



Kum-Kum Bhavnani
Telephone: (510) 987-9303
Email: kum-kum.bhavnani@ucop.edu

*Chair of the Assembly of the Academic Senate
Faculty Representative to the Regents
University of California
1111 Franklin Street, 12th Floor
Oakland, California 94607-5200*

July 24, 2020

ACADEMIC SENATE DIVISION CHAIRS

Re: Report of the Academic Council Teaching Evaluation Task Force

Dear Colleagues:

With Academic Council's endorsement, I am transmitting a report from the Academic Council Teaching Evaluation Task Force. The Task Force included the chairs of CCGA, UCAP, UCAADE, UCEP, and UCFW, and was charged with studying current best practices for student evaluation of faculty teaching, as well as issues of reliability, validity, and bias in evaluations.

The report includes six recommendations largely distilled from an analytical report on teaching evaluations produced by the UC Centers for Teaching and Learning (CTL) in 2019, as well as separate investigations into teaching evaluations conducted by individual UC campuses. The CTL and campus reports are attached as appendices, along with a separate set of reflections from UCAADE.

I ask Senate Division Chairs to forward the report to their Committees on Academic Personnel and other interested parties for review. Please note that Council does not seek to dictate a course of action, but instead wants to increase awareness of issues around teaching evaluations, and encourages campuses to consider and adopt the recommendations best suited to them.

If questions arise, please route them through Council so we can address them and respond accordingly.

Sincerely,

A handwritten signature in cursive script that reads "Kum-Kum Bhavnani".

Kum-Kum Bhavnani, Chair
Academic Council

Cc: Vice Provost Carlson
Senate Directors
Senate Analyst Abrams
Hilary Baxter, Executive Director, Academic Senate



UNIVERSITY COMMITTEE ON ACADEMIC PERSONNEL (UCAP)
John Gilbert, Chair
gilbert@cs.ucsb.edu

Assembly of the Academic Senate
1111 Franklin Street, 12th Floor
Oakland, CA 94607-5200
Phone: (510) 987-9466
Fax: (510) 763-0309

July 9, 2020

KUM-KUM BHAVNANI, CHAIR
ACADEMIC COUNCIL

RE: REPORT FROM THE 2019-2020 TEACHING EVALUATION TASK FORCE

Dear Kum-Kum,

The Teaching Evaluation Task Force established by Academic Council in March 2019 was charged with studying the current best practices for teaching and teaching evaluation, and making recommendations to Council for how to bring these into alignment with UC's personnel process. In July 2019, the chairs of CCGA, UCAP, UCAADE, UCEP and UCFW submitted a proposal to Council for an in-depth study of the reliability, validity, and bias in student evaluations. We soon discovered that a systemwide UC Center for Teaching and Learning (CTL) group was finalizing a report on its analysis of teaching evaluations at UC campuses. Upon seeing that the CTL's effort was clearly aligned with the Task Force charge, we agreed to capitalize on the significant work the Centers had already accomplished. The Task Force also learned about similar, independent assessments by the UCD, UCI, UCLA, UCSD, and UCSB divisional senates which are still at various stages of completion.

The Task Force convened its first meeting in November 2019 during which Linda Adler-Kassner, Faculty Director, Center for Innovative Teaching, Research and Learning and Associate Dean, Undergraduate Education, UCSB, discussed the highlights of the November 2019 "UC Teaching and Learning Centers Recommendations for Evaluating Teaching Effectiveness" report (Appendix 1). Members found that the recommendations put forward by the CTLs are largely consistent with those outlined in the five divisional Senate reports (the UCD and UCSD reports are public and are provided in Appendices 2 and 3 as examples). We shared the reports with our respective committees and received formal feedback from UCAADE (Appendix 4) and informal suggestions from UCEP and UCAP.

Respecting that approaches to evaluating teaching should be unique to each UC campus and recognizing the work in progress throughout the system, the Task Force decided to forward the following overarching recommendations to Council for endorsement:

Recommendation 1: The relevant divisional Senate committees should carefully consider the “UC Teaching and Learning Centers Recommendations for Evaluating Teaching Effectiveness” report and adopt the recommendations best suited to their campuses.

Recommendation 2: Campuses should adopt the broadest language possible to more fully capture the kinds of bias that may be systemic in teaching evaluations.

Recommendation 3: Divisional Committees on Academic Personnel should be educated about the systemic problems in student course evaluations including inappropriate comments and bias in the scores for women and faculty from underrepresented groups.

Recommendation 4: Campuses should adopt strategies such as assessing actual learning outcomes to contextualize student evaluations in challenging gateway courses, perhaps by examining how students perform in the next course in the sequence.

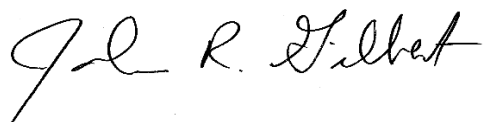
Recommendation 5: Campuses should identify robust, additional forms of evaluation, such as incorporating peer observations and reflective teaching statements for use in assessment of faculty.

Recommendation 6: UCAP and UCAADE should continue to monitor the ongoing campus activities related to improving the effectiveness and value of teaching evaluations.

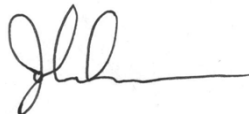
The Task Force is deeply appreciative of the fundamental contribution made by UC’s Centers for Teaching and Learning, and we are encouraged by the attention to this valuable aspect of the personnel process.

Please don’t hesitate to contact us if you have any questions.

Sincerely,



John Gilbert, Chair
UCAP



John Serences, Chair
UCEP



Ramesh Balasubramaniam, Chair
CCGA



Mona Lynch, Chair
UCAADE



Jean-Daniel Saphores, Chair
UCFW

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

UC Teaching and Learning Centers Recommendations for Evaluating Teaching Effectiveness November 1, 2019

[Executive Summary](#)

[Global Recommendations](#)

[Approaches to Evaluating Teaching Effectiveness](#)

[Student Evaluations of Teaching: Common Challenges and Recommendations](#)

[What questions to ask?](#)

[How to report the data?](#)

[How to get adequate response rates?](#)

[How to get useful feedback?](#)

[Student bias in teaching evaluations](#)

[How to promote reflection on SETs for the purpose of improving teaching?](#)

[Instructor Self-assessment: Common challenges and Recommendations](#)

[Peer Review of Teaching: Common Challenges and Recommendations](#)

[Research-based Teaching: Common Challenges and Recommendations](#)

[Towards Holistic Evaluation of Teaching: A Developmental Approach Using Multiple Measures](#)

[The Role of UC Teaching and Learning Centers in Improving Teaching Evaluation](#)

[Authors](#)

[References](#)

[Appendix A: “Documenting Teaching Effectiveness” Survey](#)

[Appendix B: Summary of Recommendations](#)

Executive Summary

Evaluation of teaching effectiveness is a challenging endeavor, yet it is compulsory for every merit and promotion case involving faculty and lecturers across the UC system. This report, by the systemwide Teaching and Learning Centers (TLC) group, synthesizes efforts that are underway on nearly every UC campus to improve teaching evaluation. The information was gathered through a UC-wide survey ([Appendix A](#)) and extensive follow-up conversations with members of campus-based Committees on Academic Personnel,

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

teaching evaluation specialists, and numerous campus faculty leaders involved with evaluating teaching effectiveness.

Our report highlights innovative practices from the UC system and beyond and provides both global and specific recommendations for challenges associated with teaching evaluation that are pervasive across the UC system. An overview of all of the recommendations contained in this report is provided in [Appendix B](#).

Global Recommendations

1. One of the strongest global themes from our work is the need for better transparency about how teaching is evaluated. Common concerns among faculty were: What forms of evidence do Senate reviewing agencies such as Committees or Councils on Academic Personnel consider to be credible? What constitutes “good or satisfactory” teaching? How acceptable is it for instructors to experiment with new teaching approaches that may result in less favorable student evaluations of teaching?¹
2. Campuses should rely on multiple methods to evaluate teaching effectiveness, balancing student evaluation of teaching (SET) scores with other measures. UCOP’s APM 210-1-d states, “More than one kind of evidence [of the candidate’s teaching effectiveness] shall accompany each review file,” yet numerical student evaluations of teaching are frequently the only form of evidence required or provided for merit and promotion cases across the UC. [UCSC](#) and [UCI](#) are actively drawing attention to the institutionalized requirement for two forms of evidence. Additional sources might include peer observations of teaching, self-assessment, or engaging in research on teaching (e.g., the Scholarship of Teaching and Learning [SoTL] or Discipline-based Educational Research [DBER]).
3. The widely-documented bias in student evaluations of teaching makes them unreliable as the primary form of evidence of teaching efficacy. We offer a series of specific recommendations for making SETs more effective in the section called “[Student Evaluations of Teaching: Common Challenges and Recommendations](#).”
4. It would be beneficial for UCs to adopt a more developmental approach to teaching in the UC, since new faculty hires need time to develop their pedagogy over their path to tenure. Such an approach creates opportunities for faculty to experiment and innovate in their courses without fear of “failing,” rewards teaching professional development, encourages critical reflection on practice, and recognizes faculty commitment to teaching/student learning.

¹ Most UC campuses are working toward providing more clear information on the types of acceptable evidence (for example, see [UCB](#), [UCLA](#) and [UCSC](#) pages).

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

5. While a shift from overreliance on SET to a more holistic evaluation of teaching presents workload implications, public research universities around the country are successfully making this transition. If UCs want to evaluate teaching effectively, we need to be willing to spend more time evaluating teaching effectiveness and expect instructors to spend more of their (finite) time improving their teaching.

Approaches to Evaluating Teaching Effectiveness

Brookfield (1998) outlines four “lenses” through which teaching can be evaluated: the perspective of students, self-reflection, colleague/peer evaluation, and research on teaching. We summarize the challenges associated with each of these approaches and offer recommendations. We then consider more holistic approaches to evaluating teaching.

Student Evaluations of Teaching: Common Challenges and Recommendations

Much has been written about what student evaluations of teaching measure and relationships between students’ evaluations of teaching and their learning. There is disagreement in the literature about whether student evaluations can measure aspects of teaching effectively (Benton and Ryalls, 2016; Boring, Ottoboni, and Stark 2016; Hornstein, 2017; Uttl, White, and Gonzalez 2017); even among researchers who argue that they can measure some aspects, there is consensus that SETs cannot measure important aspects such as curricular alignment or the instructor’s content expertise. In the following sections, we provide recommendations for how to get more useful feedback from student evaluations of teaching.

What questions to ask?

