Faculty Senate Ad Hoc Committee on Teaching Evaluation and SETs

Executive Summary

The Faculty Senate Ad Hoc Committee on Teaching Evaluation and SETs was charged with rethinking the current student evaluation of teaching process. The committee set this charge within the broader context of additional tools for mentoring and measuring teaching. Over the past year and a half, the committee has consulted with AU faculty, administrators, and students; reviewed the literature on evaluation of teaching; and circulated two draft reports.

While the present report reflects the rich community input, there were many points on which it was not possible to reach consensus, including among members of the committee. The present report notes the most critical of these points, though we acknowledge that members of the committee and of the university community at large may find that positions on which they were passionate are not represented in the current document. We are grateful for all of the suggestions and arguments, but it was not possible to include them all.

A central point upon which everyone agreed was that in evaluating teaching, it is critical to balance student input with other factors, as is stipulated in the *Faculty Manual*. Individual teaching units should continue to articulate for themselves how that balance is best achieved.

The major proposals resulting from the committee's work and from our deliberations are these:

- The student evaluation instrument should be renamed SILT (Student Input on Learning and Teaching), reflecting the role of students as contributors to the process of improving pedagogy at AU, rather than as evaluators.
- A new set of questions for the instrument is proposed, with final wording subject to pilot testing.
- The SILTs should be administered online, during class whenever feasible, following the model used in the Washington College of Law. We acknowledge the challenge of obtaining an adequate response rate from courses offered online.
- Usefulness of SILTs can be improved by
    - educating students as to the importance of providing thoughtful input on teaching and learning in the context of individual courses
    - educating faculty on ways of benefitting from student input
    - institutionally providing multiple analyses of data, including comparison with similar types of courses
    - de-emphasizing small and statistically non-significant variations in tabulated results
- In all evaluations of faculty teaching (adjunct, term, and tenure-line) for merit, reappointment, tenure, and promotion units should include measures that go beyond student input, where such measures are established by individual teaching units.

- Use AY 2015-2016 to pilot test both the new questions and the online platform, with full implementation in AY 2016-2017.

Committee Membership

The original committee served in Spring 2014, with some changeover in membership in AY 2014-2015. All members who served on the committee are listed here in alphabetical order:

Tony Ahrens (CAS), Naomi Baron (CTRL), Mark Clark (KSB), Borden Flanagan (SPA), Amanda Frost (WCL), David Kaib (OIR), Phyllis Peres (DAA), Tippi Polo (WCL), Rachel Robinson (SIS), Lenny Steinhorn (SOC), Chris Tudge (CAS), and Elizabeth Worden (CAS).

Report

Evaluation of teaching is both necessary and necessarily imperfect. The task of the Faculty Senate Ad Hoc Committee on Teaching Evaluation and SETs has been to develop improved methods of evaluating teaching given both the necessity of evaluation and the imperfections in current forms of evaluation. We examined current scholarly research on evaluation of teaching. We solicited feedback from the broader AU community: sessions at the 2014 Faculty Retreat, an online forum for faculty, meetings with deans and the Faculty Senate, a survey of undergraduate students, and face-to-face faculty open meetings.

This report, which presents the results of our work, is divided into the following parts:
1) Background on evaluation of teaching
2) Consideration of changing to electronic student evaluation of teaching
3) General SET considerations
4) Going beyond the SET in evaluation of teaching
5) Usage of various forms of evaluation
6) A summary of our recommendations
7) Pilot year and timetable for implementation
8) References cited
9) An appendix of proposed SET questions

Our recommendations are in bold text and collected in section 6.

1) BACKGROUND ON EVALUATION OF TEACHING

To evaluate the effectiveness of teaching, one needs to understand its purpose. However, there is little formal agreement on the purpose of teaching, and purposes will likely vary from professor to professor, group to group (for instance, undergraduate vs. graduate students), culture to culture, and time period to time period. Indeed, many of the most interesting comments we received were, at least in part, about the possible mismatch of assessment tool with that which is being assessed. Moreover, our own reflections on what we value from our time as students also do not map easily onto a small set of readily definable goals.

