TO: Faculty Senate
FR: Beyond SETs Task Force
     Max Friedman, College of Arts and Sciences (co-chair)
     Karen Baehler, School of Public Affairs (co-chair)
     David Banks, School of International Service
     Meg Bentley, College of Arts and Sciences
     Kiho Kim, Center for Teaching, Research and Learning
     Jason Snyder, School of Education
DT: March 26, 2019
RE: Reframing SETs

The Beyond SETs Task Force was established by the Faculty Senate to (1) address concerns about over-reliance on, and inappropriate interpretation of, Student Evaluation of Teaching (SET) surveys at American University; and (2) help teaching units more effectively implement the Faculty Manual's requirement that measures of teaching "extend beyond student evaluations of teaching" (p. 37). In order to fulfill its charge, the task force undertook a literature review, explored current trends in practice at other universities, and assessed three options against goal-based criteria.

Based on that work, we recommend that the provost's office instruct teaching units to adopt in their bylaws a customized portfolio approach to evaluation of teaching on a recurring basis for all faculty in the teaching unit, particularly for purposes of tenure, promotion, multi-year appointment, and award decisions. Portfolios should include measures of peer, student, and self-assessment selected from the menu below or devised by individual units. The interlinked measures that comprise a portfolio are meant to be viewed holistically, with SETs customarily weighted for no more than 50 percent of the portfolio.

Units are discouraged from using algorithms that reduce a portfolio to a single score. "Faculty may demonstrate teaching excellence through a variety of ways," according to the Faculty Manual (p. 37)—portfolios should preserve and highlight that variety.

The task force also suggests that units modify the recommended approach for contexts where it may be less practicable, such as annual merit reviews, one-year reappointments, adjunct contract renewals, and online teaching. Relevant subsets of faculty (such as adjuncts or online instructors) should be involved in designing such modifications. Some units are considering switching to a post-tenure cycle of merit reviews every three years instead of annually, in which case the recommended approach below may also be applicable for merit reviews.

As stated in the Faculty Manual, teaching includes both classroom teaching and engagement with students outside the classroom through mentoring, advising, supervision of capstones and theses, and related activities. Beyond self-assessment, methods for assessing teaching outside the classroom are currently underdeveloped; the task force encourages innovation across campus to fill that gap.

This report incorporates feedback from a town hall attended by 60 faculty members on November 28, 2018, email comments from faculty who could not attend the town hall, and input from the Faculty Senate Executive Committee and the full Faculty Senate at its January, February, and March meetings, 2019.

BACKGROUND

Student evaluations of teaching are imperfect, especially as the primary measure of teaching effectiveness in performance evaluations. Overreliance on SETs fails to provide a full picture of learning in the classroom. Some studies have shown that SETs correlate with student learning; other studies show they do not.

The literature also suggests that bias can be present based on faculty race, gender, age, accent, and other characteristics. No systematic mathematical formula has been proposed to correct for bias because it can be situational: students in gender studies and students in physics may penalize instructors for different apparent characteristics. Bias most affects instructors from underrepresented groups.

Other unwanted factors also can affect SET scores. For example, early morning courses, large class sizes, required courses, difficult courses, new courses, and new methods can all lower scores. Faculty who try to help students learn about racism, sexism, and other forms of discrimination may encounter antipathy, another risk to scores.

Alongside the apparent problems of bias, SETs may have distorting effects on instructors' behavior. Fear of being 'penalized' by students may discourage innovations in the classroom or induce instructors to enforce rules and deadlines less firmly. And research suggests that heavily weighted SETs can create incentives for grade inflation.

On the other hand, AU is an institution where student satisfaction and student voices are valued. Eliminating SETs altogether might produce an unwanted decline in allocation of faculty effort and responsiveness to students, given pressures in other areas. It would also remove a useful feedback tool for course improvement and for identifying faculty who need additional mentoring or support. Many faculty and administrators find that a sustained pattern of very high scores or very low scores is a meaningful and valuable signal. And other forms of evaluation, such as peer observation, are not free of potential bias.