Most of the UCs have attempted to design SET questions that can be used across all teaching contexts to facilitate the comparison of SET results for merit and promotion cases. Yet, teaching varies widely by the size and type of course, the discipline, level, location in the curriculum, and by the students who enroll. These variations make it difficult (and perhaps undesirable) to ask the same questions of all students or to directly compare instructor’s results (Stark and Freishtat, 2014). Based on our review of the literature on writing effective SET questions, we recommend the following:

Recommendations

- Campuses should clarify what aspects of teaching SET questions are attempting to evaluate. Questions that students are well-positioned to answer, for instance, include the extent to which an instructor challenges students; the degree to which students are satisfied with an instructor’s presentation; the extent to which students feel respected by instructors; the amount that students feel that they have learned. UCSB has a [framework for categorizing questions](#) that is used to facilitate the design of SET questionnaires that encompass multiple aspects of teaching.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

- Specific questions that prompt students to consider different aspects of a course and its instruction should precede more global questions, with recognition that questions asking students to evaluate the course or its instructor, “overall,” tend to be less reliable (Worthington, 2002) than specific questions.
- SETs should include open-ended questions that ask students to describe their experience in the course, what aspects facilitated their learning, and how it could be improved. Responses can provide useful, actionable feedback (Stark and Freishtat, 2014).
- If quantitative measures (such as Likert scales) are used for response items, they should be associated with questions about students’ experience of learning rather than asking questions outside of their experience. For instance: “How much would you say you have learned in this course?” is a question that focuses on students’ own experiences; “How would you rate this course/instructor” is less focused on their experience. (Becker, Bosshardt & Watts, 2012).

How to report the data?

UC campuses vary in the way that they report SET data. Three important considerations raised by respondents are 1) whether to average student scores on Likert-scale items, 2) what survey statistics to report, and 3) whether or not to provide comparative data in SET reports. The consensus is that faculty should receive clear information on how SET data are used in the evaluation process, and faculty and reviewing agencies should have consistent frames for interpreting data. This may require campuses to create shared trainings on how to interpret SET data.

Recommendations

- Members of reviewing agencies CAP members should be made aware of the widely documented bias in SET evaluations, particularly toward women and people of color (see [Student bias in teaching evaluations](#)), when considering comparative data.
- SETs should not average ratings to produce a quantitative ‘mean score.’ Instead, SETs should report the distribution of scores, the number of responses received, and response rates (Stark and Freishtat 2014).
- Campuses should not compare one instructor’s scores to those of another instructor, or to departmental averages. If comparative data are reported, reviewing agencies should have written documentation outlining how comparisons will be used (for example, see [UCSB’s Guide to Interpreting ESCI evaluations](#)) and their limitations. Reviewers might instead focus on patterns in student feedback provided via SETs over time, rather than as individual instances. (See American Sociological Association 2019.)

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

- Beran, et al. (2005, 2007) recommend training faculty on analyzing and interpreting student feedback. UC Teaching and Learning Centers can provide this support, as well as recommendations for addressing patterns of concern.

How to get adequate response rates?

Every UC campus has either switched or is evaluating switching from paper-based, in-class evaluations to online course evaluations. However, response rates can drop dramatically when SET is conducted online and is non-compulsory. In one study, rates went from an average of 70% for paper-based, in-class evaluations to 29% for online evaluations (Dommeyer et al., 2004). Nutly (2008) outlines what SET response rate is needed to allow reliable statistical inferences about the larger student populations' experience of teaching for various class sizes, with different confidence levels. Oftentimes, actual student response rates are far below these thresholds. Below are recommendations to promote higher response rates.

Recommendations

- Faculty should allocate class time for students to complete online SET evaluations (ideally at the beginning of class) on their mobile device. This signals the importance of the evaluations.
- Students should receive explanations about the importance of SET feedback for improving the course at various points in the term (Ballantyne, 2003; Hoel and Dahl, 2018).
- Students should receive both automated and personal electronic reminders about SETs, as well as verbal reminders during class (Adams, 2012; Berk, 2012).
- Campuses might offer incentives for students to complete SET surveys (e.g. early grade access, priority registration; Crews, et al., 2011; Hoel and Dahl, 2018).

How to get useful feedback?

When students provide constructive and concrete feedback, faculty find student comments helpful in improving instruction (Braskamp, et al., 1981). Providing faculty and students with resources to help students give meaningful feedback can minimize the potential for bias and reduce vague, irrelevant, or harmful comments in SETs.

Recommendations

- Use open-ended questions that invite feedback on specific aspects of teaching (e.g. What teaching methods used in this course helped you learn?), rather than general feedback about the instructor (e.g. What did you like about this instructor?).

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

- Provide faculty and students with information and resources about the importance of student feedback, as well as examples of unhelpful vs. helpful feedback (Svinicki, 2001). See for example [UC Merced's student peer education and faculty resources](#), and [UC Santa Cruz's information sheet](#) on giving useful feedback.
- Identify multiple opportunities to educate students about the importance of their feedback (when provided thoughtfully and respectfully). For example, UCLA orients students to the importance of SETs during Bruin Day, an outreach event to welcome prospective admits and their families. Students are taught that feedback on courses, instructors, and curricula is vital to documenting teaching effectiveness and better informed instruction and, in turn, leads to enriched student learning experiences.

Student bias in teaching evaluations

Numerous studies going back 40 years and continuing today have established widespread bias in SETs (for recent reviews of the literature, see Basow, et al., 2013; and Spooen, et al., 2015). Most scholarly attention within this literature has been paid to gender bias, and specifically how gender influences students' expectations for, and criticisms of, teaching performance. One leading study showed that when two instructors in an online course disguised their gender, with each instructor operating under two different gender identities, students rated the "male" identity significantly higher than the "female" identity, regardless of the instructor's actual gender (MacNell, et al., 2015). Such role expectations have special relevance in particular classroom contexts. Female instructors face less bias in small classes where individual interaction with students is the norm but are at a disadvantage in larger, more impersonal classes. Bias based on race and ethnicity has also been identified in the literature (Anderson and Smith, 2005; Bavishi, et al., 2010; Smith and Hawkins, 2011;). Overall, the research suggests that bias is pervasive and that female instructors and faculty of color are significantly disadvantaged in academic personnel reviews by these surveys, particularly when institutions rely heavily on quantitative measures to assess faculty performance.

Recommendations

While no single step, or even combination of steps, is likely to eliminate bias from the results of SETs, campuses can enact measures to lessen bias or the effects of bias.

- Eliminate heavily gendered criteria such as "helpfulness" and "warmth" from survey questions.
- Change the name from Student Evaluations to Student *Experience* of Teaching Surveys (UCSC has made this change; UCD's [2019 final report on teaching evaluation](#) recommends the change).
- Rewrite SET questions to reflect student experiences in the classroom ("the instructor defined important concepts") rather than absolute value ("grading was fair") or excessively subjective criteria ("work was returned in a timely manner").
- Shift from numerical scales to frequency scales or scales of agreement.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

- Talk to students about bias in SETs. Peterson, et al., 2019, found that including language about bias in the SET preface was successful in mitigating gender bias.
- Institute a process for faculty who wish to respond to discriminatory comments (for example, see [McGill University's Protocol for addressing hateful or discriminatory comments](#)).
- Perhaps most importantly, dilute the importance of SETs in the academic personnel review process by enforcing the requirement for other measures of teaching effectiveness.

How to promote reflection on SETs for the purpose of improving teaching?

One of the criticisms of SETs is that they are most often used as a summative metric to evaluate teaching, rather than a formative tool to help instructors improve their courses over time.

All but one UC campus reported that there was no “systematic” way that the campus encouraged instructors to review/reflect on SET results and seek help from a Teaching and Learning Center to improve their teaching. At the same time, these visits seemed largely to be motivated once a faculty received low SET scores. At that time, contact with a Teaching and Learning Center was initiated by the instructor themselves, the department chair, or a “letter of admonishment” from a Senate reviewing agency that included a recommendation that faculty work with the Teaching and Learning Center. Such framing of SETs positions them as summative assessments used to validate (or punish) teaching, rather than formative documents that can contribute to teaching improvement.

Recommendations

- Provide instructors with questions that they can use to reflect on all SET results (not just “low” or “problematic” results).
- Provide instructors with the opportunity to discuss their SET results with a teaching/learning specialist and with resources to help them overcome challenges brought to light by students’ feedback.
- In addition to end-of-quarter SETs, encourage instructors to use mid-quarter student evaluations of teaching (e.g. [UCLA's CEILS Mid-Course Evaluation](#)), which allow faculty to respond to students’ feedback before the end of the quarter/semester (Adams, 2012).
- Provide guidance on how instructors can use self-assessment to critically reflect upon, and respond to, patterns of student feedback (e.g. [UCSC's Guide to Using SET Data](#)).

Instructor Self-assessment: Common challenges and Recommendations

Currently, instructors’ self-assessments of teaching for merit and promotion (when provided) range from providing a recap of their SETs, to a summary of courses taught, to providing a reflection on their teaching development.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Recommendations

- If submitted, instructors' self-assessment should provide a reflection on their teaching *development* (focusing on growth rather than attainment). It may be important for instructors to describe their teaching context, philosophy and their insights/planned actions in response to SET feedback.
- Campuses should clearly communicate what is recommended for the teaching self-assessment, and provide examples (e.g. <http://dtei.uci.edu/the-reflective-teaching-statement/>).

Peer Review of Teaching: Common Challenges and Recommendations

Peer review of teaching (PRT), which involves seeking formative feedback from an informed colleague for the purposes of improving one's practice and/or evaluating a peer's teaching for summative purposes such as tenure and promotion, has increasingly become a common practice in higher education (Chism, 2007). Peer review of teaching is often associated with being observed while teaching a class (Bernstein, et al., 1996); however, it can also include a review of course documents such as syllabi, handouts, assignments, or other related material. The literature on PRT outlines many advantages (Seldin, 2006) and suggests that it can be especially helpful in providing formative feedback on aspects of teaching that students are not equipped to evaluate (Keig, 2000). To alleviate common fears of unequal power dynamics between reviewer and reviewee and of PRT processes being punitive, peer feedback is most helpful when used for formative purposes.

There are two prominent challenges associated with peer review of teaching: 1) Departments need a sustainable and reliable process with agreed-upon criteria appropriate to teaching in the discipline that is also seen as valid and valuable by reviewing agencies; and 2) Faculty need guidance and training to be effective evaluators.

Recommendations

- Develop structured processes for peer review (if desired). The UC Centers for Teaching and Learning can help provide resources, training, and support to departments and instructors as they develop and maintain peer review processes. While there are several potential approaches to PRT, there is common agreement that the process be faculty driven, transparent, and fair, equitable, and authentic, which may be accomplished in part by using instruments that are valid and reliable. See [UC Berkeley's Guide to Peer Review and associated forms](#), as well as [UC Irvine's Peer Evaluation of Teaching Guide](#), [UC Davis' web-based observation tool](#), and external to UC exemplars such as the [University of Oregon's holistic framework for peer review](#).