Consider some definitions of the purpose of teaching. Judith Shapiro, former president of Barnard, presented this description of the purpose of higher education: "You want the inside of your head to be an interesting place to spend the rest of your life" (Delbanco, 2012). Heidegger writes:

> Teaching is more difficult than learning because what teaching calls for is this: to let learn. The real teacher, in fact, lets nothing else be learned than — learning. His conduct,

therefore, often produces the impression that we properly learn nothing from him, if by "learning" we now suddenly understand merely the procurement of useful information. The teacher is far ahead of his apprentices in this alone, that he still has far more to learn than they — he has to learn to let them learn (Heidegger, 1968).

Similarly, when a faculty colleague asked a former student why she/he should continue to be employed, the former student declared that the faculty member should still be employed

[b]ecause four years out from college I remember specific conversations and moments from the one class I took with you that helped shape my career and education choices, more than any other educational experience I had at [that professor's school].

AU's Middle States Self-Study included the following description of the goals of education, drawing from the Association of American Colleges and Universities (AAC&U):

By "liberal education," AAC&U means, "an approach to college learning that empowers individuals and prepares them to deal with complexity, diversity, and change. This approach emphasizes broad knowledge of the wider world (for example, science, culture, and society) as well as in-depth achievement in a specific field of interest. It helps students develop a sense of social responsibility; strong intellectual and practical skills that span all major fields of study, such as communication, analytical, and problem-solving skills; and the demonstrated ability to apply knowledge and skills in real-world settings."

Some of these goals, such as having the inside of one's head be interesting, are more difficult to evaluate than others, such as demonstrating a particular practical skill. Assessment, however, is likely to focus on that which is most easily measured; some faculty may well then shape their behavior to that which is being assessed. Therefore, there is a danger that those goals that are more easily assessed will receive more attention at the expense of those goals that are more difficult to assess, crowding out efforts toward reaching potentially more important goals.

Despite these difficulties, teaching needs to be assessed, for two purposes.

First, assessment of teaching can help to improve teaching. Professors routinely engage in a variety of ways of seeking feedback so as to change how they teach, even in the middle of courses. Facilitating this sort of "formative" evaluation should improve teaching.

Second, assessment of teaching also needs to occur to evaluate the faculty engaged in teaching. In decisions ranging from merit-pay review to reappointment to tenure, teaching *must* be considered, and that cannot happen without some form of "summative" assessment.

If assessment cannot be done perfectly, it can be done better and with a sense of the limitations of those methods used, as is true also in evaluation of scholarship and service – the other two legs of the three-legged stool of academic evaluation. Our committee has several recommendations for how to improve our methods of teaching evaluation, as we describe below.

2) CONSIDERATION OF CHANGING TO ELECTRONIC STUDENT EVALUATION OF TEACHING

We recommend that SETs be done online. This shift should facilitate easy analysis of SET data and easier delivery. We will discuss some of the benefits of online evaluation in section 5, but we note here that the current paper-based system is near collapse, and so we must move to a new system in the very near future. This move affords an opportunity to switch to electronic SETs.

We will now discuss some of the considerations in changing to electronic SETs.

a) Not all forms of electronic evaluation are the same. Some universities that have switched to online SETs have experienced dramatic drops in student-response rates, with as few as 20% of students completing SET forms. Those who do complete the SET at these universities are likely to be atypical, as are those who often complete on-line reviews such as ratemyprofessor.com.

b) **We recommend that for "traditional" face-to-face courses, SETs be completed in the classroom, as has been the case for paper reviews**. For fully online courses, evaluations are of necessity done online. For these courses, one possible recommendation would be that as an incentive, students who complete evaluations would receive earlier access to their course grades than students who do not (assuming such a system is feasible). Other means of boosting response rates should also be explored (see, for example, Berk, 2012 and Jacek, 2015). The committee was of mixed minds about how best to distribute SETs in hybrid courses, but we were united in seeking a process that ensures a robust response rate.

