Course Evaluation Committee Report

December 18, 2013

Course Evaluation Committee Report

Executive Summary

The Committee was charged to develop a proposal for a new course evaluation form closely tied to student learning in order to generate better data to assist faculty in improving their classes. The charge also specified that the new form should engage students in the learning process. After polling faculty, examining current course evaluation result patterns, conducting focus groups with undergraduate and graduate students, and piloting alternative evaluation instruments, the Committee proposes a new evaluation form and process with the following key features:

1) The course evaluation form will be customizable to fit the specific characteristics of each class. Customizing the form requires the instructor to indicate, in a pre-questionnaire, those distinctive features of a class that should be evaluated. In addition, a set of core questions will be asked for all classes. Instructors will be able to add their own questions.
2) The course evaluation form will emphasize student learning, including student responsibility for learning, rather than—as is now the case—focusing nearly exclusively on the instructor.
3) The questionnaire will list the course learning goals, as designated by the instructor, and students will be asked whether the class has met those goals. Courses with "Ways" breadth designation must include at least one learning goal pertinent to that "Way," and results of that question will be communicated to the committee overseeing the requirement.
4) The form should be "smart," automatically filling in students' degree program, major, minor, year and similar demographic data points which can be obtained when students sign in to complete the form.
5) The form should be designed for use on various devices, including smart phones. The interface should be user-friendly and aesthetically pleasing. Students who submit their evaluation forms receive early access to their end quarter grades, as is currently the case.
6) Instructors should set aside 15 minutes during the last week of class to permit students to work on the form in class, although students should also be able to amend, complete and submit the form through the end of the examination period. Instructors should explain the importance of the course evaluation process as a responsibility to the community, using a standard text or their own words. Participating thoughtfully in the course evaluation process should have the same standing as respecting the Fundamental Standard.
7) Quantitative results that will be helpful to students in making course selections should be made public, as is now the case, but instead of simply reporting the mean, we recommend showing the distribution of results, i.e., x% excellent, y% good, z% fair, etc.. In addition, qualitative comments in response to the question concerning what students should know about a course should be made public in

digest form, on the basis of algorithmic analysis or staff curating. These should be available as links to the Explore Courses course entry. Outliers and offensive comments should be eliminated.

8) Sophisticated reports should be made available to instructors in an interactive format to allow for sorting by various student demographic factors (major, minor, year, prior enrollment in particular classes, etc.). However, steps should be taken to protect student identity. Instructors should also be able to review individual results, but without the demographic data, as a further protection of confidentiality. Supporting materials for interpreting the results should be developed in parallel with deploying the new form, and evaluations results should be accompanied by explanatory documents prepared by the Center for Teaching and Learning (CTL).

9) It is important to recognize that course evaluation forms are only one instrument in a wider field of evidence pertaining to student learning. They should not be used in isolation for administrative decisions but should be used with other data, e.g., student performance in end-quarter projects, student letters, and student success in subsequent courses. Mid-quarter evaluations, especially Small Group Evaluations, should be encouraged and supported. Course evaluation form results should inform department-level discussions about the overall quality of curricular structures and learning trajectories. The university should establish a framework and provide resources to support a systematic monitoring of the quality of teaching and efforts to improve it.

10) The effectiveness of the new course evaluation form and process should be monitored closely and modified appropriately, leading to a comprehensive review after two years of operation.

PROPOSED COURSE EVALUATION FORM

Course Evaluation Committee, Dec. 2013

Q 1-4. *[Learning Goals (recommend no more than 4)]*
For each course learning objective presented, the student is asked:
How well did you achieve this learning goal in this course?

Extremely well, very well, moderately well, slightly well, not well at all

Q 5. About what percent of the class meetings (including discussions) did you attend in person?

___%

[Only if instructor indicated there would be online class meetings:]
Q 5a. About what percent of the class meetings did you attend online?

___%

Q 6. How much did you learn from this course?

A great deal, a lot, a moderate amount, a little, nothing

Q 7. Overall, how would you describe the quality of the instruction in this class?

Excellent, good, fair, poor, very poor

Q 8-12. *[Course Components (recommend no more than 5)]*

How useful to you were the _____? (Blank filled from instructor's specified course components: the lectures, the discussion sections, the reading assignments, etc.)

Extremely useful, very useful, moderately useful, slightly useful, not useful at all

Q 13. What skills or knowledge did you learn or improve? [Open-Ended]

Q 14. How many hours per week on average did you spend on this course (including class meetings)?

___ hours

Q 15. How organized was the class?

Extremely organized, very organized, moderately organized, slightly organize, not organized at all

Q 16. What would you want another student to know about this class? [Open-Ended] (A summary of replies to this question will be made public; your instructor will be able to view all comments.)

[Q 17. Optional faculty-posed question]

PROPOSED SECTION EVALUATION FORM

Course Evaluation Committee, Dec. 2013

Q 1. About what percent of the section/lab meetings did you attend?

___%

Q 2. How much did you learn from the section meetings?

A great deal, a lot, a moderate amount, a little, nothing

Q 3. Overall, how would you describe the quality of the instruction in this section?

Excellent, good, fair, poor, very poor

Q 4-8. *[Section components (recommend no more than 5)]*

How useful were the _____? (Blank filled from section leader's specified section components: discussions, writing assignments, oral presentations, group projects, problem sets, lab exercises, etc.)

Q 9. What skills or knowledge did you learn or improve in this section? [Open-Ended]

Q 10. How organized was this section?

Extremely organized, very organized, moderately organized, slightly organized, not organized at all

Q 11. Please comment on the overall performance of the section leader. [Open-Ended]

[Q 12. Optional section leader-posed question]

Contents

The Provost's Course Evaluation Committee was chaired by Prof. Russell Berman, Comparative Literature and German Studies, and met from October 2012 through November 2013. Committee members were:

- Tom Black, Registrar
- Prof. Sarah Church, Physics
- Robyn Dunbar, Ph.D., Director of the Center for Teaching and Learning
- Visiting Prof. Tom Ehrlich, Education
- Prof. Margot Gerritsen, Earth Sciences
- Prof. Pam Grossman, Education
- Prof. Caroline Hoxby, Economics
- Prof. Jon Krosnick, Communication, Political Science, Psychology
- Michele Marincovich, Ph.D., Senior Advisor to the Vice Provost for Undergraduate Education
- Corrie Potter, Institutional Research & Decision Support
- Prof. Sheri Sheppard, Mechanical Engineering

1. Introduction

In recent years, Stanford has underscored its strong commitment to teaching and learning. Both the Study of Undergraduate Education at Stanford (SUES) (January 2012) and the report to the Western Association of Schools and Colleges (October 2012) demonstrated Stanford's focus on student learning and on continuous reflection on the quality of teaching programs and practices. It is worth noting that this prominence of a teaching mission has not always been the case in research universities. Still, much remains to be done, as SUES reminded us.