Because the SETs are an imperfect instrument, whose incentives can work for and against the teaching mission and the mission of fostering diversity and inclusion, some universities are adopting new models that decenter SETs. The University of Oregon is switching from numerical to qualitative SETs and faculty reflections on teaching as part of a broader evaluative method. The University of Southern California removed SETs from high-stakes performance evaluation decisions altogether. Instead, faculty receive SET scores and narratives for their own use, and describe in teaching reflection statements how they used student feedback for improvement. Peer classroom observation and internal peer review of course materials, design, and assignments, along with instructors' teaching reflection statements, are used in performance evaluation for tenure and promotion. USC administrators also may consult SETs to identify faculty who do "an outstanding job at engaging students, faculty who may need some support in that area of their teaching, or problematic behaviors in the classroom that require further inquiry."[1] But they do not use them in promotion and tenure decisions. The University of Washington, after faculty of color met with the provost to discuss the problem of bias, adopted a portfolio approach that includes SETs but is more defined than AU's current general instructions in the Faculty Manual (p. 37) that evaluation of teaching performance "must extend beyond student evaluations of teaching."

---

[1] Colleen Flaherty, "Teaching Eval Shake-Up," *Inside Higher Ed,* May 22, 2018.

Peer review is a feature of all of these reformed systems, in keeping with the standard of peer review applied to evaluating a faculty member's scholarship. Reflective practice is another distinctive feature of most reforms, with an emphasis on supporting improvements in teaching and sharing teaching wisdom among colleagues.

OPTIONS

The Beyond SETs Task Force sees three possible options, and recommends the third.

Option 1: Status quo
The Faculty Manual requires that measures of teaching "extend beyond student evaluations of teaching" (p. 37) but does not specify what other measures should be used. In practice, many reviewers focus entirely or primarily on SET scores, and in particular, single-number aggregations of those scores. An effort could be undertaken through messaging to faculty and reviewers to be sure to include additional measures as per the current rules, and to apply appropriate statistical interpretation of scores.

Option 2: Portfolio (as described in Option 3) without SETs
We note that in addition to the potential for bias in SETs, bias based on pedagogical style preference as well as faculty characteristics can enter other mechanisms within the portfolio, such as student and peer observation. The university would need to monitor for this as well as other possible unintended consequences described above and in the table below.

Option 3: Defined portfolio with constrained SET use
This option, in line with reform efforts at other universities, attenuates some of the problems associated with SETs while turning assessment of faculty performance into a cooperative process aimed at continual improvement of teaching rather than a numerically-based, backward-looking test. CTRL would provide training and expert resources to support portfolio-related activities.

> **I. Portfolio**
> When applying for reappointment, tenure, or promotion (tenure line faculty); appointment to a multi-year contract[2] or promotion (term faculty); or teaching awards (tenure line and term), faculty should submit a teaching portfolio containing at least one item from each of the five categories that follow, or a comparable list of items adapted by units as appropriate to their disciplines and stated in their bylaws. Page limits (TBD) are strongly recommended for portfolios.
>
> > A. **Teaching statement** (as currently required in Files for Action - comprehensive narrative): reflect on performance of courses (what worked, what actions to change/improve, etc.) and address achievements, including engagement with students beyond the classroom and any new curricular initiatives
> >
> > B. **Self-assessment of pedagogical activities**
> > > 1. **Annotated syllabus**: describe your design and innovation, especially new courses, revisions

---

[2] A portfolio would be required only for the individual's first multi-year contract and not for multi-year renewals.