c) **We recommend a pilot project for SETs to be administered exclusively with electronic devices, if possible.** Individual faculty or units may also collect qualitative evaluations in paper if they wish. Electronic administration should be designed to maximize participation. However, we suggest that faculty encourage students to complete the SETs on a laptop instead of a smart phone or other small device. We suggest, too, that there be a 48-hour window after a course's initial administration of the SET in which students can still submit their responses.

d) Some non-traditional courses (e.g., team-taught, laboratory) are not well handled by the current system. We recommend that solutions to SET usage in such courses be examined during

the pilot and implementation phases of the project. At a minimum the new SET system should be flexible enough to accommodate all types of courses taught at AU, allowing all faculty to be effectively evaluated.

e) **We recommend that a joint group of faculty, representatives from OIT, and the Registrar's office manage the electronic system.** This should include development of specifications for the system in light of this report.

3) GENERAL SET CONSIDERATIONS

We propose new SET questions. These will be presented at the end of this report in an appendix. Before considering specific questions we will discuss SETs in general, so as to provide context. Using SETs to evaluate teaching presents a variety of problems. We describe three of these problems here.

> Predictive Value for Subsequent Student Performance
> One recent study (Braga, Paccagnella, & Pelizzari, 2014) suggested that higher SETs were predictive of poorer performance in subsequent classes. Students were, to some degree, randomly assigned to different sections of the same course. Outcome measures were grades in subsequent (also somewhat randomly assigned) sections of linked courses. Students in sections that received better SETs did better in the initial class but less well in subsequent classes. This suggests that faculty who "teach to the test" might have students who value the course-specific progress they make and do better in the short run, but miss the deeper learning achieved by those in classes that are less favorably reviewed (for which the gains might be less immediately visible to students). However, this study examined courses in only three disciplines at one university in Italy, and there were difficulties in random assignment, rendering interpretation of the study difficult.
>
> Potential Sources of Bias
> Student evaluations of teaching are likely affected by a variety of factors that are unrelated to how much students learn. For instance, the largest correlate of student evaluations may be whether students have a prior interest in a course (Wright & Jenkins-Guarnieri, 2012). Thus, it is plausible that professors teaching lower-level required courses might receive lower ratings than those teaching higher-level optional courses even if there are no differences in objective learning. In addition, there may be biases in evaluations based on race, gender or other demographic variables. For instance, one review suggested that "gender appears to operate in interaction with other variables, such as the gender of the rater, the gender-typing of the field in which one teaches, one's gender-typed characteristics, and status cues" (Basow & Martin, 2012). There may also be biases based on actions faculty take that are unrelated to

learning. Indeed, one article, drawing on correlational studies, suggested twenty methods faculty might use to raise SETs without increasing learning (Neath, 1996). **We recommend that there be an examination of potential biases arising from such factors as gender, race, country of origin, and age of the instructor.** This examination could take the form of an individual or committee periodically (perhaps annually?) and systematically evaluating the accumulating SET data at AU for various elements of bias.

Student Estimates of Own Learning
It is difficult to assess what we do and do not know. For instance, sometimes the only way we can understand our own ignorance would require us to not be ignorant. Thus, the least competent are often the most likely to overestimate their performance (Kruger & Dunning, 1999). Students who have learned little might judge themselves to have learned as much as those who have learned a great deal. If students rely on erroneous estimates of what they have learned to make their evaluations, those evaluations will be of dubious validity, as suggested by the Heidegger (1968) quote above.

These are all serious problems, and they do not exhaust the list of problems to be considered when using student evaluations. However, they do not doom student evaluations as a tool. That the evaluations are affected by some things other than learning does not mean that they are uninfluenced by learning. And the most recent review we could find of meta-analyses of the validity of SETs concluded, drawing upon 193 studies, that SETs were in fact valid (Wright & Jenkins-Guarnieri, 2012). Across a large number of studies, courses that were better rated also demonstrated more learning.