A robust culture of teaching and learning that would engage students in metacognitive reflection while enabling instructors to refine their teaching practices would depend on many components, including candid discussions at the program and department level about curricular organization, resources to support teaching (such as are available at Stanford through the Center for Teaching and Learning), clear indications that the institution values teaching in hiring, promotion and compensation, and more. One indispensable tool to support teaching is the student course evaluation, which has a long history at Stanford. It is an important vehicle both to provide teachers information and to invite students to evaluate their learning.

Charged to design a new course evaluation form, this Course Evaluation Committee (CEC) has been able to confirm that the widespread doubts regarding the current evaluation form are warranted. The course evaluation system is broken; it does not provide reliable information, and many students as well as faculty members give it little credence, for reasons detailed below. We should do better.

Our recommendations include
- the adoption of a new course evaluation form that can be customized to match specific characteristics of the respective course;
- a form that shifts to a focus on the success of the course in promoting student learning instead of the current nearly exclusive focus on judging the instructor; and
- a form that explicitly asks whether a course has met instructor-designated learning goals.

In order to customize forms and to include learning goals, the process will require a preliminary step in which instructors provide relevant information that can be built into the form that students complete.

Yet it is not only the form itself that is at stake but also the modalities of its implementation that need improvement. We recommend that time should be set aside during the last week of instruction for students to begin the evaluation in class, in order to underscore the importance that the university attaches to the evaluation process. We also

recommend a form that can be deployed easily on various devices, including smart phones; while the evaluation should begin in class, we recommend that a window of time remain open until after the end of the examination period for students to complete their reflections on the course.

We also recommend that the form be "smart," so that students need not fill in demographic information on their own; that a sophisticated analysis of evaluation results be performed in order to explore relationships within and across courses; and that special attention be paid to preserving the confidentiality of student responses.

We furthermore recommend that the results of course evaluations be made accessible to the university public. Quantitative results are currently available online and should remain so, although we recommend that the distribution of student responses be reported, rather than a single mean. Some qualitative comments should also be published, as for example through links at Explore Courses. Publication of course evaluation was historically the case at Stanford. However, results for new instructors, first-time course offerings, and courses with fewer than eight enrollees should not be published.

Course evaluations at Stanford have a long past, primarily as a student initiative. Students circulated qualitative evaluations of courses long before the Faculty Senate endorsed the principal of universal course evaluation. While shifting course evaluations to an online environment has achieved some remarkable efficiency, it has also made the process less visible, especially as the qualitative comments disappeared. This may explain the doubts and cynicism many students express about the current course evaluation system. Our recommendations are designed to address these apprehensions in order to strengthen the learning partnership on which the teaching mission depends.

2. The Charge to the Committee

The charge to the CEC (Appendix 1), embedded in the invitation to participate extended by the Provost, the VPUE and the VPGE, proceeds from SUES and its call to provide faculty with "reliable information about how well students are learning in their classes." The charge also notes that "there is wide agreement that the current course evaluation system is inadequate for this purpose. It has not been updated in many years, so this effort is certainly due." As is demonstrated below, we have been able to document that many faculty and students believe that the current course evaluation process, including the form itself, no longer functions well. In addition, the CEC examined the quantitative results generated by the current form and found that the data corroborate the subjective perceptions that the current process does not yield useful information.

The charge also suggests specific goals, which have guided the work of the CEC. The relevant paragraph reads:

"Our goal is to develop an instrument that is more closely tied to student learning, rather than just satisfaction, and that is tailored to the specific class and its format, teaching

practices, and learning goals. It should also provide better data about student learning across classes and departments. An improved evaluation system has great potential to inform for the better what faculty can do to improve learning in the classroom. Equally important are questions of how we engage students in this process. We hope to develop a new course evaluation that provides better information to faculty, encourages students to reflect more thoughtfully about their educational experiences, and enhances the learning partnership."

This framework proved crucial for our work. The charge references the common complaint that student evaluation of teaching may often respond to "satisfaction," in the pejorative sense, implying a superficial popularity of the instructor or a consumerist stance on the part of the student. An alternative to documenting such mere satisfaction would require a superior metric of student learning that would go beyond student reports. To measure student learning objectively, tools such as evaluations of the quality of end-quarter projects, success rates in final exams, or student grades in follow-up courses would be useful. The importance of relying on a diverse set of tools to analyze student learning is central to the CEC's vision of developing a more robust culture of teaching and learning, in which the analysis of student end-quarter responses to the quality of individual courses is just one element.

While we recognize the importance of that wider field of indicators of student learning, the CEC's mission has been the design of a new course evaluation form, which depends inescapably on the student's perspective, and, in particular, the student's self-assessment of his/her learning (rather than on other, objective data). Shifting attention specifically to student learning, as mandated by the charge, stands in marked contrast to our current course evaluation form, which primarily asks the student to judge the instructor (rather than asking about the learning process or the student's own contributions to it). Insisting on the importance of this Copernican turn in the evaluation process—from a focus on the instructor to a focus on the learning—has been central to the CEC's thinking.

The charge also asks for the development of an instrument that can be customized to the specific character of each class. This goal can be reached, we believe, by utilizing more effectively the potentials of the online environment. The charge calls furthermore for more effective data analysis, "across classes and departments," which points to the broader context of implementation, which the CEC has also addressed: we have discussed not only the shape of a new course evaluation form but also how it can be utilized most effectively. At stake is the potential to use the course evaluation process in a way that goes beyond a report card on an individual faculty member in a single course in order to develop an informed understanding of the character of learning in departments and more broadly. An improved course evaluation form will have a significant impact if it is part of a process to develop an extensive culture of teaching and learning throughout the university.

The conclusion of the quoted charge paragraph has proven particularly important for the work of the CEC. "We hope to develop a new course evaluation that provides better information to faculty, encourages students to reflect more thoughtfully about their

educational experiences, and enhances the learning partnership." These three points have been decisive. First, we need to provide better information because, as we demonstrate below, the current system does not provide sufficient information and because faculty striving to fashion a more effective learning environment need to know more. Second, the current evaluation process rarely engages students in a reflective evaluation of their learning. On the contrary, we have ascertained a degree of cynicism and dismissiveness regarding the course evaluation system that does not contribute productively to the culture of learning at Stanford. Third, our ultimate goal involves a renewal of the "learning partnership" by fashioning a new evaluation instrument that invites both faculty and students to focus on the quality of student learning.

3. Course Evaluation at Stanford

This report does not presume to offer a comprehensive history of course evaluation practices at Stanford, their trajectories in different Schools, or their relations to changing understandings of the relationship between instructors and students. The School of Engineering in particular has its own history of course evaluation through Tau Beta Pi. Nonetheless, a brief review of the development at Stanford may be useful.