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

- For campuses new to PRT, consider developing PRT pilots in interested departments. See, for example, [UC San Diego's consultation and support services for establishing a PRT process](#) that include: 1) Guidelines for establishing a formal PRT process; 2) Customizable instruments informed by evidence-based teaching and teaching evaluation practices 3) Workshops on conducting effective classroom observations; 4) Evaluation tools to assess impact of PRT process.
- Ensure that reviewing agencies value the benefits of PRT as part of a holistic review of teaching. When done well, a process for formative peer review of teaching can help improve teaching and inform summative decisions.

Research-based Teaching: Common Challenges and Recommendations

As faculty members in research universities, UC faculty are well-positioned to engage in a research-based approach to teaching, to conduct research on teaching, and to share their findings with colleagues.

Recommendations

- UCs can support opportunities for faculty to *experiment* and *innovate* in courses using a *research-based approach to teaching*, whereby instructors innovate, gather feedback/data, reflect and re-design. To encourage experimentation, reviewing agencies can acknowledge that student evaluations of teaching (SET) may be less favorable during the initial stages of teaching innovation, allowing instructors to experiment and innovate without fear of “failing” (Gillman, et al., 2018).
 - Examples of UC-based programs that foster research-based pedagogies include:
 - [UCSC's Digital Instruction Project](#)
 - [UCSB's Instructional Improvement Grants](#)
 - [UCLA's Transforming Teaching in the LS Core](#) and
 - [UCD's Scholarship of Teaching and Learning Conference](#).
- Teaching Professors (L(P)SOE) and Unit 18 lecturers, in particular, should be encouraged to generate new research and publications on teaching and to share their findings with colleagues using systematic approaches like the Scholarship of Teaching and Learning (SoTL) and Discipline-Based Educational Research (DBER). UCs can support this by providing resources for professional development, such as funds to attend pedagogical conferences and course release to engage in significant pedagogical innovation and redesign.
- Teaching professors and Unit 18 lecturers should be encouraged to share their teaching insights and innovations by considering in merit and promotion cases: invited talks, teaching presentations, publications in campus news outlets (newspapers, blogs and vlogs), record of undergraduate and graduate mentorship, and service on campus committees related to teaching, learning and outreach.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Towards Holistic Evaluation of Teaching: A Developmental Approach Using Multiple Measures

Holistic evaluation of teaching considers *multiple sources* of evidence for teaching efficacy and values faculty *commitment* to teaching excellence. It is a *developmental* approach that promotes research-based teaching approaches and engagement in teaching professional development. The overarching purpose of holistic evaluation is to improve teaching.

Holistic evaluation encourages reflection on practice, for example, through faculty assembling a teaching portfolio containing a narrative about their teaching trajectory, with artifacts to illustrate their growth.

Some specific examples of innovative work toward holistic evaluation of teaching include:

- [UC San Diego's Academic Senate "Holistic Teaching Evaluation Workgroup report,"](#) September, 2019, recommends, "The adoption of a holistic teaching portfolio and accompanying instructor self-reflection as the primary means of evaluating teaching in academic files" as one of its four core recommendations.
- UC Irvine encourages the use of [Weiman's rubric](#) for research-based course design.
- UC Davis is developing a HHMI-funded teaching portfolio tool, the "[Multidimensional Instructional Development for Achievement and Success \(MIDAS\)](#)" and the "[Know Your Students \(KYS\)](#)" web based application to improve inclusive instruction across STEM courses."
- The University of Oregon re-envisioned teaching evaluation as "[The Continuous Improvement and Evaluation of Teaching System](#)" and has developed extensive supporting documentation and resources to support it.
- Kansas University's NSF-funded "[Benchmarks Project](#)," uses a holistic, rubric-based approach to examine seven dimensions of teaching practice.
- Numerous campuses nationwide and an entire university system abroad have already moved toward a more holistic evaluation of teaching using [The Career Framework for University Teaching](#).

Recommendations:

- Recognize engagement in teaching professional development activities (e.g. consultation and training by campus-based Teaching and Learning Centers, professional development through NSF grants or professional organizations) as evidence of teaching effectiveness.
- Encourage use of teaching portfolios, such as [Weiman's rubric](#) or KU's [Benchmarks Project](#) rubric, to more accurately evaluate teaching and to promote teaching improvement.
- Even as campuses consider movement toward a holistic evaluation, campuses should consider what constitutes APM-mandated "at least two measures of teaching

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

effectiveness” to encourage a minimally more holistic approach to teaching.

- For instance, giving guidance on how to use the personal statement to document changes made to a course or to instructional practice provides opportunity for reflection on teaching, which can be more effective than using the statement to merely repeat what’s in the SETs.
- Campuses could consider offering the option (within departments or potentially in collaboration with a campus Teaching and Learning Center) of peer evaluation.
- Campuses could use a common protocol and possibly a common rubric to add depth to the teaching portion of the review.
- Encouraging teaching mentors for early-career faculty can nudge campuses to a more intentional and developmental approach to teaching development.

As the national pendulum swings increasingly toward educational accountability, and as faculty are being asked to prioritize student success alongside their research and service loads, shifting from a culture of single measured evaluation to one of supported and holistic teaching development will benefit faculty and students alike.

The Role of UC Teaching and Learning Centers in Improving Teaching Evaluation

The UC Teaching and Learning Centers are well-positioned to support the recommendations outlined in this report. With guidance from campus-based CAPs, the Teaching and Learning Centers can leverage their expertise to:

- Support the campus and departments in improving student evaluations of teaching;
- Facilitate the development of templates and protocols for peer observation of teaching;
- Provide professional development opportunities for faculty to improve their teaching and learning;
- Support Teaching Professors and Unit 18 lecturers in conducting research on teaching and learning (e.g. SoTI/DBER) and facilitate the sharing of their work on campus;
- Support CAP in how to utilize non-traditional evaluation instruments;
- Work with campus-based CAP and UCOP to update the policies for teaching evaluation.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Authors

- Lisa Berry, Sr. Instructional Consultant, Instructional Development, UC Santa Barbara
- Carolyn Sandoval, Associate Director, Teaching & Learning Commons, UC San Diego
- Jody Greene, Associate Vice Provost for Teaching and Learning and Director, Center for Innovations in Teaching and Learning, UC Santa Cruz
- Marc Levis, Director, Center for Educational Assessment, UC Los Angeles
- James Zimmerman, Interim Vice Provost and Dean, Undergraduate Education, UC Merced
- Linda Adler-Kassner, Faculty Director, Center for Innovative Teaching, Research and Learning and Associate Dean, Undergraduate Education, UC Santa Barbara

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

References

- Adams, C.L. (2012). Online measures of student evaluations of instruction. In M. E. Kite (Ed.), *Effective evaluation of teaching: A guide for faculty and administrators*. Retrieved from the Society for the Teaching of Psychology web site: <http://teachpsych.org/ebooks/evals2012/index.php>
- Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences*, 27(2), 184-201.
- Ballantyne, C. (2003). Online evaluations of teaching: An examination of current practice and considerations for the future. *New directions for teaching and learning*, 2003(96), 103-112.
- Basow, S., Codos, S., & Martin, J. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal*, 47(2), 352-363.
- Bavishi, A., Madera, J. M., & Hebl, M. R. (2010). The effect of professor ethnicity and gender on student evaluations: Judged before met. *Journal of Diversity in Higher Education*, 3(4), 245.
- Benton, S. L., & Ryalls, K. R. (2016). Challenging Misconceptions about Student Ratings of Instruction. IDEA Paper# 58. *IDEA Center, Inc.*
- Becker, W. E., Bosshardt, W., & Watts, M. (2012). How departments of economics evaluate teaching. *The Journal of Economic Education*, 43(3), 325-333.
- 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(2), 98-107.
- Bernstein, D. J., Jonson, J., & Smith, K. (2000). An examination of the implementation of peer review of teaching. *New Directions for Teaching and Learning*, 83, 73-86.
- Braskamp, L. A., Ory, J. C., & Pieper, D. M. (1981). Student written comments: Dimensions of instructional quality. *Journal of Educational Psychology*, 73(1), 65-70. doi.org/10.1037/0022-0663.73.1.65
- Brookfield, S. (1998). Critically reflective practice. *J. Contin. Educ. Health Prof.*, 18: 197-205. doi:[10.1002/chp.1340180402](https://doi.org/10.1002/chp.1340180402)
- Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

- Chism, N. V. N., & Chism, G. W. (2007). Peer review of teaching: A sourcebook. Bolton, MA: Anker Pub. Co.
- Crews, T. B. & Dylan F. Curtis, D. F. (2011). Online course evaluations: Faculty perspective and strategies for improved response rates. *Assessment & Evaluation in Higher Education*, 36:7, 865-878, DOI: 10.1080/02602938.2010.493970
- Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29(5), 611-623.
- Gillman, R. A., Hensel, N. H., Salomon, D. A., & Wilhite, S. C. (2018). REFINING THE PARADIGM.
- Hoel, A., & Dahl, T.I. (2018). Why bother? Student motivation to participate in student evaluations of teaching. *Assessment and Evaluation in Higher Education*, DOI: 10.1080/02602938.2018.1511969.
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), 1304016.
- Keig, L. (2000). Formative peer review of teaching: Attitudes of faculty at liberal arts colleges toward colleague assessment. *Journal of Personnel Evaluation in Education*, 14(1), 67–87. doi: 10.1023/A:1008194230542
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303.
- Nutly, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment and Evaluation in Higher Education*, 33(3, June), 301–314.
- Peterson, D. A., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PloS one*, 14(5), e0216241.
- Seldin, P. (2006). *Evaluating faculty performance: A practical guide to assessing teaching, research, and service*. Bolton, MA: Anker Pub. Co.
- Smith, B. P., & Hawkins, B. (2011). Examining student evaluations of Black college faculty: Does race matter?. *Journal of Negro Education*, 80(2).
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642.
- Stark, P. B., & Freishtat, R. (2014). *An evaluation of course evaluations*. doi:10.14293/S2199-1006.1.SQR-EDU.AOFRQA.v1Stark. Retrieved from Science Open:

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

<https://www.scienceopen.com/document/vid/42e6aae5-246b-4900-8015-dc99b467b6e4?0>

Svinici, M. D. (2001). Encouraging your students to give feedback. *New Directions for Teaching and Learning*, 87, 17-24.

"Statement on Student Evaluations of Teaching," American Sociological Association, August 2019

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42.

Worthington, A.C. (2002). The impact of student perceptions and characteristics on teaching evaluations: A case study in finance education, *Assessment and Evaluation in Higher Education*, 27(1), 49–64.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Appendix A: “Documenting Teaching Effectiveness” Survey

Documenting Teaching Effectiveness

CTL Directors Survey, University of California

There has been a nationwide movement, over the past decade, re-evaluating the way that teaching effectiveness is measured and documented in higher education. The UC-wide Directors of Teaching and Learning Centers seek to understand how each UC campus currently measures instructors’ teaching effectiveness, how that information is used in tenure and promotion cases, and what steps are being taken to address challenges associated with traditional approaches, such as student evaluation of teaching (SET), for the purpose of sharing and developing best practices.