In addition, SETs give students an opportunity to provide insights into a course. This is important not only because student input can improve teaching and help identify stronger and weaker teachers, but also because the role of a student calls for the reflection and voice present in the SET exercise. There can be an unfortunate tendency to reduce "students" to "customers." Having students supply input about their learning experience will, at its best, help students reflect on their activities and experience and by doing so enter more deeply into their roles. Having students voice their input serves as a reminder that their role in the scholarly community is not passive but, rather, active.

All these considerations—the drawbacks, validity, and usefulness of SETs—influenced our decisions regarding the SET. In the appendix, you will find those revised questions. We will discuss modified usage of SETs in section 5.

Different teaching units and individuals had different inclinations for which questions they would like to use. We propose that the new SET allow teaching units and individuals to add a small number of customized questions, as is the case with the current SETs.

Significant changes to both questions asked and delivery systems might well have unforeseen consequences. Therefore, **we recommend that the questions and delivery system be pilot tested in AY 2015-2016 in classes taught by a group of senior full professors**. Results of this pilot would be reviewed by the Senate before extending the new questions and delivery system to all of campus in AY 2016-2017.

4) GOING BEYOND THE SET IN EVALUATION OF TEACHING

Effective teaching has many dimensions, so our evaluation system must incorporate multiple means of capturing evidence of teaching effectiveness. In fact, our *Faculty Manual* mandates that we do so.

Presently, all academic units have narratives (listed on the DAA website[1]) identifying the variety of ways in which their own unit looks beyond SETs in evaluating (and hopefully mentoring) good teaching, primarily for promotion and tenure. However, in reviewing these criteria in their present form, as well as drawing upon CTRL's survey of academic unit practices, it was difficult to discern how such multiple criteria are actually applied. In addition, this is of particular concern for term and adjunct faculty, for whom there are few evaluation criteria for reappointment and merit.

While it is important to solicit and consider student feedback (e.g., as represented by SETs) in the educational process, we believe that it is critical for teaching units to develop a more inclusive, transparent, multifaceted assessment of teaching effectiveness. The choice of measures and weight they are given may well vary with type of review (e.g., tenure line versus term faculty, merit review versus promotion or tenure). All faculty (tenure line, term, and adjunct) within each unit should be given the opportunity to participate in this development process, and the results of the process should be widely disseminated.

Examples of "beyond SET" criteria identified across units as well as proposed by our Committee include:
- o Peer observation and mentoring of teaching
- o Teaching portfolio
- o Course syllabi
- o Innovations in teaching, including teaching a new course
- o Publications in/presentations at pedagogical journals/conferences

---

[1] http://www.american.edu/provost/academicaffairs/unit-guidelines.cfm#ten, accessed March 2014.

- Letters from former students and advisees
- Publications by students whose research faculty have supervised (faculty may or may not be co-authors)
- Self-description of course goals and self-evaluation of achievement of course learning objectives
- Examples of feedback provided on student work
- Fulfillment of course and program learning outcomes
- Preparation for advanced courses in the program

As with SETs, each of these has its limitations. For instance, peer observation of classes may capture only a small portion of a colleague's teaching and is likely to be affected by who is reviewing. Different peers might well give diverging feedback (a phenomenon familiar from peer review of scholarly manuscripts or grant applications). A given classroom visit will be difficult to contextualize absent knowledge of what has previously occurred in a course. And peer review on a large scale would be labor intensive. Letters from students will likely favor those with many students or those who encourage their students to write letters. Other items on the list also need to be critically evaluated for both strengths and weaknesses. But such alternatives must be considered to yield a richer, more holistic evaluation than that provided by SETs alone.

The committee was of mixed minds regarding whether to have units set a cap on the amount of weight that SETs have in overall evaluation of faculty (adjunct, term, tenure line) for merit, reappointment, promotion, and tenure. However, we all agreed on the importance of balancing SET information with the other factors (beyond SETs), that each department, school, and college has already identified on the Dean of Academic Affairs website and as mandated in the *Faculty Manual*.