The evaluation of courses at Stanford appears to have developed as an independent student initiative, which then prompted faculty action. As early as the late 1950s, a student publication, *The Scratch Sheet*, provided substantive descriptions of courses with qualitative comments. No effort was made to poll all students. Instead the editors relied on information "from a number of students who are familiar with the courses and whose judgment we respect." No quantitative data were reported.

In the early 1970s, the ASSU was publishing a Course Review with summary descriptions of selected courses, with quantitative data eventually appearing alongside qualitative comments. A turning point was marked by the 1977 Report of the ASSU Task Force on Tenure and Teaching Quality, chaired by then graduate student Larry Diamond. It cited widespread student concern that the university paid insufficient attention to teaching quality. It called for mandatory evaluation of all courses large enough to preserve confidentiality. The Faculty Senate's C-AAA Subcommittee on the Evaluation and Improvement of Teaching (SCEIT) took up this proposal, leading to Senate endorsement of universal course evaluation on May 12, 1977. SCEIT recommended that all quantitative data be released, including to the ASSU Course guide for publication, while qualitative comments were to be released only with the instructor's permission.

The ASSU Course Guide, publishing both quantitative and qualitative course evaluations, became the main vehicle for disseminating such information to the university public. Yet the SCEIT, chaired by Lee Shulman in April 1995, noted "poignantly consistent findings" from its interviews and focus groups that "neither faculty nor students believe[d] that current teaching evaluations [were] taken seriously." The report recommended that teaching be evaluated on a number of issues; in mid-course as well as end-course evaluations; and that chairs discuss evaluation results with all faculty members on a

regular basis. Even though the principle of universal evaluation had been established, dissatisfaction remained regarding the limited attention paid to evaluation results.

The online evaluation system was introduced in December 2006, and the existing form was carried over into the new environment. Quantitative evaluation results have been available online. Evidently the publication of qualitative results disappeared in this transition. The ASSU Course Guide, which had previously made results available for more than thirty years, ceased publication in early 2007 in light of the online reporting.

The current course evaluation form, found in Appendix 2, inquires methodically and comprehensively into the character of courses, particularly with regard to instructor performance. Its structure represented a reasonable strategy to understand the quality of teaching fifteen years ago. Since its adoption, however, key changes in the learning environment have occurred, undermining the form's effectiveness. The range of class types, from traditional lectures to new online formats, has diversified, rendering a single, standardized evaluation format insufficient. Student familiarity with technology and especially portable digital devices has grown enormously, which can allow for new protocols. An extensive national discussion has unfolded regarding student learning, which should inform evaluation processes. In light of these changes, the development of a new evaluation form is timely.

The current form consists of three parts:

a) Eight queries that collect a mixture of demographic data (year, area of study, reason for taking the course) and reports on the student's behavior (attendance, expected grade, time spent on course work outside of class). Two items can be construed as evaluations of the course itself (type of instruction and % [of time] that was valuable).

b) A series of 18 closed-ended questions, divided into six sections, with options to answer: Ex, Vg, G, F, P, N/A. Of the 18 questions, 17 explicitly involve an evaluation of the instructor, while the last question involves the integration of the section or lab. There is also a separate form to evaluate the section or lab.

c) Four pairs of boxes for open-ended comments, in which an invitation to comment on "strengths" is followed by an opportunity to offer "suggestions for improvement." The topics here concern the effectiveness of the instructor, the textbooks and readings, the value of the assignments, and "any additional comments."

The grounds for faculty and student dissatisfaction with the form are discussed at length below. For now, the following comments on each of the three parts of the form should suffice:

a) Some of the demographic data (e.g., year, area of study) could be populated automatically if the form were refashioned to be "smart," i.e. plugged into the rest of the Stanford information system. Because students have to log in to complete the form, the system ought to be able to designate the student's year and area of

study automatically. Asking students about their "expected grade" may not provide useful information and, in any case, certainly adds to student concern with grades; we recommend dropping that question.

b) All these questions (and most in c, as well) are instructor-directed, rather than learning-directed. They position the student in a judgmental, potentially adversarial relationship to the instructor. None of these questions directs student attention toward their own behavior or their responsibility for their learning processes.

c) The open-ended questions, too, focus the student on the instructor or the course, while not asking about his/her own contribution. Instructors sometimes report benefiting more from the open-ended or qualitative answers, rather than from closed-ended or quantitative scores. However, the placement of the comment boxes at the end of the form reduces the likelihood of significant student engagement in providing answers. Students generally report experiencing the form as too long, so by the time they reach the end, they're unlikely to provide insightful commentary. Survey fatigue is undercutting the quality of the information.

4. Faculty Perspectives on the Current Form and Process

The CEC sent email requests to all academic council members and to lecturers soliciting their opinion on the course evaluation protocol. A total of 66 responses were received—a low return rate but one which nonetheless provided some indication of faculty sentiment. A summary of the faculty suggestions is provided in Appendix 4 to this report.

Respondents expressed doubt as to whether students are well equipped to answer some of the questions, especially those having to do with the instructor's knowledge of the field. 45% doubted that students could determine whether the faculty member had "selected course content that was valuable and worth learning," and some indicated concern that the meaning of "valuable" may not be clear. In addition, 27% questioned whether the student could judge whether the instructor "displayed a thorough knowledge of course material."

In the CEC discussion, members noted that these and similar questions focus the student on the appearance of faculty authoritativeness in a way that privileges certain traditional pedagogical styles—"the sage on the stage"—and disadvantages innovative teaching techniques which might be more Socratic, discussion-based, or project-based. Similarly, placing a premium on the appearance of authoritativeness may tap into students' stereotypes that can disadvantage instructors who are women or members of minority groups. In general, the faculty responses suggest significant dissatisfaction with at least parts of the current evaluation process.

The responses also suggest that faculty members do not view the evaluation process as a source of productive information. 31% indicated little or no understanding of the purpose of the evaluations, with some regarding it merely as a sanctioned "venue for complaints and griping." That number seems to dovetail with the 71% who indicated that they utilize the evaluation results, although some explicitly ignore the open-ended answers and focus only on the numerical results (but the opposite preference is also present). While one can treat these results as indication that many instructors make use of their evaluation reports, it is also clear that close to a third ignores them. In fact, the fraction of the faculty that disregards the evaluation results may be larger, since the cohort of respondents was very likely biased toward instructors who take the evaluations seriously in the first place: faculty members who do not care about evaluations at all may have chosen not to respond to the evaluation committee's request for input. It is fair to say then that many faculty members do not utilize the evaluation results and that therefore the evaluation process is not currently providing all faculty members with the information with which they could improve their teaching.

