportunity to comment on the online experiment.

Shannon Jackson
Chair

SJ/al

February 8, 2013

CHAIR CHRISTINA MASLACH
Berkeley Division, Academic Senate

Re: COCI comments on Online Evaluation of Courses Mid-Term Progress Report

Dear Chair Maslach,

COCI appreciates the potential usefulness of better-designed student course evaluation forms and is encouraged by the high response rates achieved in the first online test period. The more (and better) information that instructors can gather about student responses to course materials and course presentation, the better. Learning quickly what is working and what isn't can help to improve teaching at every level. In addition, the reduction in workload for department staff is a major advantage of the online initiative.

We have two major concerns about the project, however. The first is the question of response rates. It seems unlikely that a one-term rate can successfully predict future online response rates in light of the generally weak longer-term rates found at other institutions, and we see little that is obvious to explain why the new Berkeley system should produce significantly better rates of return over time. The second concern is more general: what uses are the data gathered by this method to be put? Assemblages of numbers have a seductively solid appearance, but even the most carefully assembled data are of little use if they are counting something that is improperly described or non-existent. Even large numbers of UFO sightings by otherwise responsible adults do nothing to demonstrate the existence of extra-terrestrial visitors in the absence of other evidence. "Fifty-five percent of students rated this instructor 'excellent'" means literally nothing in the absence of information about what 'excellent' might mean to this particular group of students. No matter how excellent the data collection, the potential use of these evaluations as the sole evidence of teaching effectiveness in matters of promotion and tenure is disquieting, particularly since what ought to be measured is the degree of student learning, although few methods of measuring that quality have been devised as yet.

For a recent discussion of the potential validity of student evaluations versus student outcomes see:

Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors
Scott E. Carrell
University of California, Davis and National Bureau of Economic Research
James E. West

U.S. Air Force Academy
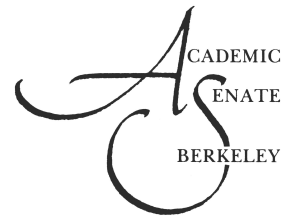*Journal of Political Economy,* 2010, vol. 118, no. 3

www.econ.ucdavis.edu/faculty/scarrell/profqual2.pdf

Sincerely,

Daniel F. Melia
Chair, Committee on Courses of Instruction

DFM/scq

Academic Senate Berkeley

320 STEPHENS HALL
UNIVERSITY OF CALIFORNIA

February 6, 2013

CHAIR CHRISTINA MASLACH
Berkeley Division of the Academic Senate

Re: Committee on Educational Policy Comments on the
Online Evaluation of Courses Initiative Report

Dear Chair Maslach,

At its January 30, 2013 meeting, the Committee on Educational Policy reviewed the Online Evaluation of Courses Initiative Report. Members were supportive of continuing a pilot to demonstrate effectiveness of the program.

Committee members discussed ways to have students participate in the online evaluations in class by using electronic devices. Members noted that encouraging students to complete online evaluations using smartphones and laptops in class might increase participation of the cohort that is actively participating in the course. However, some members felt that using electronic devices in class would exclude students that do not own a smartphone or laptop and suggested a 24-hour window.

With best regards,

Ronald Cohen, Chair
Committee on Educational Policy

RC/lc

February 7, 2013

PROFESSOR CHRISTINA MASLACH
Chair, Berkeley Division of the Academic Senate

Re: *Online Evaluation of Courses Initiative mid-term progress report*

Dear Chair Maslach,

At its February 4, 2013 meeting, the Graduate Council (GC) discussed the mid-term progress report for the Online Evaluation of Courses Initiative. Members were generally satisfied with the results documented in the report, but had the following observations.

Firstly, classes with lower enrollments (this would include graduate courses) will provide a different learning experience for students than large enrollment classes. Surveys should be developed that are tailored to smaller classes. From the report provided, it appears that language and writing courses, which are largely led by graduate student instructors, had much lower response rates in the online pilot. Variation in the success of the pilot between types of courses should be monitored and causes should be explored.

Secondly, the campus has approved two online degree programs and is considering a proposal for a third online degree program. Students enrolled in these courses will have to complete online course evaluations, so an online capability is required for them.

Finally, it is important to include graduate student instructors in course evaluations. We trust that they are included in the initiative, but the report did not make this clear.

Sincerely,

Mark Stacey
Chair, Graduate Council