Teaching units should develop materials and training opportunities to support "beyond SETs" options, working with CTRL or other resources as appropriate. The primary function of this support should be to develop and mentor good teaching, rather than to serve as an evaluation metric. Current support systems include:

CTRL and Unit Programs, Conferences, and Workshops
Both CTRL and several of the academic units have events and personnel in place to help support good teaching. We strongly suggest that academic units work to shift their emphasis from asking faculty up for reappointment to "raise their SET scores" and instead nurture an environment of helping them improve their teaching.

Peer Observation of Teaching

Individual units should continue to have the option of whether or not to conduct peer observations. Similarly, units should continue to decide for themselves whether feedback will be used strictly for mentoring purposes or for evaluation purposes as well. However, we urge that the emphasis be put on mentoring, not evaluation. CTRL has prepared a handbook on peer observation and will continue to assist departments and programs in developing peer observation of teaching methods tailored to their particular needs.

Explore Creation of a University-Wide Teaching Mentoring Program

While some parts of the university have initiated programs to mentor faculty teaching (e.g., the pilot CTRL-CAS Partners in Teaching Program; programming in SIS), others have not. We propose that the university explore the usefulness and feasibility of establishing a university-wide cadre of faculty mentors available to faculty (and academic units) who choose to work with them.

5) USAGE OF VARIOUS FORMS OF EVALUATION

To some extent the difficulties with evaluation of teaching arise from the *usage* of particular methods of teaching evaluation rather than the methods themselves. Evaluation shapes behavior, and so, for instance, evaluating teaching by SETs likely changes behavior that will influence SETs. Faculty have been known to say that they have to raise SETs and that they must go to CTRL to do so, rather than with the intent to do a better job of teaching. To the extent that faculty are trying to raise SETs, are they doing so with an opportunity cost of not doing better as teachers?

For all these reasons, evaluation of teaching must change to ensure that faculty do not do the equivalent of "teaching to the test." Such a change may be more important than the specific questions used in any SET instrument.

But deans and others who evaluate teaching only have so much time for evaluation. All full-time faculty are evaluated on an annual basis for merit pay. To do deep analysis of every teacher on an annual basis would consume a huge amount of time.   So for some purposes there is a need for relatively simple evaluation, but one that minimizes distortion arising from the form or schedule/frequency of evaluation.

The difficulties in interpretation of teaching-evaluation information also arise from the challenge of self-assessment. It is difficult for people in general to understand their shortcomings; if understood, they would be corrected (Kruger & Dunning, 1999). As a result, people generally believe that they are above average (Dunning, Heath, & Suls, 2004). For instance, in one study over 90% of faculty considered themselves to be above average compared to others at their

institution (Cross, 1977). This makes conflict between self-evaluation of teaching and evaluation by outside observers both common and fraught. Having ways of dealing with at least some of these problems would be useful.

Here are some suggestions for how to improve the usage of SETs.

a) Develop better comparison groups.
Right now courses are compared within units (e.g., department and college). They are also compared within broad categories of classes (e.g., 100-400 vs. 500- and above). But courses vary within these groupings. For instance, an introductory course with over 100 students is likely to draw a set of students with different commitments than a 400-level seminar with 15 students. Within a unit some courses generally get lower ratings. Chairs then have difficulty recruiting faculty for these courses as faculty do not want to risk receiving lower SETs.

We recommend the development of better, more fine-grained comparison groups. For instance, general education courses that many students take simply to fulfill a requirement might be separated out, as might small graduate seminars that draw students intensely interested in the topic. (Of course, some students take general education courses out of desire, and some students take courses that are not part of general education for reasons other than desire. However, general education courses are disproportionately likely to draw students who at least initially do not have that desire.) We suggest that general education courses for each of the five areas be grouped for purposes of comparison. We also recommend that other groupings be considered (e.g., courses taught by first-year faculty, social-science research methods courses, and required courses) and note that a move to electronic SETs should facilitate this process.