Faculty respondents also offered a wide range of specific suggestions that were discussed by the CEC in the course of its deliberations. Many proposed developing a form that is customizable to reflect the learning goals and particular character of specific courses, a feature that became central to the committee's work. Many suggestions involved adding further questions in general, in addition to the prospect of instructors' being able to add their own questions. This interest in expanding the questionnaire indicates that faculty members are in fact looking for more information about the effectiveness of their classes, which maps well onto the CEC mission.

However, as became clear in the course of our deliberations, lengthening the list of questions in the pursuit of more information could be counter-productive, as survey fatigue sets in among students the longer the instrument grows. More questions may in fact yield less reliable information, unless the overall experience of the evaluation process could be significantly improved. In general, the real challenge involves increasing the collection of information while shortening the questionnaire.

A similar trade-off pertains to other suggestions from faculty members who asked for more precise information about student respondents in order to relativize their evaluations, e.g., whether low attendance correlates with low evaluation scores. While finer-grained detail about student identities might very well shed light on evaluation results, the committee discussed how expectations of greater self-identification on the part of students may undermine the claim that the process preserves student anonymity. Such perceived loss of anonymity may either reduce student participation in general or reduce the candor of the responses. As discussed below, students express considerable concern regarding the confidentiality of their comments.

The CEC also discussed the results of the VPUE 2013 Faculty Survey on Teaching with regard to course evaluation questions. That survey found that only 63% report reviewing the written comments "very often," or 61% for the numerical ratings. This compares to

the 71% result we obtained. However, in the VPUE report, an additional 24% indicated reviewing the results (written and numerical) "often."

As many as 66% report modifying their course in response to the evaluation "often" or "very often." However, they rarely discuss their course evaluation results. Only 28% discuss the content of evaluations with colleagues "often" or "very often;" and only 12% discuss them with their department chair "often" or "very often." Results vary across schools, with notable differences also between tenure-line faculty and instructors ineligible for tenure. The results of this survey corroborate our finding that the current course evaluation process has only a limited impact on teaching at Stanford.

5. Student Responses: Focus Groups

The Committee conducted a series of focus groups to gain insight into the student perception of the course evaluation process and to inform our design of a new instrument. In all of the sessions, some students reported having refrained from filling out the current form for at least one course they had completed (although the rate of non-participation was greater among undergraduates than among graduate students, many of whom displayed a more complex understanding of the process by virtue of their having served as Teaching Assistants and therefore having been themselves subjected to evaluations). The fact that students reported that they had avoided completing such evaluations formed the basis for a line of questioning into their dissatisfaction with the current form and process.

Certain themes recurred frequently. Students regard the current form as too long and therefore requiring too much time, an issue exacerbated by the scheduling of the evaluation process just before and during examination period, when students feel themselves under particular time pressure. In addition, questions on the current form are often seen as repetitive, especially when students are asked to fill out multiple forms for the same course: different forms for each of the members of a teaching team in team-taught courses, plus a form for the section or lab. In its appropriate effort to pursue precise points of information, the current form in fact frustrates students, discouraging them from thoughtful participation. Students would prefer to answer questions specifically tailored for the class or class type, rather than what is perceived as generic questioning. Students also criticized the interface (insufficiently "smart") and not aesthetically appealing.

In addition to the criticisms of the form itself, focus group participants mentioned aspects of the overall process that work against full participation and therefore reduce the quality of the information provided to the instructor. The timing of the evaluation at the end of the quarter provokes dissatisfaction (although moving it earlier or later would both present other challenges). Students also claim to have little understanding of what is done with the forms—despite a clear policy statement placed on the questionnaire itself. There is cynicism with regard to the question as to whether the professor reads them or responds to them or whether they genuinely matter in salary setting or promotion

(although some students also expressed a reluctance to make negative comments about some instructors precisely for fear of harming them). A further reason given to explain non-participation is the invisibility of the evaluation results to students: although the quantitative results are in fact available on Axess, many students do not know it, and others complain that the results are presented in a way that works against accessibility.

Qualitative comments are not available to the student public at all, in a marked contrast to the student comments available on CourseRank (https://www.courserank.com/stanford/main) (to which many students report turning, precisely because of their interest in these candid remarks)—and in contrast as well to the former practice at some colleges and universities, including Stanford, to publish a "course guide," with collections of student comments on courses.

It is particularly important to note that students repeatedly expressed doubts as to whether their evaluation forms and, especially, their qualitative comments remain anonymous. In small classes or small programs, so students worried, professors are likely to be able to deduce who may have made a particular comment, and they fear consequences at a later date, for example, in a subsequent class or otherwise in their careers within a program or department.

This largely negative attitude toward the course evaluation form and the process associated with it leads to student behavior that undermines the information value collected by the form. Students repeatedly stated choosing not to fill out the form unless they felt very positively or very negatively toward a course. This bimodality of responses means that the students in the middle who might provide a nuanced perspective drop out, leaving only the enthusiasts and the naysayers. Many students also report, that, due to time constraints, they pay attention to the first couple of questions, and then just select the same answer for the rest of the survey, in order to complete it quickly: so, while the questionnaire design strives for precision, the students' response strategy ends up blurring the picture. Finally, because students face multiple forms, which they already regard as excessively lengthy, at a particularly stressful time of the quarter, at least some students report completing them all one after another, sometimes even forgetting which course they are in the process of evaluating.

The focus groups were also asked what questions or topics they would like to see, ideally, on an evaluation form. These results are discussed below in section 9 as a key element in the formulation of the new instrument.

6. Data Analysis of Course Evaluation Results

As reported in detail in Appendix 5, we analyzed course evaluation results for a single academic year, 2010-11. Our goal was to determine the value of the information collected by the evaluation forms. We wanted to investigate whether an analysis of the data would corroborate the widespread expressions of dissatisfaction with the form.

The analysis found a lack of variation across items and straight-line responding: In the course evaluation focus groups, we heard students talk about picking a response category on the course evaluation, (e.g., 'Excellent,' 'Good') and marking the same category all the way down the form, regardless of the specific question. In fact we observed that 35% of students mark the same category for every item on the evaluation form. As a result, there is little meaningful variation across the different items in an evaluation at the class level.

We also conducted a factor analysis to examine relationships between the items and identify distinct dimensions (factors) underlying the responses. We found four unique factors: teaching, grading, content, and instructor responsiveness to students. These findings indicate that students' responses to all seventeen of the substantive questions are being driven primarily by their feelings on just those four dimensions, and support a dramatic reduction of the number of questions on the survey.

We found compression in the results at the course level. 75% of courses were rated 'Very Good' or better. While one can greet this result as indication of high quality teaching, it also means that the evaluation process provides little grounds to distinguish among the quality of different courses or teachers, and that the information it yields is insufficient to assist instructors in improving their teaching.