The results from this questionnaire will be shared in three stages. First, results will be shared with the Directors of Teaching and Learning Centers across the UC. This group’s primary task will be to refine the presentation of the results so that all campuses benefit from the survey responses in finding legitimate, practical ways to document teaching and gather formative feedback. Second, results will be shared with campuses’ Senates. Campus Senates are primary users of teaching effectiveness data and are best positioned to share and promote “best practices” in teaching evaluation. Third, the results will be shared with system-wide groups, such as system-wide Senate committees, Vice Provosts and Deans of Undergraduate Education, Graduate Deans, and the Provost’s Academic Planning Council. We anticipate that as we move through these stages of result sharing the data will increasingly be presented in aggregate form so as to avoid the perception of evaluating individual campuses. If results are presented in which a campus can be identified, then we will not share the results without that campus’ permission.

We recommend that you read through the set of questions first, as the answers may require input from various campus units, such as a current or past Chair of the Committee on Academic Personnel (CAP), Academic Personnel Office (APO), administrative personnel in Divisions/Departments, Department Chairs, the campus administrators of student evaluations of teaching, etc.

We’re using Qualtrics for this survey, so need to point out a few quirks of that system. Importantly, the **survey can be completed on only one computer**. Survey responses are **saved automatically (there is no Save button)**. **To return to the survey, you must use the same browser and computer that you started it on.**

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

How is teaching effectiveness currently documented, beyond SET?

According APM 210-1-d1, Criteria for Appointment, Promotion, and Appraisal of Teaching, "More than one kind of evidence [emphasis added] shall accompany each review file.... All cases for advancement and promotion normally will include: (a) evaluations and comments solicited from students for most, if not all, courses taught since the candidate's last review..."

Beyond end of course student evaluations of teaching (SET), what other approaches are used on your campus to document teaching effectiveness for CAP tenure and promotion cases?

Self-reflection

- Teaching statement
- Teaching portfolio
- Teaching philosophy statement
- Course syllabi
- Course materials

Student feedback

- Mid-quarter feedback
- Feedback from graduating seniors
- Focus group report

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Colleagues' feedback

- Peer assessments/observations
- Departmental perspective on overall teaching contribution

Scholarship of Teaching and Learning

- Instructor's research on their teaching (published or unpublished)
- Participation in teaching professional development

Please describe any additional approaches that were not listed above.

Please identify any significant challenges associated with these methods and describe how your campus is addressing these challenges (e.g. bias, low response rates, grade inflation).

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

For the boxes that you checked, above, please provide samples or exemplars. **Note:** we realize that these can be difficult to find, but this will likely be the most useful information gathered from this survey.

To upload multiple files, you must first combine the files into a .ZIP file, then upload the single .ZIP file using the button below. Alternatively, you can email the files to lisa_berry@ucsb.edu

Q10 For the boxes that you checked, above, what assistance or training is available to faculty/departments in preparing these materials?

Q24 How do approaches to documenting teaching effectiveness vary across departments/colleges on your campus?

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Q25 How do expectations for documenting teaching effectiveness vary by instructor title (e.g. tenure track faculty vs. lecturers with potential security of employment vs. Unit 18 lecturers).

Q26 Student Evaluations of Teaching

Q27 How are student evaluations of teaching (SETs) conducted on campus?

- On paper (1)
- Online (2)
- Combination of paper and online (3)
- Other (4) _____

Q28 Who is responsible for administration of SETs?

- Center for Teaching and Learning (1)
- Office of Institutional Research (2)

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

- Office of Institutional Assessment (3)
- Office of the Provost (4)
- Executive Vice Chancellor's office (5)
- Student group (6) _____
- Other (7) _____

Q29 What questions are common to all surveys, campus-wide? Please copy/paste questions, below or upload file.

Q30 Click Browse to upload file of questions that are common to all SET surveys, campus-wide.

Q31 If your campus has a standardized SET survey that is different from the questions required on all surveys, above, please copy/paste the survey below, or upload a file with the standard survey questions.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Q32 Click Browse to upload file of your standardized SET survey.

Q33 Departments on some campuses customize their SET surveys. Please comment on the variability in SET surveys on your campus.

Q34 How are results of SETs presented to instructors?

- Results are aggregated by question (1)
- Distributions of student responses are shown for each question (2)
- A single, aggregate numerical score is computed for each question (3)
- Comparative data is provided (4)
- Instructors can view individual student responses (without names) in addition to aggregated responses (5)

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Q35 Please describe any additional features of how SET results are presented to instructors:

Q36 Please upload a sample report showing how instructors see their SET results, and any associated documentation. If you have multiple files to upload, please combine the files into a zip file, then upload the zip file below.

Q37 How are SET results used?

Q38 Who has access to instructors' student evaluations of teaching?

- Department Chair (1)
- Faculty in department (2)
- Students (3)

Q40 How important are SETs, relative to other forms of teaching evidence, in tenure and promotion cases, for

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

	Much more important (1)	A little more important (2)	Equally important (3)	A little less important (4)	Much less important (5)
Tenure track faculty (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lecturers with (potential) security of employment (LSOE/LPSOE) (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unit 18 lecturers (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q41 Comments about the importance of SET, relative to other forms of teaching evidence, in tenure and promotion cases, for tenure track faculty, LSOE/LPSOE and Unit 18 lecturers?

Q42 How are instructors encouraged to review/reflect on their SETs, for the purpose of improving teaching?

**APPENDIX 1 - CENTERS FOR
TEACHING AND LEARNING**

Q43 Please describe any initiatives that are underway on your campus to improve the quality of SETs.

Q44 What next?

Q45 What additional discussions/actions are happening on your campus around improving the documentation of teaching effectiveness that we have not asked about?

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Q46 How can the UC-wide Directors of the Centers of Teaching and Learning support your campus in documenting teaching effectiveness?

End of Block: Default Question Block

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Appendix B: Summary of Recommendations

Global Recommendations

One of the strongest global themes from our work is the need for better transparency about how teaching is evaluated. Common concerns among faculty were: What forms of evidence do Senate reviewing agencies such as Committees or Councils on Academic Personnel consider to be credible? What constitutes “good or satisfactory” teaching? How acceptable is it for instructors to experiment with new teaching approaches that may result in less favorable student evaluations of teaching?

Campuses should rely on multiple methods to evaluate teaching effectiveness, balancing student evaluation of teaching (SET) scores with other measures. UCOP’s APM 210-1-d states, “More than one kind of evidence [of the candidate’s teaching effectiveness] shall accompany each review file,” yet numerical student evaluations of teaching are frequently the only form of evidence required or provided for merit and promotion cases across the UC. [UCSC](#) and [UCI](#) are actively drawing attention to the institutionalized requirement for two forms of evidence. Additional sources might include peer observations of teaching, self-assessment, or engaging in research on teaching (e.g., the Scholarship of Teaching and Learning [SoTL] or Discipline-based Educational Research [DBER]).

The widely-documented bias in student evaluations of teaching makes them unreliable as the primary form of evidence of teaching efficacy. We offer a series of specific recommendations for making SETs more effective in the section called “[Student Evaluations of Teaching: Common Challenges and Recommendations.](#)”

It would be beneficial for UCs to adopt a more developmental approach to teaching in the UC, since new faculty hires need time to develop their pedagogy over their path to tenure. Such an approach creates opportunities for faculty to experiment and innovate in their courses without fear of “failing,” rewards teaching professional development, encourages critical reflection on practice, and recognizes faculty commitment to teaching/student learning.

While a shift from overreliance on SET to a more holistic evaluation of teaching presents workload implications, public research universities around the country are successfully making this transition. If UCs want to evaluate teaching effectively, we need to be willing to spend more time evaluating teaching effectiveness and expect instructors to spend more of their (finite) time improving their teaching.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Student Evaluations of Teaching
<p>What questions to ask?</p>
<p>Campuses should clarify what aspects of teaching SET questions are attempting to evaluate. Questions that students are well-positioned to answer, for instance, include the extent to which an instructor challenges students; the degree to which students are satisfied with an instructor’s presentation; the extent to which students feel respected by instructors; the amount that students feel that they have learned. UCSB has a framework for categorizing questions that is used to facilitate the design of SET questionnaires that encompass multiple aspects of teaching.</p>
<p>Specific questions that prompt students to consider different aspects of a course and its instruction should precede more global questions, with recognition that questions asking students to evaluate the course or its instructor, “overall,” tend to be less reliable (Worthington, 2002) than specific questions.</p>
<p>SETs should include open-ended questions that ask students to describe their experience in the course and how it could be improved. Responses can provide useful, actionable feedback (Stark and Freishtat, 2014).</p>
<p>If quantitative measures (such as Likert scales) are used for response items, they should be associated with questions about students’ experience of learning rather than asking questions outside of their experience. For instance: “How much would you say you have learned in this course?” is a question that focuses on students’ own experiences; “How would you rate this course/instructor” is less focused on their experience. (Becker, Bosshardt & Watts, 2012).</p>
<p>How to report the data?</p>
<p>Members of reviewing agencies CAP members should be made aware of the widely documented bias in SET evaluations, particularly toward women and people of color (see Student bias in teaching evaluations), when considering comparative data.</p>
<p>SETs should not average ratings to produce a quantitative ‘mean score.’ Instead, SETs should report the distribution of scores, the number of responses received, and response rates (Stark and Freishtat 2014).</p>
<p>Campuses should not compare one instructor’s scores to those of another instructor, or to departmental averages. If comparative data are reported, reviewing agencies should have written documentation outlining how comparisons will be used (for example, see UCSB’s Guide to Interpreting ESCI evaluations) and their limitations. Reviewers might instead focus on patterns in student feedback provided via SETs over time, rather than as individual instances. (See American Sociological Association 2019.)</p>

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Beran, et al. (2005, 2007) recommend training faculty on analyzing and interpreting student feedback. UC Teaching and Learning Centers can provide this support, as well as recommendations for addressing patterns of concern.

How to get adequate response rates?

Faculty should allocate class time for students to complete online SET evaluations (ideally at the beginning of class) on their mobile device. This signals the importance of the evaluations.

Students should receive explanations about the importance of SET feedback for improving the course at various points in the term (Ballantyne, 2003; Hoel and Dahl, 2018).

Students should receive both automated and personal electronic reminders about SETs, as well as verbal reminders during class (Adams, 2012; Berk, 2012).

Campuses might offer incentives for students to complete SET surveys (e.g. early grade access, priority registration; Crews, et al., 2011; Hoel and Dahl, 2018).