b) Use better statistics to summarize course outcomes.
Right now SET ratings are looked at in terms of averages, or as percentage of ratings in a particular range (e.g., 6 or 7 on a 7-point scale). While use of averages has clear advantages, it also has limits. For instance, in a small class one or two very low ratings might well pull down averages. **We recommend that the statistical output of SETs for each course include a variety of summary statistics.** For instance, medians and deciles yield information not present in means. Standard deviations may also be helpful. Two courses with similar means might differ, with one having ratings clustered tightly about the mean and the other with ratings more widely distributed. Identifying extreme variations in distributions might provide a richer understanding of the evaluations. We also suggest that summary statistical information about SET ratings for department/college/type of course be distributed to those involved in merit-pay review. This should help those who are assigned to courses that typically draw lower SETs, while still minimizing the reporting burden on faculty. Electronic SETs will likely facilitate the use of a more sophisticated statistical approach.

c) De-emphasize the importance of small differences in SETs.
There is faculty concern that small differences in SETs, though not meaningful, might be perceived as meaningful. For instance, some units review SETs for whether they are above or below the mean for the relevant department and college. But it is unclear that, for instance, being above and below a department's average is an important distinction. For example, we know of a case in which a faculty member received an average instructor rating of 6.07 for a large general education course section but was nonetheless below the department average for the instructor rating. As a result faculty, especially term and pre-tenured faculty, worry a great deal about receiving SETs that are even barely below comparison means. This problem also leads some faculty to try to avoid teaching courses that are particularly likely to draw low SETs. Our recommendations in sections 5a and 5b partially address this problem. **We recommend that comparison of SET numbers across courses use a variety of summary statistics.** For instance, they might note how many standard deviations an average SET is from the mean for the relevant type of course and how close to the ceiling.

d) Conduct further study of the role of grades in evaluation of teaching.
Students in some courses rightly receive better grades than students in other courses. For instance, graduate students in very selective programs are likely to do particularly strong work and so receive particularly good grades. But faculty reports suggest that there is pressure to give relatively good grades, regardless of student performance, so as to bring about good SETs. At the same time, some evaluating bodies examine grade distributions as they conduct their own reviews of teaching performance, but these examinations are not done systematically or consistently across or even within units. We recommend further study of the role of grades in evaluation of teaching in order to produce recommendations for more consistent consideration of grades when evaluating teaching.

e) **Train faculty to make a case for themselves in files for action, merit files, etc.**
Because the SET instrument cannot capture all of the details of a course, faculty should fill in those gaps by describing innovations in their teaching, revisions to courses, "experiments" that did or did not succeed, problematic situations and their responses to them, and new pedagogical approaches. Units should create the opportunities for these descriptions by requesting them as part of files for action and merit files and as appendices to FARS reports. Units should also provide guidance to faculty—particularly new term faculty and junior faculty—in effective methods of advocating for themselves as teachers.

f) Identify models of excellent evaluation of teaching.
One of the reasons that SETs are so heavily used by administration is that units sometimes do not provide detailed information about faculty teaching to deans. There are many reasons for the failure to look beyond SETs in the evaluation of faculty teaching. As noted above, it takes time to evaluate, and faculty time is often scarce. It can also be uncomfortable for faculty to give

negative feedback to colleagues with whom they regularly interact, and often, non-SET measures are insufficiently standardized or transparent enough to use in evaluations. It is understandable, then, that units sometimes engage in analyses lacking in detail and constructive negative feedback.

We recommend that deans work with divisions and departments in their units to develop meaningful ways of providing faculty evaluations that reflect the breadth of faculty teaching activity. Several programs on campus, such as the College Writing Program, offer models that can function as examples.