We also found noteworthy demographic patterns in how students respond. We found associations between student characteristics and their mean responses. For example, seniors rate classes more highly than freshmen. This could imply that freshmen are "tougher graders," that seniors take the evaluation process less seriously, or that seniors (understandably) have a better understanding of their interests and therefore select classes more appropriate to their interests. It might alternatively suggest that freshmen are less accustomed to college-level discourse and its culture of exploration, while seniors have been integrated into the critical thinking of the academy. We also learned that those who attend a higher percentage of classes rate classes more highly, and courses taught in the spring are rated more highly than courses taught in the fall.

In general, it appears that results from the current evaluation form are less than satisfactory because of straight-line responding. The prominence of the four factors in student responses suggests that the questionnaire could be shortened. Compression allows for little distinction in evaluations of quality, and the demographic factors further undermine the validity of results.

7. Practices at Peer Institutions

Ten peer institutions shared information about their student evaluation forms, and in many cases the forms themselves. A detailed discussion is included in Appendix 6.

Despite some variation, certain patterns emerged:

Mid-term evaluation forms are shorter and largely open-ended, while end-term evaluations are generally longer, with primarily closed-ended questions.

Questions about sections or labs are not frequently included on the main evaluation form but are reserved for a separate questionnaire. Questions about other topics appear with regular frequency: course overall (65%), course components (71%), instructor (70%), TA (64%), student learning (61%), student effort (69%). It is noteworthy that even within this narrow range, questions about the instructor (70%) and course components (71%) are at the high end, while student learning (61%) is at the low end. In other words, current practice elsewhere, as at Stanford, emphasizes evaluating the instructor over evaluating student learning. Based on the charge to the committee, the CEC has aspired to elevate the status of student learning and to steer away from placing primary emphasis on evaluating the instructor.

The Committee particularly noted a document provided by the Dartmouth Center for the Advancement of Learning, designed to assist faculty in the interpretation and use of student evaluations. It is included in Appendix 7, along with the current statement from Stanford's CTL (Appendix 8). The Committee recognizes that more is at stake than merely designing a new form: the evaluation process has to be explained to faculty, students and department chairs. This will require various strategies to embed the reformed evaluation process in a strengthened culture of teaching and learning. We need to be thoughtful about explaining the process to students, supporting faculty in their interaction with and understanding of course evaluations, and making clear to department chairs and Schools that the results of the course evaluation form should be treated as only one data point in a larger field of evidence regarding student learning and successful teaching.

8. Scholarly Literature on Evaluation Forms

There is considerable scholarly literature on course evaluation forms, let alone on the wider topics of student learning and survey design. A small sample of this literature is summarized in Appendices 9 and 10.

Three concerns in particular emerge from the literature.

First, there is good reason to believe that student assessment of instructors through course evaluation is a contributing factor to grade inflation. Instructor concern that course evaluation results may impact their compensation or career prospects can induce them to signal to students the prospect of a high grade, in the hope of increasing student satisfaction. That satisfaction, presumably, is expected to translate into high course evaluation results. It is difficult to see how this connection can be eliminated fully, but its impact may be dampened through a careful choice of questions and through administrative awareness of this problem. The more instructors and students view the

evaluation process as a vehicle to provide information to faculty as to how to improve their teaching rather than primarily as a basis for administrative decisions, the weaker the pressure toward inflated grading may grow.

Second, course evaluations may invite students to judge courses on the basis of their entertainment value, the personal popularity of the instructor, or other factors that are independent of objective learning gains. This too may be unavoidable, to some extent, although choosing questions carefully may reduce this factor somewhat. However, a robust evaluation of student learning would correlate course evaluation results with other evidence, such as student performance on final examinations or in follow-up courses. For departmental or school evaluation of the effectiveness of an instructor, a holistic review of student accomplishment should contextualize the results of course evaluation forms. This, however, would require resources and a framework to support systematic improvement of teaching that goes beyond reporting end-quarter results of individual courses.

Third, the committee notes that the best available evidence shows that course evaluation ratings are *not* related to learning as measured by performance in later classes, a sobering reminder of the status of the course evaluation process in general. The value of course evaluation results might be improved if more appropriate questions are asked and if the process itself is reformed, along the lines we suggest in this report. Nonetheless, individual faculty members, department chairs and School leaders should keep this cautionary note in mind. The real goal is the improvement of student learning, holistically and over time. Single course evaluation results are just small pieces of a larger picture and should be viewed in that context, rather than in isolation. For example, single course evaluation results should be combined with evaluations of overall department or program teaching effectiveness.

9. Process

The Committee reviewed input from faculty members, survey data, student focus groups, practices at peer institutions and the scholarly literature. Regarding the focus groups, participants were asked what topics they would like to see addressed in an evaluation form, and the results for all the focus groups were tabulated; they are collected in Appendices 13 and 14 (online only, at http://goo.gl/10GhfO and http://goo.gl/5u20vt respectively).

The focus group suggestions formed the basis of an initial draft survey, which was then modified by the committee to reflect findings from the faculty survey, the literature review on survey design, and what modern learning science says about student feedback. Two different models for a new questionnaire emerged, one somewhat longer form with a preponderance of closed-ended questions, and one shorter form with an emphasis on open-ended questions (although each form included both question types). These questions were then revised through a process of cognitive pretesting to determine student understanding of the phrasings, leading to further editing.

The two drafts of the new form were pilot-tested on 2,364 students from 19 classes, with the permission of the instructors. One third of the participating students were presented with the current form as a control. In addition to presenting the students with the forms—the closed-ended longer draft, the open-ended shorter draft, and the current version—all students were asked a set of final questions about their estimation of the course evaluation form that they had just completed.

The results of the pilot test shed important light on the challenges in evaluation design.

Our interest involved especially the student evaluations of the different forms. When asked how difficult it was to interpret the meanings of the questions in the questionnaires, students found the traditional questionnaire more difficult than either the closed-ended or the open-ended form, while the closed-ended and open-ended forms seemed to pose the same degree of difficulty. In other words, either of the new forms would appear to be superior to the traditional form, at least in terms of difficulty of comprehension.

When asked to indicate how difficult it was to generate their own answers to the questions on the questionnaire, students found the closed-ended form easier, and ranked the open-ended form closer to the traditional form. The preference for the closed-ended form is intuitive, since students largely needed only to select an option among multiple choices, rather than generating their own response to open-ended questions.

More students thought that the open-ended form was the right length to evaluate the course fully. In other words, significantly greater percentages of respondents thought that both the closed-ended form and the traditional form are longer than they should be.

See Appendix 11 for additional detail.