How to get useful feedback?

Use open-ended questions that invite feedback on specific aspects of teaching (e.g. What teaching methods used in this course helped you learn?), rather than general feedback about the instructor (e.g. What did you like about this instructor?).

Provide faculty and students with information and resources about the importance of student feedback, as well as examples of unhelpful vs. helpful feedback (Svinicki, 2001). See for example [UC Merced's student peer education and faculty resources](#), and [UC Santa Cruz's information sheet](#) on giving useful feedback.

Identify multiple opportunities to educate students about the importance of their feedback (when provided thoughtfully and respectfully). For example, UCLA orients students to the importance of SETs during Bruin Day, an outreach event to welcome prospective admits and their families. Students are taught that feedback on courses, instructors, and curricula is vital to documenting teaching effectiveness and better informed instruction and, in turn, leads to enriched student learning experiences.

Student bias in teaching evaluations

Eliminate heavily gendered criteria such as "helpfulness" and "warmth" from survey questions.

Change the name from Student Evaluations to Student *Experience* of Teaching Surveys (UCSC has made this change; UCD's [2019 final report on teaching evaluation](#) recommends the change).

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Rewrite SET questions to reflect student experiences in the classroom (“the instructor defined important concepts”) rather than absolute value (“grading was fair”) or excessively subjective criteria (“work was returned in a timely manner”).
Shift from numerical scales to frequency scales or scales of agreement.
Talk to students about bias in SETs. Peterson, et al., 2019, found that including language about bias in the SET preface was successful in mitigating gender bias.
Institute a process for faculty who wish to respond to discriminatory comments (for example, see McGill University’s Protocol for addressing hateful or discriminatory comments).
Perhaps most importantly, dilute the importance of SETs in the academic personnel review process by enforcing the requirement for other measures of teaching effectiveness.
How to promote reflection on SETs, for the purpose of improving teaching?
Provide instructors with questions that they can use to reflect on all SET results (not just “low” or “problematic” results).
Provide instructors with the opportunity to discuss their SET results with a teaching/learning specialist and with resources to help them overcome challenges brought to light by students’ feedback.
In addition to end-of-quarter SETs, encourage instructors to use mid-quarter student evaluations of teaching (e.g. UCLA’s CEILS Mid-Course Evaluation), which allow faculty to respond to students’ feedback before the end of the quarter/semester (Adams, 2012).
Provide guidance on how instructors can use self-assessment to critically reflect upon, and respond to, patterns of student feedback (e.g. UCSC’s Guide to Using SET Data).
Instructor Self-assessment
If submitted, instructors’ self-assessment should provide a reflection on their teaching <i>development</i> (focusing on growth rather than attainment). It may be important for instructors to describe their teaching context, philosophy and their insights/planned actions in response to SET feedback.
Campuses should clearly communicate what is recommended for the teaching self-assessment, and provide examples (e.g. http://dtei.uci.edu/the-reflective-teaching-statement/).

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Peer Review of Teaching (PRT)

Develop structured processes for peer review (if desired). The UC Centers for Teaching and Learning can help provide resources, training, and support to departments and instructors as they develop and maintain peer review processes. While there are several potential approaches to PRT, there is common agreement that the process be faculty driven, transparent, and fair, equitable, and authentic, which may be accomplished in part by using instruments that are valid and reliable. See [Appendix C](#) for example materials as well as UC Irvine's [Peer Evaluation of Teaching Guide](#), UC Davis' [web observation tool](#), and external to UC exemplars such as the University of Oregon's holistic framework.

For campuses new to PRT, consider developing PRT pilots in interested departments. See, for example, [UC San Diego's consultation and support services for establishing a PRT process](#) that include: 1) Guidelines for establishing a formal PRT process; 2) Customizable instruments informed by evidence-based teaching and teaching evaluation practices 3) Workshops on conducting effective classroom observations; 4) Evaluation tools to assess impact of PRT process.

Ensure that reviewing agencies value the benefits of PRT as part of a holistic review of teaching. When done well, a process for formative peer review of teaching can help improve teaching and inform summative decisions.

Research-based Teaching

UCs can support opportunities for faculty to *experiment* and *innovate* in courses using a *research-based approach to teaching*, whereby instructors innovate, gather feedback/data, reflect and re-design. To encourage experimentation, reviewing agencies can acknowledge that student evaluations of teaching (SET) may be less favorable during the initial stages of teaching innovation, allowing instructors to experiment and innovate without fear of "failing" (Gillman, et al., 2018).

- Examples of UC-based programs that foster research-based pedagogies include: [UCSC's Digital Instruction Project](#), [UCSB's Instructional Improvement Grants](#), [UCLA's Transforming Teaching in the LS Core](#) and [UCD's Scholarship of Teaching and Learning Conference](#).

Teaching Professors (L(P)SOE) and Unit 18 lecturers, in particular, should be encouraged to generate new research and publications on teaching and to share their findings with colleagues using systematic approaches like the Scholarship of Teaching and Learning (SoTL) and Discipline-Based Educational Research (DBER). UCs can support this by providing resources for professional development, such as funds to attend pedagogical conferences and course release to engage in significant pedagogical innovation and redesign.

APPENDIX 1 - CENTERS FOR TEACHING AND LEARNING

Teaching professors and Unit 18 lecturers should be encouraged to share their teaching insights and innovations by considering in merit and promotion cases: invited talks, teaching presentations, publications in campus news outlets (newspapers, blogs and vlogs), record of undergraduate and graduate mentorship, and service on campus committees related to teaching, learning and outreach.

Towards Holistic Evaluation of Teaching: A Developmental Approach Using Multiple Measures

Recognize engagement in teaching professional development activities (e.g. consultation and trainings by campus-based Teaching and Learning Centers, professional development through NSF grants or professional organizations) as evidence of teaching effectiveness.

Encourage use of teaching portfolios, such as [Weiman's rubric](#) or KU's [Benchmarks Project](#) rubric, to more accurately evaluate teaching and to promote teaching improvement.

Even as campuses consider movement toward a holistic evaluation, campuses should consider what constitutes APM-mandated "at least two measures of teaching effectiveness" to encourage a minimally more holistic approach to teaching.

APPENDIX 2 - UC DAVIS

REPORT OF THE SPECIAL COMMITTEE ON EVALUATION OF TEACHING

JUNE 7, 2019

TABLE OF CONTENTS

EXECUTIVE SUMMARY2

INTRODUCTION AND OVERVIEW3

EVALUATION OF TEACHING AT UC DAVIS5

EVALUATION OF TEACHING AT PEER INSTITUTIONS9

LITERATURE REVIEW12

**RECOMMENDATIONS THAT MAY BE ACTIONABLE IN THE SHORT
TERM.....16**

**RECOMMENDATIONS FOR ACTIONS THAT MAY BE TAKEN OVER THE
LONGER TERM19**

REFERENCES.....22

COMMITTEE MEMBERSHIP.....23

COMMITTEE CHARGE.....24

APPENDIX 2 - UC DAVIS

EXECUTIVE SUMMARY

The Special Committee on Evaluation of Teaching (SCET) was charged with assessing practices for the evaluation of teaching at UC Davis, at peer institutions, and within the published literature, in order to suggest immediately actionable and longer-term potential changes to improve the evaluation of teaching at UC Davis. In ten meetings from January through May 2019 and considerable time outside of meetings, SCET collected, discussed and distilled these sources of information into this comprehensive report.

It is important to note that UC Davis is not the only institution of higher education that is engaged in evaluating its practices for the evaluation of teaching at this time. Many other universities currently are making, or recently have made, minor or major changes in their practices, and indeed, the UC has recently struck a system-wide task force on this topic. The fact that the UC and other universities are engaged in this process does not mean that actions could or should not be taken locally at this time. There were several points of convergence between the findings of published research, innovations at peer institutions and best practices within specific departments and schools at UC Davis that highlighted what changes could be made across UC Davis to improve the evaluation of teaching.

Our recommendations for these short-term and longer-term changes are presented in the final two sections of this report. Four of the most important recommendations are:

1. There needs to be a culture change regarding the perceived purpose of the evaluation of teaching at UC Davis from being primarily or exclusively for *summative* purposes of determining merit and promotion. The evaluation of teaching also can and should serve the *formative* purpose of developing instructors' professional skills and competencies as educators and the *pedagogical* purpose of enhancing teaching effectiveness and students' learning outcomes.
2. To be effective, the evaluation of teaching must be informed from multiple sources, including (but not limited to) student evaluations of teaching (SET), peer evaluations of teaching, and self-assessment of teaching through reflective practices.
3. SET should be recognized and framed as students' *experiences* of their courses rather than *evaluations* of instructors, with the explicit purpose of students' ratings and open-ended feedback being to enhance the quality and effectiveness of instruction.
4. Peer evaluations should be conducted as interactive collaborations between the peer evaluator(s) and an instructor, with meetings prior to and following any in-class observation in order to discuss goals, expectations, performance and recommendations.

APPENDIX 2 - UC DAVIS

INTRODUCTION AND OVERVIEW

In January 2019, the Academic Senate (AS) formed the Special Committee on Evaluation of Teaching (SCET), with the following charge:

“The Special Committee on Evaluation of Teaching will be to (1) evaluate the current practices for evaluation of teaching for merits and promotions at UC Davis; (2) consider practices for evaluation of teaching at peer universities (e.g., wording of items used in SET; methods to improve response rates for SET; alternative ways to evaluate teaching outside of SET such as peer review); and (3) consider research literature on best practices for evaluation of teaching in higher education. The final report should recommend possible near-term minor revisions to the current practices as well as potential long-term major revisions.”

An important distinction should be noted in the title and charge of SCET from the committee struck in 2010, Special Committee on Student Evaluation of Teaching (SCSET). Specifically, “Student” was not in SCET’s title, and the “current practices for evaluation of teaching for merits and promotions at UC Davis” include more than student evaluations. Thus, SCET did not limit its scope to considering students’ reports on their in-classroom experiences of courses. Student reports are an important component of the evaluation of teaching, but they are one component and do not in and of themselves constitute a comprehensive or sufficient means of evaluating university teaching. Peer observations, self-reflection and other procedures also inform the evaluation of teaching.

The SCET members also recognized at least three distinguishable yet complementary goals in the evaluation of teaching. The charge from AS focused on what could be regarded as the summative goal: The collection and presentation of information on which to judge an instructor’s performance as an educator for the purpose of determining appropriate merit and promotion. The evaluation of teaching also can, and we would argue should, serve a formative goal: Informing an instructor of those aspects of course design and delivery that could be improved in order to advance the instructor’s professional development and enhance teaching effectiveness. That attention to teaching effectiveness draws focus to the third and possibly most important goal: Increasing the pedagogical quality of educational experiences for our students such that their learning outcomes are improved.