We recommend that processes be developed for educating faculty and administrators about more effective use of SETs. This would include, for instance, methods for comparison of SET results to other similar courses (for instance, particular general education areas, or research methods classes across units, or classes of a certain size).

g) **We recommend creating a separate set of questions that faculty could use for formative purposes around mid-semester.**
Many faculty already engage in mid-semester exercises to obtain student feedback on their courses. It has been suggested that students take more seriously their end-of-semester evaluation role when they perceive faculty listening to, and acting on, a mid-semester evaluation (Jacek, 2015). We recommend developing a set of questions or templates that faculty could draw upon for use around mid-semester, though administration of mid-semester evaluations would remain the faculty member's option. The outcomes of these formative SETs would only be available to the faculty member. As appropriate, faculty members could seek mentoring advice from their colleagues or from CTRL.

h) **Modify FARS to better record and represent teaching effectiveness.**
FARS should be modified to include space at the <u>beginning</u> of the document where faculty members can summarize their most important accomplishments (in teaching, research, and service) for the year. Note that at present, the only open-ended space in FARS (found under "Teaching") limits faculty to talking about pedagogical innovations or student outreach activities "*that are not connected with a specific course.*"

i) Educate all members of the university about the functions of teaching evaluations.
Information about the purpose, components, and uses of American University's SET system of evaluating teaching effectiveness should be shared with students, faculty, and administrators in a systematic and repeated fashion. This process should ensure that
   o   students understand the role SETs play in faculty evaluation, and therefore the importance of responding thoughtfully.

- faculty evaluation committees are aware of unit policies regarding the balance between SETs and other criteria in evaluating teaching, along with the vital role of mentoring.
- new faculty are introduced to the FARS system, including how it is used administratively.
- administrators and/or university-wide committees making final decisions on merit pay, promotion, and tenure remain cognizant of differential academic-unit policies.

j) Rename SETs.
We recommend that the SET questions be renamed, "Student Input on Learning and Teaching (SILTs)". We believe this change better reflects and frames what students are doing. (Note that we will continue to refer to them as SETs for the rest of this report, for ease of reading.)

High-quality teaching is critical to the mission of American University. Therefore, we offer these recommendations to include multiple aspects of teaching effectiveness in our evaluation and support systems for the continued development and practice of high-quality teaching.

6) A SUMMARY OF OUR RECOMMENDATIONS

1. We recommend that SETs be done online.
2. We recommend that for "traditional" face-to-face courses, SETs be completed in the classroom, as has been the case for paper reviews. **We also recommend that the window in which students can complete responses be kept open if response rates are low.**
3. We recommend a pilot project for SETs to be administered exclusively with electronic devices, if possible.
4. We recommend that a joint group of faculty, representatives from OIT, and the Registrar's office manage the electronic system.
5. We recommend that there be an examination of potential biases arising from such factors as gender, race, country of origin, and age of the instructor.
6. We recommend that the questions and delivery system be pilot tested in AY 2015-2016 in classes taught by a group of senior full professors.
7. We recommend that teaching units should develop materials and training opportunities to support "beyond SETs" options, working with CTRL or other resources as appropriate.
8. We recommend the development of better, more fine-grained, comparison groups.
9. We recommend that the statistical output of SETs for each course include a variety of summary statistics.
10. We recommend that comparison of SET numbers across courses use a variety of summary statistics.
11. We recommend training faculty to make a case for themselves in files for action, merit files, etc.

12. We recommend that deans work with divisions and departments in their units to develop meaningful ways of providing faculty evaluations that reflect the breadth of faculty teaching activity.

13. We recommend that processes be developed for educating faculty and administrators about more effective use of SETs.

14. We recommend creating a separate set of questions that faculty could use for formative purposes around mid-semester.

15. We recommend that FARS be modified to better record and represent teaching effectiveness.

16. We recommend that information about the purpose, components, and uses of the SET system of evaluating teaching effectiveness be shared with students, faculty, and administrators.

17. We recommend that the SET questions be renamed as "Student Input on Learning and Teaching (SILTs)."

7) IMPLEMENTATION TIMETABLE

As noted above, the proposals will need to be pilot tested before full implementation.
We suggest the following as the timetable toward full implementation:
Summer 2015: Identify an online vendor.
Fall 2015 or Spring 2016: Pilot the new SET.
Spring 2016: Do focus groups with students and faculty; bring emendations to Faculty Senate if possible.
Summer 2016: Bring emendations of questions and instrument to Faculty Senate Executive Committee.
Fall 2016: Run full-scale emended questions online.