For the two classes involved in the pilot test in which the numbers of participating students were large enough to maintain anonymity, we presented the results to the instructors. Specifically, we provided each instructor with the tabulated results from the student answers to the closed-ended pilot form, the open-ended pilot form, and the current form as a control. The instructors were asked which form generated the most valuable information for them, and which questions generated the most valuable answers.

One faculty member provided feedback, emphatically preferring the new forms to the traditional one, particularly with regard to the question concerning the student's perception on how well the specific learning goals were achieved. This colleague indicated moderate preference for the open-ended form and considered the closed-ended form too long. The value of the closed-ended form would rest with an ability to undertake correlations across questions.

10. Proposed New Forms

Course Form

The proposed new form follows below. While the currently used form includes 34 discrete items, the proposed form has 17, assuming that instructors abide by the recommendation to include no more than 4 learning goals and no more than 5 course components. Yet even if they exceed those limits, this form is considerably shorter than the current one. We believe it will facilitate student participation, invite greater self-reflection, and yield better information for the faculty members. However, it is crucial to remember that the end-quarter evaluation instrument should be viewed as only one among several sources of feedback for the faculty member. More detailed feedback is likely to emerge from mid-quarter and small group evaluations.

There was some disagreement in the CEC concerning Q5: percent of meetings attended. Some members look forward to collecting longitudinal data on this point, while others doubt the reliability of self-reporting. We have included the item but consider it important to monitor its results.

There was also some disagreement about Q14: hours per week spent on a course. This information is useful because it can be correlated to the number of units, and certain results could lead to an adjustment of units assigned to a course. Some CEC members noted, however, that the question of hours per unit has been complicated by federal policy, and that many of our peers no longer work on a unit basis at all. We have included the question, but we refer it to the administration for consideration.

Q 1-4. *[Learning Goals (recommend no more than 4)]*
For each course learning goals presented, the student is asked:
How well did you achieve this learning goal in this course?

    Extremely well, very well, moderately well, slightly well, not well at all

Q 5. About what percent of the class meetings (including discussions) did you attend in person?

    ___%

[Only if instructor indicated there would be online class meetings:]
Q 5a. About what percent of the class meetings did you attend online?

    ___%

Q 6. How much did you learn from this course?

    A great deal, a lot, a moderate amount, a little, nothing

Q 7. Overall, how would you describe the quality of the instruction in this class?

    Excellent, good, fair, poor, very poor

Q 8-12. *[Course Components (recommend no more than 5)]*

How useful to you were the _____? (Blank filled from instructor's specified course components: the lectures, the discussion sections, the reading assignments, etc.)

      Extremely useful, very useful, moderately useful, slightly useful, not useful at all

Q 13. What skills or knowledge did you learn or improve? [Open-Ended]

Q 14. How many hours per week on average did you spend on this course (including class meetings)?

      ___ hours

Q 15. How organized was the class?

      Extremely organized, very organized, moderately organized, slightly organize, not organized at all

Q 16. What would you want another student to know about this class? [Open-Ended] (A summary of replies to this question will be made public; your instructor will be able to view all comments.)

[Q 17. Optional faculty-posed question]

Section Form

Many courses at Stanford include sections, labs or discussion groups, that supplement the primary instruction and that are led by instructors of various ranks, including graduate student teaching assistants and lecturers. These sections are currently evaluated with the "Course and Section Evaluation" form, included in Appendix 3.

While CEC has focused on the main course evaluation form, we believe that the section form also requires similar revisions. It includes 19 discrete items: 11 closed-ended questions and 8 open-ended. In light of the problem of survey fatigue, this form too should be shortened.

All instructors can benefit from student feedback; this is particularly true for graduate student teaching assistants. It is therefore particularly important that departments take TA training very seriously through ongoing mentorship, faculty and peer visits to sections, mid-quarter evaluations and other strategies to support graduate students as teachers. The section evaluation form is just one instrument and is only useful in conjunction with other sources of data. Course professors as well as section leaders should see these results for mentoring purposes.

Shortening the form may reduce the tendency for students to "straight line" their answers. The proposed alternative form below also shifts attention toward student learning while maintaining room for comments on the section leader's performance.

Although we have not submitted the section form to a piloting process, we propose revising it in ways that mirror the changes for the course form.

Q 1. About what percent of the section/lab meetings did you attend?

    ___%

Q 2. How much did you learn from the section meetings?

    A great deal, a lot, a moderate amount, a little, nothing

Q 3. Overall, how would you describe the quality of the instruction in this section?

    Excellent, good, fair, poor, very poor

Q 4-8. *[Section components (recommend no more than 5)]*

    How useful were the _____? (Blank filled from section leader's specified section components: discussions, writing assignments, oral presentations, group projects, problem sets, lab exercises, etc.)

Q 9. What skills or knowledge did you learn or improve in this section? [Open-Ended]

Q 10. How organized was this section?

    Extremely organized, very organized, moderately organized, slightly organized, not organized at all

Q 11. Please comment on the overall performance of the section leader. [Open-Ended]

[Q 12. Optional section leader-posed question]

Just as we recommend that the effectiveness of the main course form be monitored closely and revised where necessary, so too should the effectiveness of this new section form be tracked and modified appropriately.

11. Implementation parameters

The Committee believes that a successful strategy for course evaluation means more than adopting a new form. It means that the university should pay close attention to the process through which the new form is deployed. This process includes multiple dimensions. First, in order to achieve the goal of customizing the form to correspond to the specific character of individual courses, faculty members will be asked to make some choices about the form that students in their class will receive. This point is discussed immediately below. Second, we propose changes to the context in which students fill out

the form. Third, we emphasize that the course evaluation form is only one measurement and, on its own, is insufficient to determine either the success of the instructor or the learning trajectory of the student. To do so requires the development of a robust culture of teaching and learning, especially at the department level.

Customizing the Form and the Pre-Questionnaire

The charge to the Committee emphasizes the importance of tailoring the new form to the specific character of the course. Faculty comments and student focus group results both confirm this imperative. Faculty members want to be able to ask their own questions, and students resent generic questionnaire items that appear irrelevant to the particular course; the mismatch between questions and student experience reportedly discourages students from thoughtful participation, diminishes the quality of the results, and may contribute to a cynicism regarding the evaluation process.

The hard-copy questionnaires that predated the move online included obvious opportunities for instructors to add their own questions to the otherwise standardized form. The current process online has, in fact, the capacity for faculty to add questions of their own, but this function is not visible, and the Registrar's office reports that it is used very rarely.