Paralleling our attention to these three goals was a broadly shared concern for accuracy, fairness and equity in the evaluation of teaching. The common perspective of SCET’s members is that there is widespread dissatisfaction with current practices for the evaluation of teaching at UC Davis because they are seen as susceptible to inaccuracy, unfairness and inequity. Presumably that is why SCET was formed by AS, and it is a theme that we heard from our academic colleagues across the university. There are at least two facets to this concern. First, there is a remarkable degree of variation (or to put it another way, a striking lack of consistency) in evaluation practices across the departments, colleges and schools at UC Davis; naturally this raises questions about whether instructors in some programs are being advantaged by local practices while instructors in other programs are being disadvantaged. Second, demographic, identity and personal characteristics of instructors are seen as affecting the evaluation of teaching in ways that advantage some instructors and disadvantage others; more specifically, implicit

APPENDIX 2 - UC DAVIS

gender, racial and ethnic biases may differentially affect the evaluations of our diverse instructors. Our communications with other institutions and our examination of the academic literature indicate that concerns about gender, racial and ethnic biases are widely-held and may be justified. An essential element of developing “best practices” in the evaluation of teaching at UC Davis is to ensure that said practices improve accuracy, fairness and equity; indeed, the summative, formative and pedagogical goals of the evaluation of teaching cannot be achieved without improving accuracy, fairness and equity in the evaluation of teaching.

SCET’s work should be positioned within state and national contexts. When SCET began its work in January 2019, we learned that UC Santa Barbara was surveying other universities about their teaching evaluation practices. Other universities across the country had recently changed, or were in the process of examining, their teaching evaluation practices, particularly with respect to students’ ratings of courses and instructors (e.g., U Oregon: <https://senate.uoregon.edu/tag/teaching-evaluations/>). In May 2019, the UC Academic Senate struck the UC Course Evaluation Task Force to discuss issues of reliability, validity, and bias in student evaluations. Thus, we are in a period of broad interest in, active examinations of, and dynamic adjustments to teaching evaluations. SCET and the UC Davis AS should be mindful of this context, as we can engage with and learn from these parallel and ongoing efforts.

The next three sections of this report present our considerations of the three sources of data on the evaluation of teaching that we were charged with assessing: Local practices at UC Davis; practices at peer institutions; and published literature on best practices. Based on our integration of the common themes and most striking observations from these sources, this is followed by two sections presenting the conclusions we drew: Our recommendations for minor changes to the UC Davis teaching evaluation practices that potentially could be implemented in the short term; and our identification of more substantive changes to teaching evaluation that could be considered for further study and future implementation over the longer term.

APPENDIX 2 - UC DAVIS

EVALUATION OF TEACHING AT UC DAVIS

The landscape of current approaches to the evaluation of teaching at UC Davis could be described as “highly variable.” There is very little consistency or commonality across units. In part, this reflects the reality of the differences in instructional practices and goals across very different disciplines. Why would one expect the evaluation of teaching to look the same in, for example, Art and Art History, Mechanical Engineering, and Nursing? Yet, all instruction involves some common elements, such as selection of course materials for appropriate depth and breadth of content, accuracy of instructor knowledge, teaching effectiveness for student learning and engagement, etc. For evaluation of teaching for summative purposes (merits and promotions) to be fair and equitable across instructors and units, there need to be broadly agreed-upon standards of what should be evaluated and how evaluations should be conducted. Although there has been less emphasis at UC Davis on the formative (enhancing professional development) and pedagogical (improving student learning) purposes of the evaluation of teaching, there also are mechanisms for improving these efforts that could share common elements across units.

In this section, we summarize the current practices for evaluation of teaching at UC Davis and identify some of the “best practices” that could be considered for broader implementation.

Student evaluations of teaching (SET)

Regulation 534 of the Davis Division of the Academic Senate pertains to Course Evaluations and states:

"In every course designated by the Committee on Courses of Instruction, all instructors must implement a course evaluation procedure in such a manner as to afford to each student the ability to evaluate the instructor and the course. Such evaluations shall be made available to the instructor after grades for the course have been submitted. The evaluation procedure shall, to the fullest extent possible, preserve the anonymity of the student and restrict the identification of the course instructor to authorized persons only, including the Committee on Academic Personnel and others involved in the academic personnel process and in the selection of course instructors. (En. 4/17/2012)"

As a result of regulation 534, current teaching evaluation practices at UC Davis are primarily driven by COCI (Committee on Courses of Instruction) policy that mandates registered students be given the opportunity to evaluate courses offered for academic credit. COCI defines the minimum elements of Course Evaluation as the following 2 questions, as written, in addition to an opportunity for comments:

1. Please indicate the overall teaching effectiveness of the instructor.
(5 = excellent; 4 = very good; 3 = satisfactory; 2 = fair; 1 = poor)
2. Please indicate the overall educational value of the course.
(5 = excellent; 4 = very good; 3 = satisfactory; 2 = fair; 1 = poor)

However, there are numerous Memoranda of Understanding (MOUs) with COCI that delegate oversight of course contents to a school or college, as well as the course evaluation process to the

APPENDIX 2 - UC DAVIS

respective school or college. As a result of these MOUs, and individual choices in various schools and departments, there is marked heterogeneity across campus in how course and instructor evaluations are executed. Some units ask students the minimum two questions only, however, other units ask as many as twelve questions regarding the course and up to ten questions regarding the individual instructors. In those units that ask more than the required two questions, there is variation in whether those questions appear at the beginning, end, or somewhere in the middle of the set of questions, which could affect how the questions are interpreted and responded to by students. Some units also ask questions about teaching assistants and some ask students to reflect upon their own efforts and accomplishments. Many units provide the opportunity for open comments at the end of the evaluation form, often being prompted by such things as - " Please comment on what worked well for you and what could be changed in the future to better aid learning ", "Please provide feedback to this instructor on their teaching effectiveness – What worked well? What could be improved upon? ", "For any of the above where you scored "Strongly Disagree", please provide comments", "Do you have any additional comments not captured above? ".

There is great variation across campus in the completion rates of SET. Variables that appear to influence completion rates include provision of dedicated time in class for completion, use of paper versus on-line SET forms, and whether inducements or penalties are attached to completion. Unlike some other institutions (Stanford), release of grades is not tied to completion of student evaluations by any unit at UC Davis. The School of Veterinary Medicine has very high rates of completion of student evaluations, as it links completion to expected standards of professionalism that are articulated to students. Evaluations are tracked, and repeated (3 or more) failures to complete evaluations result in visits to the Students Affairs Committee.

Although the use of SET as a feedback tool for instructors' own use (formative purpose) is not controversial, there is widespread instructor concern about the use of SET in the merit and promotion process (summative purpose). These concerns pertain to the potential for a variety of implicit and explicit biases to affect SET, including gender and racial bias, and to long-standing debate about the validity of SET for assessing accuracy and completeness of course content, teaching effectiveness and student learning. In some merit dossiers, SET are the only supporting information provided for teaching competency, despite APM 210 mandating that, regarding teaching, "more than 1 kind of evidence shall accompany each review file". There is marked departmental variation in mean SET ratings. It is unlikely that this reflects differences in the quality of teaching by different disciplines, but more likely differences in the student populations (lower division versus upper division, undergraduate versus graduate student and professional school) or disciplinary norms about the meaning of anchor points on rating scales.

As such, the consensus view of SCET is that it would be more appropriate to refer to SET as students' ratings of their "experiences" rather than as students' evaluations. Further, if SET are to continue being used for formative purposes, student ratings should be used with caution and only in conjunction with other measures of educational effectiveness when evaluating instructors.

APPENDIX 2 - UC DAVIS

Peer evaluations of teaching

Best practice should dictate that instructors receive feedback not just from students but from other experienced instructors who can comment on course content and design, as well as delivery, i.e., peer observation. Peer Observation of Teaching is one tool that provides rich, qualitative evidence for teachers, quite different from closed-ended or open-ended SET. When Peer Observation of Teaching is incorporated into university practice and culture, and is conducted in a mutually respectful and supportive way, it has the potential to facilitate reflective change and growth for teachers (Siddiqui et al., 2007).

Some departments use a form of peer observation whereby one colleague is asked to observe a lesson and to provide a narrative evaluation of an instructor's teaching that is included in the dossier based on that single observation. However, this type of peer observation has little constructive or formative utility, as the instructor does not have an opportunity to discuss elements of the course and its teaching (e.g., course design, required materials, instructional approach, etc.) with the observer. Further, with respect to summative utility, such reports typically lack credibility as they often are written as exclusively and effusively positive (a “crony report”) and provide no useful information to FPC or CAP when making merit and promotion decisions.

There are local examples of academic units that have begun to implement more comprehensive peer evaluation procedures that are more likely to be effective for summative, formative and pedagogical purposes. The MCB department in CBS has instituted a more robust peer review process for their faculty, especially those who are coming up for promotion (tenure step mostly). This process involves selection of two colleagues who have some familiarity with the curriculum in the course, a pre-observation meeting with the “candidate”, discussing their approaches and course resources, observation of approximately two appropriate lectures, and subsequent synoptic write up that accompanies the dossier.

Some of the professional schools have begun to provide a more holistic review of instruction, with greater emphasis on “enhancing” teaching effectiveness rather than (only) “evaluating” instructors. The School of Veterinary Medicine (SVM) currently runs a program of peer observation and coaching with the aim to provide instructors with formative feedback on their classroom sessions (both large and small group). The goals of the SVM program are twofold.

1. To provide a process to enhance teaching (and document progress)
2. To provide a framework for formative and summative peer discussion and self-assessment of teaching strategies.

SVM uses a Peer Observation instrument developed in conjunction with the Teaching Academy of the Consortium of West Region Colleges of Veterinary Medicine (<https://teachingacademy.westregioncvm.org/>). The instrument is based on pedagogical best practices and was developed by a working group specifically tasked to do so. The instrument consists of 3 parts (see appendix). The primary goal of the process is formative but the Post-Observation component has a summary paragraph and a categorical overall perception of teaching – Emerging, Evident, Exemplary – that accompanies the dossier. There are two

APPENDIX 2 - UC DAVIS

observers for each “observation,” ideally, one familiar with education best practices and one content expert / content familiar observer. SVM provides (workshop) training for observers. Two observations are recommended prior to tenure, and an additional observation is recommended prior to promotion to Full Professor. The first observation is a "range-finder", with the intent that progress / enhancement is demonstrated (and documented) with subsequent observations.

The School of Medicine (SOM) is currently working on implementing a similar program that uses content experts and educational specialists to work with instructors to improve teaching effectiveness and vertical integration of the curriculum.