2. **Professional development** related to teaching, including CTRL events attended
3. **Examples of feedback** to students such as comments on their work
4. **Written self-evaluation of video** of teaching a class: video need not be submitted
5. **Written self-evaluation of teaching outside the classroom**

C. **Peer (faculty) assessment of teaching**
   1. **Peer classroom observation** and follow up conversation, at least twice before each major review: Reviewers from inside or outside the teaching unit, selected by chair in consultation with faculty member. Use unit-developed template or rubric. (CTRL has examples.) Faculty member writes reflection in response (half-page maximum). (Pilot with senior faculty.)
   2. **Peer observation of classroom video** by one or more colleagues and follow up conversation, at least twice before each major review: Reviewers, templates/rubrics, and response as above. Video need not be submitted.
   3. **Review of course materials** and report by appropriate standing or ad hoc committee at teaching unit level designated by the chair: Materials to include syllabi, course assignments, lecture notes, or other materials chosen by faculty member to convey pedagogical approach and quality.

D. **Student assessment of teaching**
   1. **Student observer committee report:** CTRL and School of Education train a group of students, perhaps drawn from Peer Advisors, to observe classes. Faculty member receives the report and may respond. (Pilot with senior faculty.)
   2. **Focus group** of faculty member's students led by colleague or CTRL instructional staff, followed by report using unit-developed template or rubric. At least twice before each major review. Response as above.
   3. **Narrative portions of SETs:** If any narratives are submitted for a course, all narratives from that course must be included. Reviewers are encouraged to weigh numerical scores in light of any expressions of bias related to physical characteristics and the like. Faculty are encouraged to respond constructively to issues raised in the narrative comments with ideas or steps for addressing valid student concerns.

E. **SET scores** with new constraints and improvements as per II below

II. Constraining and Improving Use of SETs

a. Do not use SETs as the sole or predominant indicator of faculty effectiveness. Teaching units shall develop guidelines for applying integrative and holistic judgment to portfolio reviews, including maximum proportions for SET scores to count within the teaching portfolio, customarily up to a ceiling of 50%. When individual faculty receive very low SET scores consistently over time (the "fire alarm" scenario), other components of the individual's portfolio may provide clues to diagnosing and addressing the problems.

b. Report median scores, frequency distributions, and highest/lowest scores, rather than means (averages).[3]
c. Develop standards for what is a meaningful deviation from the unit and school median scores. Small deviations shall not be held to be significant.
d. Use SETs primarily to assess improvement over time by faculty on same course, not to compare faculty to one another individually.
e. Take into account course characteristics (e.g., disciplinary field, class size, required/elective, lower division/upper division, etc.) when interpreting scores.
f. Note new course preps; use AU innovation exemption policy to drop SETs in advance.
g. Eliminate lowest student score from each course to avoid the potential "one irate student" phenomenon that gives outliers undue influence. (Number of scores dropped may vary by class size.)
h. To increase response rate, return to policy of mandatory use of class time[4]; online evaluation period should begin by default during the penultimate class and extend until the exam time for absent students or those without portable devices, and the evaluation period can be adjusted by faculty member.
i. Provide text for faculty to read aloud or summarize explaining the purpose of SETs (modeled on WCL script, see Appendix).
j. Work with Registrar to delay grade availability to students who do not complete SETs, i.e. those who complete SETs see their submitted grade by date X, other students by date Y. Even a minor delay could increase response rates.[5] Reframe as early access to grades for those who complete SETs, using incentive rather than punishment approach, in part to set a more positive tone at the moment of SET completion.
k. Continue to make SETs available to chairs, deans, and appropriate committees to identify faculty who need additional support or mentoring, as well as for merit and award processes.

III. Adaptation for other processes
As noted, the portfolio would be assembled as part of any major faculty action; it is not meant to be a routine, annual exercise. Therefore, units should develop modified processes for annual merit review, reappointment of term faculty on one-year contracts, and adjunct contract renewals.

---

[3] This is because the scale of potential answers associated with each SET question (the numbers 1-7 or 1-5) represents an ordinal ranking in which any particular number, like 5 or 6, and intervals between numbers, may mean different things to different respondents. By contrast, cardinal numbers are open to less interpretation because they correspond with a real phenomenon and a recognized metric, like age in years, income in dollars, or distance in miles or kilometers. The mean (average) is typically used to measure the midpoint of a series of cardinal numbers, and therefore not appropriate for SETs reports. The median is most appropriate for measuring the midpoint of answers on an ordinal scale.