In order to customize the questionnaire, the Committee proposes that instructors fill out a "pre-questionnaire" as part of the process by which a course is entered into the Registrar's system. A sample pre-questionnaire from our pilot test is in Appendix 12. Just as instructors are currently asked to indicate whether a class is a lecture or seminar, to estimate the size, etc.—in order to facilitate room assignment—they will also be asked to select from a list of potential course components those which are relevant for the specific course evaluation. For example, selections could be made for: lecture, discussion section, lab, guest speakers, field trips, service-learning component, textbook, etc. Students will be asked to evaluate the quality of the selected course components.

Moreover, instructors will be asked to indicate one or several learning goals. These learning goals will appear on the course evaluation form. Students will be asked how successfully each goal was met. In particular, courses that carry "Ways" breadth certification will be expected to include at least one course goal that corresponds to the particular way or ways. How well the course is evaluated as fulfilling that particular breadth area should be communicated to the committee overseeing the requirement.

Instructors will also be invited to add questions of their own.

Despite the customizing capability, we recommend that there be a small number of questions constant across courses to allow for comparisons. Similarly, the overall structure of the form should remain stable across courses, so that students can become familiar with it. Finally, in order to have students respond positively to the evaluation form, close attention has to be paid to the character of the interface, including the aesthetics of an attractive and appealing graphic design.

Two specific types of courses deserve special attention as part of the project to customize the evaluation form.

Team-taught courses are a source of irritation in the current evaluation process because students are asked to fill out full separate forms for each of the instructors. This multiplies the students' workload during the stressful period in which they are preparing for or taking exams. Team-taught course evaluations should instead be tailored so that the few instructor-specific questions can be answered for each of the instructors separately, while the rest of the course questions need to be asked only once.

Lecture-series courses, i.e. courses consisting solely of a series of guest speakers, do not lend themselves well to the current evaluation protocol, and students have expressed disaffection at the mismatch between the course structure and the assumptions of the current form. Such lecture-series courses should be exempted from evaluation on the grounds that they do not genuinely constitute a cohesive course.

Context of the Evaluation

Focus group comments suggest that having students complete the form during the pre-exam study period or during exam week is not optimal, adding to student stress. It is worth remembering that prior to the move online in 2006, students completed the paper form in class during the last week of classes, with instructors setting aside a period of time for this activity. Shifting to the online medium was combined with a shift in the location of the exercise, from in-class to out-of-class, which may have seemed natural in a period before the current wide dissemination of laptops, tablets and smart phones. In today's environment, we can maintain the electronic format, but return the evaluation process to the classroom. Devoting class time would provide an opportunity for the instructor to demonstrate that he/she regards the evaluation process as important.

However, because not all students are in attendance and because not all students may have an appropriate device in class, we propose a long window of time to fill out the form, stretching from the in-class session during the last week of classes until the two days after the end of the examination period. Once a student fills out the form, he/she should have the opportunity to revise it until the deadline.

For large classes, there may be limits of available bandwidth in the classroom, which would have to be rectified. Asking students to complete the evaluation during class also raises issues of access to computers. We propose that the new format be designed so that it can be completed on a smartphone, a tablet or a laptop. Students should be asked to bring a device to class for the purposes of filling out the evaluations. Students who do not bring a device to class should take the opportunity to draft their responses to the open-ended questions. Students who submit their evaluation forms receive early access to their end quarter grades.

In general, asking students to complete, or at least begin to work on, their course evaluations during class is an opportunity for the instructor to underscore the importance of the process and to remind the student that participating in the evaluation represents a

responsibility to other students—by helping to improve instruction and by generating information that will be shared with the student public. The goal is to build a culture of responsibility for learning and partnership with the instructor, and moving the setting of the evaluation into the classroom can underscore these values. As Committee member Caroline Hoxby put it, "You can't really benefit from a great university's education unless you as a student take a lot of responsibility for your learning." Putting the evaluation process back into the classroom—with the additional opportunity to complete it outside of class—underscores this public responsibility.

The CEC proposes that instructors be given an exemplary text to read when they introduce the course evaluation time during the last week of class, although they should be free to choose their own phrasing. The statement should explain:

    a) why the course evaluation is important for improving student learning
    b) who reads the course evaluations and where the results are available
    c) protection of anonymity
    d) student responsibility to the instructor and to other students.

Taking the course evaluation seriously is crucial to the partnership between instructors and students, just as it is an important contribution by the student to the well-being of future students. Completing the course evaluation form should be taken as seriously as the Fundamental Standard.

Course Evaluation and the Culture of Teaching and Learning

The Committee's charge involves designing a new form that provides better information to instructors and that focuses on student learning, while engaging students themselves in the process. The Committee noted that the charge does not refer to a goal of a summative judgment on the quality of an instructor for department or school administrative purposes, such as salary setting or promotion.

Yet for many instructors—especially those without tenure, and even more so for lecturers not on the tenure line—course evaluations can be seen as a potential threat to job security. There is considerable anxiety that poor results will have negative impacts on one's career. Practice on this point varies across schools. In any case, only a very small number of courses receive noticeably low scores; most teaching at Stanford is rated between 4 and 5, i.e., between good and excellent.

Interestingly, some student comments in the focus groups indicated a belief that course evaluations have a strong impact on instructor employment status; some students therefore expressed reluctance to criticize the quality of teaching for fear of doing harm to their instructors. More common, however, is the view the evaluations have no consequence at all, and that even poor teaching will not be corrected. This cynicism is detrimental to the well-being of the intellectual community of the university.

In order for any evaluation instrument to be successful, the university should emphasize that genuine success involves building a process in which the student responses

productively inform the continuing growth of the instructor as a teacher. To make this case requires shifting attention away from the question of the summative judgment; while course evaluation results will surely be consulted in the future with regard to promotion, they are only one piece of evidence, balanced by other, often more important material, such as letters from students, reports from peer visits, and the trajectory of one's teaching repertoire, as evidenced, for example, in the development and quality of syllabi. The main function of the course evaluation lies elsewhere, as a catalyst to improved teaching.

For that catalyst to work effectively, it must be integrated into a wider culture of teaching and learning. Instead of regarding course evaluation results as the private concern of the instructor, we envision department level discussions of the overall quality of teaching. Just as an organized and articulated curriculum is a departmental responsibility, so too is the quality of teaching and learning. Teaching effectiveness, based on the aggregated evaluation scores, should be the topic of departmental discussions, on at least an annual basis. Evaluation scores should be addressed between mentors and assistant professors, although broader deliberations would be productive as well. Departments should similarly sponsor talks or round tables on pedagogy issues relevant to their disciplines. Departments as well as individual faculty should consult with CTL to understand the results of the course evaluations, and CTL should continue to develop appropriate material to advise faculty regarding evaluations.