The peer observation programs in both the SVM and SOM utilize pre-observation and post-observation meetings with individual instructors, in addition to classroom observation. The pre-observation meeting helps instructors reflect on the learning objectives/outcomes of the session being reviewed. Additionally, the instructor is asked specific questions about the session with regards to use of new pedagogies or materials. It is important for a professional school (where courses tend to be team-taught) to also have the instructor consider how the session fits into the overall course/curriculum. The post-observation meeting asks instructors to reflect on the teaching session(s), consider any challenges they encountered and how they might enhance the session going forward. Both peer observation/review programs aim to help instructors become more reflective instructors and to encourage use of effective (evidence-based) pedagogies. The ultimate goal of these programs is to enhance student learning and outcomes.

Self-assessment of teaching

Currently there is no consistency in expectations for or documentation of self-assessment of teaching across instructors or units. The teaching statement within merit and promotion dossiers is where most instructors might be expected to provide a summary of their practices. Yet, some instructors simply list the courses taught, their mean SET scores, and perhaps the number of trainees mentored. This does not necessarily mean that these instructors have not engaged in self-reflection in their efforts to enhance teaching effectiveness; they simply may not see that as something to be reported. Other instructors provide extensive detailing of their pedagogical philosophies, the professional skills development workshops or exercises that they have undertaken, their methods for mid-quarterly assessments of student learning and teaching effectiveness, and how they have used such activities to enhance the in- and out-of-classroom experiences of their students and mentees. CEE provides numerous online, individualized and group-based resources and opportunities for instructors to strengthen their self-assessment and enhance their teaching competencies. The extent to which instructors engage in such activities at CEE, at professional conferences or in other venues varies considerably within units, but also between units as the local departmental pedagogical culture appears to shape individual instructors' attitudes and practices for self-assessment.

APPENDIX 2 - UC DAVIS

EVALUATION OF TEACHING AT PEER INSTITUTIONS

UC Davis is not alone in its efforts to revisit and revise practices for evaluation of teaching. We hoped to gain inspiration and feedback from other institutions engaged in this effort. To this end, we gathered information from other institutions in three ways. First, we solicited feedback from 10 peer institutions (letter in Appendix) and received direct responses from two: UC Irvine (UCI), which shared the results of a recent study comparing two SET formats, and Yale, which recently completed a revision to the SET used in Yale College. Second, our own Center for Educational Effectiveness (CEE) and members of our committee provided information about teaching evaluation at the University of Kansas (KU) and Harvard Medical School. Finally, we reviewed public websites from University of Oregon, University of Southern California (USC) and Vanderbilt which describe those institutions' policies or reviews. Other institutions that we contacted did not respond in time to be included in this report. We have reviewed these materials with a focus on student evaluations, peer observation, self-reflection and other holistic approaches to evaluation of teaching.

Student evaluations of teaching (SET): Oregon, UCI, USC, Vanderbilt, Yale

SET involves a series of questions, presented in survey format to students. This is the most widespread tool for evaluating teaching, but as outlined elsewhere in this report, there are a number of concerns about this approach, particularly as the sole form of instructor evaluation. Other institutions are taking numerous and varied steps to improve SET and use student survey results as a source of information about student experience rather than instructor effectiveness.

- **Changes to SET structure or design**
 - Emphasize reflection on learning outcomes by placing open-ended questions about this first, or early on in the evaluation (UCI, Yale).
 - Indeed, UCI compared two SET formats within 35 courses and over 5,000 students. They found that their newer format resulted in slightly lower ratings for instructor behaviors, but that largely, there was little difference in outcomes, despite numerous changes to the design. Instructor satisfaction was slightly higher with the new form, but the bias identified in other studies (men > women, non-STEM > STEM, smaller courses > larger ones) was detectable in both versions.
 - Focus on the course rather than the instructor: at Yale, only 1 out of 10 questions asks about teaching effectiveness, as opposed to characteristics of the course such as workload, intellectual challenge, and organization.
 - Yale did not compare their old and new forms; they redesigned and now use the updated version.
 - USC recommends SET should include elements addressing: a) instructional design, b) instructor characteristics, c) learning experiences, d) assessments and feedback, e) diversity and inclusion practices.

- **Changes to SET presentation to students**
 - Include introduction about unconscious bias before students take survey (Yale).
 - Set aside time in class for SET, including introduction to the importance of the process, as part of a culture of commitment to teaching (Vanderbilt).

APPENDIX 2 - UC DAVIS

- Increase the frequency of SET by centrally administered midterm student experience surveys (Oregon).
- **Include questions about students or tie SET results to student outcomes**
 - Include questions about students' effort or commitment to the course (UCI).
 - Include questions about the perceived level of challenge associated with the course (Yale, UCI).
 - Link SET results to student outcomes such as a) project samples as part of student portfolios and performance; b) learning outcomes (recommended at USC); or (c) grades (tried by UCI in their study of 35 courses).
- **Changes to use and presentation of SET findings**
 - Change the name or clearly acknowledge that SET describe student experience or satisfaction, rather than evaluation of teacher effectiveness (Oregon, USC).
 - Make SET course ratings more transparent and easily available, which has resulted in more referrals of instructors to their teaching center (Yale).
 - Include a “warning label” and guidance about bias when SET results are presented in the M&P process (Oregon).
 - Stop using SET as the only form of instructor teaching evaluation (Oregon), or implement a multi-modal evaluation requirement for lecturer promotion process (Yale).
- **Improve response rates**
 - Offer some form of incentive to students for completion (UCI).
 - Suggest tying SET to release of course grades (USC).
 - Set aside time to complete SET in class (Vanderbilt).

Peer evaluations of teaching: Harvard, KU, UCI, USC

As noted earlier in this report, peer observation of teaching has been a component of UCD teaching assessment for tenure and promotion, but there is great variation in its implementation across campus. Other institutions have taken the following actions to develop and improve the peer observation process as part of enhancing, and assessing, teaching effectiveness.

- Harvard Medical School has a handbook and rubric for peer-evaluation process, which mirrors one of the recommendations in Section 5 (recommendations for the short-term) about use of a standardized form.
- USC recommends incorporating peer evaluation into *formative* teaching evaluation, echoing recommendations in Section 5 about the value of peer evaluation in a formative approach to improving instructional quality.
- USC recommends incorporating peer review of *instructional design* (see elements of teaching portfolio below) as part of both the formative and summative evaluation of instructors.
- UCI provides resources for best practices in peer evaluation to be used either on a voluntary basis or as part of merit and promotion that includes some of the recommendations outlined in Section 5, including a pre and post observation meeting between the instructor and observer(s) and use of a standardized form. Other recommendations outlined in Section 5 are not included in the UCI recommendations, as they indicate that their entire process should take only 3 to 4 hours.

APPENDIX 2 - UC DAVIS

Self-assessment of teaching: KU, UCI, USC, Vanderbilt

Self-assessment can include reflection statements, teaching portfolios, or midterm feedback from students. The degree to which any of these are mandated by the institution varies.

- USC and Vanderbilt both emphasize development and use of a teaching portfolio which can include a statement of teaching philosophy, sample syllabi, instructional plans, assignments with grading rubrics, sample work from students, and a self-reflection statement.
- KU and Vanderbilt ask for consideration of how the instructor's teaching has changed over time.
- UCI encourages the use of self-reflection statements and a Teaching Practices Inventory as part of the merit and promotion process.
- Oregon has increased the frequency of SET by adding centrally administered midterm student experience surveys whose results are only available to instructors for formative, not summative, purposes (e.g., to be made aware of course elements that are not working out as planned in order to adapt them accordingly in the second half of the course).

Other approaches: beyond SET, peer observation and self-assessment towards a more holistic approach to instruction quality

Several institutions value evaluation of teaching as a component of campus climate and may also use it as a tool to engage the wider community of educational institutions on a national level.

- USC outlines five specific recommendations about promoting a culture of excellence in teaching on their campus, including:
 - systemic review of course evaluation
 - articulating components and levels of what constitutes teaching quality
 - tying teaching evaluation to their campus-wide diversity plan
 - incentivizing professional development to promote teaching excellence
 - use of their equivalent of our Center for Educational Effectiveness as the platform for implementing change
- KU includes 7 elements of evaluation of teaching and provides a rubric for evaluation of each:
 - goals, content and alignment
 - teaching practices
 - achievement of learning outcomes
 - classroom climate and student perceptions
 - reflection and iterative growth
 - mentoring and advising
 - involvement in teaching service, scholarship or community
- KU and Vanderbilt assess instructor contributions to the broader, off campus teaching community.
- KU shares their teaching effectiveness benchmarks with other institutions engaged in similar efforts.

APPENDIX 2 - UC DAVIS

LITERATURE REVIEW

We reviewed the literature to identify best practices in the evaluation of teaching and to outline factors departments should consider when reflecting on teaching quality. We approached this review having already identified the need for a cultural shift that moves the evaluation process away from one in which teaching evaluations are solely used to sum up the delivery of instruction to a focus on improving teaching and learning. As Hattie (2015) explains, this cultural shift should move departments to focus on “seeking evidence to support interpretations about impact, having collective discussions about this impact, what the magnitude of this impact should be, and how pervasive is this impact on the students” (p. 89-90). That is, we view the evaluation process as an opportunity for instructors to reflect on their instruction with the explicit goal of improving student learning—to be effective, teaching evaluation systems must focus on improvement.

Evaluation of teaching is a process separate from the instrumentation used to gather information about instruction. Evaluation should involve many components, with some results shared summatively—attached to a merit or promotion dossier—and other components used formatively—for instructors to use to reflect on (and to improve) their instruction before the quarter has ended. Formative assessments may be developed and administered externally (e.g., classroom observation tools that are scored by colleagues) or involve the instructor’s own data collection for self-assessment, to privately reflect on their teaching.

Finally, we present general principles for the evaluation of teaching but do not prescribe a particular approach in recognition of the fact that each unit will need to develop or tailor a method that is appropriate for their particular context. We share recommendations for three assessment approaches that should be used in conjunction with each other below: student evaluations of teaching surveys, classroom observation instruments, and self-assessment.

Student evaluations of teaching (SET)

SET are surveys that ask students about their experiences in a given class and are used to summarize student experiences. They are one of the most widely used and most studied educational measures, yet little consensus exists about their quality. Hattie (2015) and Marsh (2007) report that SET are reliable and valid indicators of instruction. However, others have noted tremendous variability between students in rating the same instructor (Clayson, 2018), variability in the honesty of evaluations because students sometimes think that ratings could affect their own grades (McClain, Gulbis & Hays, 2018), and potential sources of bias associated with the gender, race, ethnicity, or culture of the instructor (Linse, 2017; Macnell, Driscoll, & Hunt, 2014).

Despite extensive research, the extremity of bias is unknown. As Linse (2017) argues, while bias clearly plays a role in the ratings provided by some students, it is less likely that bias is pervasive enough in most situations to cause persistently low ratings across all courses taught in a review period. Linse goes on to note severe methodological flaws in many studies of bias in SET that lead to incorrect conclusions. This variability in the rigor with which studies are conducted is one factor contributing to the lack of clarity about the extent to which bias substantially pollutes

APPENDIX 2 - UC DAVIS

scores (see also McClain, Gulbis & Hays, 2018 who concur with this interpretation of the literature).