[4] "Dissatisfied students are more likely to use out-of-class time to complete the assessment" - Mitchell, Ojmarrh, and Morales, "The Effect of Switching to Mandatory Online Course Assessments on Response Rates and Course Ratings," Assessment & Evaluation in Higher Education 43, no. 4 (May 19, 2018): 629–39, 630.

[5] From Johns Hopkins University student evaluations instructions to faculty: "students will also not be able to view their course grades in SIS [student information system] prior to submitting all of their course evaluations; therefore, your cooperation in this regard is important. Do not tell students their grades in your courses. Your students will get their course grades through SIS after all of their course evaluations have been completed. If you post course grades in Blackboard, please be sure that the relevant column in your grade book is 'hidden' so it is not released to students."

ANALYSIS

Table 1 (below) reports the results of the task force's assessment. Based on analysis against the criteria presented in the table, the task force recommends Option 3 to require multiple measures of assessing teaching effectiveness and to constrain the weight applied to SETs in any high-stakes decision. This option will help reduce bias and other unwanted factors in the evaluation process, and better align practice with the expectations stated in the Faculty Manual.

Table 1: Assessing Options against Criteria

CRITERIA	OPTIONS		
	1 Status Quo	2 Portfolio without SETs	3 Portfolio with SETs
Emphasizes importance of teaching quality over popularity		√+	√
Reduces incentives for grade inflation		√+	√
Encourages and enables instructional innovation and continuous improvement of teaching		√	√
Recognizes and seeks to mitigate potential for bias		√	√
Facilitates comparison of uniform performance measures across time periods and between individuals, units, etc.	√		√
Balances SETs with other indicators as inputs to tenure, promotion, reappointment, and award decisions	√		√+
Easily identifies teachers who are struggling so that constructive assistance can be provided	√		√+
Identifies exemplary teachers for purposes of recognition	√	√	√+
Gives students meaningful voice and input into assessment of teaching	√		√+
Requires use of multiple measures to capture multiple dimensions of teaching		√	√+
Allows flexibility for units to choose the best mix of measures to accommodate various modes of teaching	√	√	√+
Builds collegiality around a culture of reflective teaching		√	√
Shifts focus of system away from summative evaluation toward formative feedback		√	√
Imposes reasonable time and $ burdens on units and reviewers of files for action	√+	√	√

Appendix: Washington College of Law student evaluations procedure yielding 80% response rate.

Step 1: Notify students one class in advance that they should bring a WiFi-enabled device to the next class.

Step 2: Reserve approximately 15 minutes to conduct the survey, preferably at the start of class.

Step 3: Have students open MyWCL and locate the appropriate course link on their home page (each faculty member will have an independent link.) Read the Statement to Students:

At the AU Washington College of Law, student evaluations are a critical component in our assessment of faculty members and courses. They affect salaries, promotions and retention. They are instrumental in curricular decisions including the scheduling of courses and the sequencing of courses from one semester to the next. They are reviewed by the Dean, Academic Deans, and faculty committees evaluating full-time and adjunct faculty. They are NOT made available to professors until final grades are submitted and there is no method by which faculty or administrators can identify the author of a response. We ask that you please take time to fill out the evaluation carefully. A separate evaluation link must be completed for each professor if the class you are taking has more than one professor. **Please note that a rating of "5" is positive and "1" is negative.** Responses are not transmitted until the submit button at the bottom of the evaluation is selected. Responses cannot be saved, nor can they be edited, reassigned or retracted after submission.

Step 4: Depart the classroom while students are completing the evaluations.

Step 5: Send an all-class email message later that day or evening asking any students who have not yet completed the evaluation to do so ASAP.

Step 6: Monitor response rates on your MyWCL home page. If response rates appear to be far short of the 80% response goal, prompt students to finalize. Academic Affairs can reopen course evaluations on a subsequent date (prior to exams) if necessary.