Making the best use of course evaluation results requires moving beyond an analysis of the individual course—as important as that level is—in order to develop broader perspectives on teaching quality. Faculty should receive reports that reflect the longitudinal development of their teaching effectiveness, while also placing their results in relation to results in similar courses. Meanwhile, we should use computational capacity better, to analyze how different groups of students—majors, non-majors, students with or without certain prerequisites, etc.—respond to certain courses. In order to improve curriculum design—a higher order version of improving individual teaching—a textured understanding of student progress through the departmental curriculum is needed. CTL or another university office should be charged with providing this analytic service to departments.

Carrying out this kind of data analysis, however, should not endanger the confidentiality of student responses, an issue particularly germane in small courses or small programs. Instructors should also be able to review individual results, but without the demographic data, as a further protection of student identity. Demographic analyses of course evaluation results for larger classes or groups of classes should similarly take steps to protect the confidentiality of student responses.

The end-quarter course evaluation is just one element in the large field of evidence concerning student learning. For learning to be successful at the end of a course, however, the committee recommends more frequent use of mid-quarter evaluations, including the Small Group Evaluations (SGEs) provided by CTL. Increasing the number of SGEs may require training additional staff to carry them out. In some focus groups, students advocated for course evaluations to take place during the term, when they might

still contribute to an improvement of their own learning environment; that is one rationale for mandating mid-quarter evaluations. In contrast, an end-quarter evaluation is compelling only on the basis of student altruism for others, an admirable virtue, but less immediate than the self-interest inherent in the possibility of improving a course only halfway over. Therefore, department chairs should encourage faculty to contact CTL to arrange for SGEs or otherwise to undertake mid-quarter evaluations, e.g., the online mid-quarter evaluations CTL also offers. Resources should be made available to support an increase in the number of SGEs, if there is demand.

Department-level analysis of teaching quality should supplement the results of course evaluations with other data, such as questions on the senior survey, the assessment results of seniors completing the major, analyses of course enrollment trajectories and performance by students, and alumni surveys. This will require a considerable coordination of different information sources. By developing a holistic understanding of teaching performance at the department level, the contributions and potential of individual faculty as teachers can be understood more effectively, and successful teachers can be paired with others in order to share their skills. For example, an instructor with lower evaluation scores might be invited to observe the class of a better performing teacher.

Making teaching effectiveness a public topic at the department level raises the important question as to the accessibility of the course evaluation results. Currently, the quantitative results of the course evaluations (but not the qualitative responses to open-ended prompts) are available online for the university public (in the Teaching Center on Axess). Although students are not generally aware that they can view this information, the Committee endorses this public access and wants it to continue. We recommend, however, reporting the distribution of results—x% excellent, y% good, z% fair—rather than solely the mean. Providing the broader range of responses can avoid students focusing on relatively minor differences between single figures assigned to different instructors.

In focus groups, students repeatedly indicated an attraction to CourseRank for, among other reasons, the qualitative comments from other students. It should be remembered that, for decades, the ASSU Course Guide regularly published qualitative comments on individual courses. While Stanford students currently provide qualitative comments on the course evaluation forms, these comments are not made available to anyone aside from the instructor and the administrative ladder. In the interest of candor, one would want to make these comments available as well, but there are legitimate concerns regarding excessively harsh or abusive formulations.

The CEC therefore recommends that some qualitative comments be made available to the university public, either through a curatorial process—a staff member would have to eliminate offensive terminology or outlier comments—or via an algorithmic analysis. In either case, the faculty member and the administrative ladder should receive the full original comments. There may be no more effective step toward countering student cynicism regarding the evaluation process—and the university's commitment to teaching

more generally—than making the qualitative results available to the student public. Students want to read other students' words about courses. In particular, and with the above qualifications, we recommend making public the student responses to the open-ended question: "What would you want another student to know about this class?"

Courses by first-year instructors and new courses, taught for the first time, should be exempt from the mandate to make the numerical or the qualitative results public. The same holds for small courses, with enrollment below 8.

Our investigation clearly demonstrates that a new course evaluation form is needed. However, the success of our proposed new form depends on attention to the wider process of evaluation and on a university commitment to pursue improved teaching and learning in a systematic way. Such a commitment will require support from university leaders at all levels. Yet chairs, deans and vice-provosts have many other concerns, and advocacy for teaching and learning can be lost in the shuffle.

We therefore urge consideration of establishing a unit—or charging an existing one—to provide systematic monitoring of the quality of teaching and efforts to improve it. This responsibility could be met in various ways. It could take the shape of regular reports dedicated to teaching, including annual reports from deans to the provost, as was proposed in 1973 by Dean of Undergraduate Studies James Gibbs to Provost William Miller. Alternatively it could involve establishing a liaison position between VPUE and VPGE to review departments' teaching agenda, in coordination with CTL, and to provide proactive support for teaching improvement.

However the university decides to manage a broad and robust teaching improvement strategy, it is important that the efficacy of the new course evaluation form be monitored closely from the start, that modifications be made where necessary, and that a comprehensive review take place after two years of implementation. In particular, we recommend a focus group process parallel to the one carried out by this committee. It would be particularly desirable to compare those future focus group results with those reported here.

Steps such as these toward building a culture of teaching and learning will provide opportunities for individual faculty members to grow as teachers, and they will demonstrate to the students that departments and the university as a whole take the quality of its teaching seriously. Improving the evaluation of courses should be embedded in a wider agenda that also considers the structure of curricula and promotes the value of teaching within Stanford's research environment. Such a foundation is indispensable to achieve the learning partnership worthy of a great university.

List of Appendices

1. The Provost's Charge/Letter of Invitation (http://goo.gl/FmLiGA)
2. Stanford's Current Course Evaluation Form (http://goo.gl/HNDlva)
3. Stanford's Current Section Evaluation Form (http://goo.gl/vD7UxU)
4. Summary of Input from Stanford Faculty and Instructors (http://goo.gl/LbxS0f)
5. Analysis Showing Insufficiency of Current Form (http://goo.gl/DvMtOv)
6. Summary of Input from Peer Institutions (http://goo.gl/iWKe8p)
7. Dartmouth Suggestions to Faculty on Interpreting Evaluation Results (http://goo.gl/LuSKHK)
8. CTL Suggestions to Instructors on Interpreting Evaluation Results (http://goo.gl/3DS0AK)
9. Literature Review on Student Evaluations of Teaching (http://goo.gl/8DvAfD)
10. Course Evaluation Literature Review (http://goo.gl/W7nSFA)
11. Pilot Results
12. Sample Instructor Pre-Questionnaire for Proposed Course Evaluation (http://goo.gl/gY1GJx)

Online-Only Appendices

13. Graduate Focus Group Results (http://goo.gl/10GhfO)
14. Undergraduate Focus Group Results (http://goo.gl/5u20vt)

Also Referenced:

Study of Undergraduate Education at Stanford University (SUES) Report: (http://www.stanford.edu/dept/undergrad/sues/SUES_Report.pdf)