8) REFERENCES CITED

Basow, S. A., & Martin, J. L. (2012). Bias in student evaluations. In M. E. Kite (Ed.) *Effective evaluation of teaching: A guide for faculty and administrators*. Retrievable at the Society for the Teaching of Psychology site: http://www.teachpsych.org/ebooks/evals2012/index.php

Berk, R., A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching and Learning*, *8*, 98-107.

Braga, M, Paccagnella, M, & Pelizzari, M., (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, *41*, 71-88.

Cross, K. P. (1977). Not can, but will college teaching be improved? *New Directions for Higher Education*, *17*, 1-15.

Delbanco, A. (2012). *College: What it was, is, and should be.* Princeton, NJ: Princeton University Press.

Dunning, D., Heath, C., & Suls, J., (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69-106.

Heidegger, M. (1968). What is called thinking? (J. G. Gray, Trans.). New York: Harper & Row.

Jacek, L. (2015). Tactics to increase course evaluation response rates. *College & University (May, 2015), 90(2)*, 13-19.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121-1134.

Neath, I. (1996). How to improve your teaching evaluations without improving your teaching. *Psychological Reports*, *78*, 1363-1372.

Wright, S. L., & Jenkins-Guarneri, M. A. (2012). Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use. *AssessmentandEvaluation in Higher Education*, *37*, 683-699.

9) AN APPENDIX OF PROPOSED SET QUESTIONS

Proposed Scale

Proposed online SILT (5 point scale) =
0 – Not applicable
1 – Definitely not
2 – No, mostly
3 – Somewhat
4 – Yes, mostly
5 – Yes, definitely

QUESTIONS:

I. The Course

1. I am now more knowledgeable about the subject matter of this course.

2. This was an academically challenging course.

3. The readings and other course materials enriched my learning.

4. The graded assignments (such as papers, projects or other required work) enriched my learning.

5. The overall course experience enriched my learning.

6. I worked hard in this course.

II. The Professor and the Learning Environment

7. The professor was well organized.

8. The professor presented and explained the material clearly.

9. The professor treated all students with respect.

10. The professor created a positive learning environment.

11. The professor provided me with the concepts, insights, and/or skills to engage successfully with difficult and complex ideas.

12. The professor was open to student questions and comments.

13. The professor was appropriately available for conversations or questions outside of the designated class time.

14. The professor provided constructive feedback on papers, tests, or other assignments.

Note: We present here two alternative options for piloting.
15a. Given the opportunity, I would take another course with this professor.
15b. I would recommend this professor to other students.

III. Your Engagement with the Course [Note use of different scales]

16. Excluding class time, the average amount of time I spent on the course per week was:
0-2 hours
3-5 hours
6-8 hours
9-11 hours
12 or more hours

17. In terms of my attendance in this class this semester:
I never missed a class session
I missed only one class session
I missed 2 or 3 class sessions
I missed more than 3 class sessions

18. The grades I received on the course assignments and/or exams fairly reflected the quality of my work.
Yes, my grades fairly reflected the quality of my work.
No, my grades were too high given the quality of my work.
No, my grades were too low given the quality of my work.
Not certain.

IV. Background and Reasons for Taking the Course

19. My class level is:

First year
Sophomore
Junior
Senior
Master's
Ph.D.
Other (please specify)

20. The primary reason(s) I took this class was/were (choose all that apply):

I was interested in the subject matter
The course was required for my major or minor
The course satisfied a General Education requirement
The course satisfied a university requirement
The course satisfied a MA, PhD or certificate program requirement
I heard the professor was good
None of the above
Other (please specify)

Open-Ended Responses
(Only the professor will see these responses)

1 What was the best part of this class?

2. What changes to this course do you suggest?

3. What changes to the professor's teaching style or methods do you suggest?

4. Additional comments