Despite these concerns, SET are one important component of a balanced teaching evaluation system. Recommendations from the literature on SET include:

- Consider class context when interpreting ratings. Things like class size, whether the course is for undergraduate or graduate/professional students, and course topic can influence ratings (Linse, 2017).
- Examine the entire collection of SET ratings instead of focusing on ratings from only one class (Linse, 2017).
- Take active steps to increase response rates, such as: providing students with access to the course evaluation ratings, regularly collecting feedback from students in a variety of formats throughout the quarter so that students understand the value of feedback to instructors, verbally explaining the ways in which feedback will be used and its value to instructors, or administering ratings live in class instead of expecting students to complete them outside of class (Linse, 2017).
- Remember that SET provide important information about the student experience, but that students are unqualified to reflect on important aspects of instruction like the accuracy of instructional content. Recent meta analyses have found inconsistent relations between SET ratings and measures of student learning including course grades, which tend to be small or nonsignificant when examined using high quality SET and learning outcomes measures (Clayson, 2009; Uttl, White, & Gonzalez, 2017).

Peer evaluations of teaching

Peer observations of teaching can also yield helpful information about teaching quality, and they have the added benefit of creating an opportunity for colleagues to jointly reflect on potential approaches for improving instruction. Observations can be useful in their own right as a supplement to SET to provide multiple sources of information about teaching. However, peer observations are much more valuable if the observer and instructor being evaluated meet both before and after an observation session to discuss instructional goals, what happened during the lesson, and next steps in instruction. This formative approach is likely to improve instruction, but requires a commitment on the part of instructors to approach observation as more than checking off a box of tasks necessary to complete a dossier.

Recommendations from the literature on observational approaches to evaluate teaching (discussed more fully in Fletcher, 2018) include:

- Hold a pre-observation meeting to discuss instructional goals, and things the instructor would like the observer to focus on to ensure feedback improves instruction.
- Use a standardized form, rooted in pedagogical best practices, to summarize observation findings. This ensures that important facets of an observation are addressed and ensures consistency across instructors — consistency in observation is especially important if they part of the impact merit and promotion process.

APPENDIX 2 - UC DAVIS

- Remember that instructional quality varies from lesson to lesson and class to class. Studies of classroom observation ratings in the K-12 setting have found that at least five observations are needed to gather reliable ratings (Hill, Charalambos, & Krat, 2012; Kane, Kerr, & Pianta, 2014). While it is unlikely that units will conduct multiple observations per instructor, it is important to keep this in mind in drawing conclusions based upon only one observation.
- Hold a post-observation meeting to discuss the lesson. Feedback should be positive and supportive, inform next steps in instruction, and focus on pedagogy. Observers should address a limited number of issues, ones that are most likely to improve student experiences. Providing too much feedback will be overwhelming for the person being observed and thus limit its utility.

Self-assessment of teaching

Self-assessment is essential to becoming an effective instructor. As Hattie (2015) explains “To be successful, university teachers need to think of themselves as evaluators and ask about the merit, worth, and significance of the impact of their interventions— essentially, successful educators actively practice the Scholarship of Teaching and Learning (SoTL).” (p.80). That is, good teaching is based in evaluating how your instruction is going for all students.

Self-assessment can take many forms, but always involves collecting evidence, reflecting on it, and making instructional changes based on the results. That is, it involves gathering feedback about how instruction is going with the explicit intent of improving instruction. From a promotion, tenure, and merit standpoint, it is essential that instructors write about their self-assessment practices in their teaching statement and that they emphasize what they learned about their teaching and how their instruction changed as a result. Summaries of self-assessment findings are secondary to how instructors respond to results.

Recommendations from the literature on self-assessment include:

- Collect self-assessment evidence during instruction to allow for instructional adjustments. As Hattie (2015) explains, “The most critical mind frame is ‘know thy impact’--when an academic walks into a teaching situation their fundamental question needs to be ‘how will I know my impact today.’” (p. 89)
- Ensure the anonymity of students in the self-assessment process. As McClain, Gulbis and Hays (2018) point out “...some student responses vary according to which format they think best protects their identity.” (p. 381)
- Make sure to include opportunities to share open-ended feedback so key concerns (or compliments) are captured. One approach, recommended by Boston University’s Center on Teaching and Learning is to ask three questions: (a) “What should I start doing?” (b) “What should I stop doing?” and “What should I continue doing?” (<http://www.bu.edu/ctl/teaching-resources/start-stop-continue/>) Other approaches to collecting self-assessment data can be found here: http://www.crlt.umich.edu/gsis/p9_1
- Analyze results and decide how to act upon the results. Share both results and action steps with students during the class meeting immediately following data collection. See

APPENDIX 2 - UC DAVIS

<http://www.bu.edu/ctl/teaching-resources/start-stop-continue/> for an example of how to do this.

- Consider the use of teaching portfolios (dossiers) as they are one promising method for presenting self-assessment evidence. Developing a portfolio is an extensive process which involves:
 - Creating a repository to collect work samples. At this stage, units must make decisions about what kinds of work samples to collect and what competencies to address in the portfolio.
 - Developing a procedure to provide feedback on the work samples so that they can be revised, or new work samples added, before the portfolio is finalized.
 - Curating the portfolio. Instructors use feedback to select a portion of work samples for inclusion in the final portfolio (dossier). They also write a reflection that discusses how the work samples provide evidence of meeting the competencies being evaluated (Clarke & Boud, 2016; for details of how to do this and example forms see: <https://cft.vanderbilt.edu/guides-sub-pages/teaching-portfolios/>)

Conclusions

As Benton and Young (2018) explain:

Units that take a balanced approach recognize the challenges in evaluating teaching effectiveness. The accumulated evidence must come from multiple sources and include materials such as descriptions of teaching activities, modifications made to courses, adoption of new teaching strategies, participation in professional-development activities, and contributions made to better the unit's overall instruction. Multiple measures increase the likelihood that the evaluation will encompass all dimensions of teaching, including course design, course delivery, course assessments, instructor availability, and course management." (p.8)

Treating student evaluations of teaching, peer observation, and self-assessment as three legs of a balanced assessment system is a good first step in creating an environment that encourages support and growth in order to improve teaching practice. It would also improve our conceptualizing of what teaching involves and would more comprehensively capture instructional endeavors.

APPENDIX 2 - UC DAVIS

RECOMMENDATIONS THAT MAY BE ACTIONABLE IN THE SHORT TERM

General

1. Clearly and consistently document in all communications that teaching evaluation requires information from multiple sources, including (but not limited to) SET, peer observation and self-reflection, and that no one source of information should be considered sufficient for the summative purpose of evaluating educator performance.
2. Begin working on culture change from teaching evaluation as *summative* for the purposes of merit and promotion to *also* being *formative* for instructors' development of professional skills and *pedagogical* for enhancing teaching effectiveness and students' learning outcomes. This can begin with simple language changes in the framing and presentation of the various components of teaching evaluation, e.g., wherever possible, replace 'evaluate' with 'enhance' or 'benefit' when describing the purpose (e.g., peer observation for the purpose of *enhancing* teaching performance).
3. Require departments to vote separately on each aspect of evaluation (teaching, research, service) as its own content area prior to voting on recommended merit step for the candidate. Some departments already do this, but not all.
4. Drawing on the expertise of Affirmative Action and Diversity Committee, STEAD, Vice-Chancellor of Diversity and Inclusion and other sources of expertise on diversity and bias, include standard language preceding and introducing teaching dossiers for FPC and CAP members conducting merit and promotion reviews.

Student evaluations of teaching (SET)

1. Rather than "evaluation", characterize SET as "Student *experience* of teaching", "Student *experience* of the course" or similar (Given AS policy, this may be a long-term recommendation). Assuming the language of the regulation cannot be changed easily, include standard language introducing and framing the online and paper SET forms as being "for the goals of sharing your experience of the course in order to enhance the quality and effectiveness of instruction."
2. Recognizing that we are in a period of local (UCD), system-wide (UC) and nation-wide examinations of the use of SET for summative purposes, AS may want to consider including a letter to campus FPCs and CAP explaining the issues relating to reliability and validity of SET for all merit and promotion decisions until there is greater consensus on their appropriate design, administration and interpretation.
3. Encourage all instructors to provide in-class time for students to complete SET, even if doing so online, in order to increase participation. Encourage instructors to introduce SET by discussing the value of SET in helping them to improve future versions of the class (formative and pedagogical goals).
4. Encourage instructors and departments to use more than the minimum two SET questions required by SET, and to precede those two globally evaluative questions with more concrete questions about specific aspects of the course (e.g., clarity of presented materials, effectiveness of text or other instructional materials, etc.), instructor (e.g., organization, timeliness, responsiveness, etc.), and student (e.g., attendance, completion of assignments, participation, expected grade). Which specific questions are appropriate

APPENDIX 2 - UC DAVIS

are likely to vary by course and discipline, but knowing which aspects of the course or instruction were less versus more well-received by students is important for the formative goal, and to scaffold the students' understanding and consideration of the two globally evaluative questions.

5. Drawing on the expertise of Affirmative Action and Diversity Committee, STEAD, Vice-Chancellor of Diversity and Inclusion and other sources of expertise on diversity and bias, include standard language introducing online and paper SET forms reminding students of the possibilities of implicit biases when completing SET.
6. Currently the SET scantron forms include demographic information but the online forms do not. This inconsistency should be resolved in one of two ways: Either remove from the scantron forms or add to the online forms. A primary argument for removal would be eliminating the possibility of identifying students in small classes. A primary argument for inclusion would be to increase the potential for instructors and the Center for Educational Effectiveness (CEE) to do research on SET responses in order to identify constituents for whom instruction is working less well and adapt the course correspondingly (formative and pedagogical goals). If demographic is included, currently Gender only has binary "female/male" options; add a non-binary option (at least in class sizes where this would be unlikely to identify individuals).

Peer evaluations of teaching

1. Encourage peer observers to meet with instructor prior to observation to establish expectations for in-class experience, and again after in-class observation to discuss and provide feedback.
2. Encourage peer observers to examine course portfolio (syllabus, materials, Canvas site, etc.) comprehensively prior to in-class observation and as part of overall evaluation.
3. Drawing on CEE resources, academic units should provide a rubric or template of content areas, competencies and pedagogical goals to guide in-class observations and peer evaluations and to increase consistency of peer evaluations across instructors.
4. Re-brand peer evaluations as a "coaching model" of enhancing teaching effectiveness (see above note about culture change and language).
5. Drawing on the expertise of Affirmative Action and Diversity Committee, STEAD, Vice-Chancellor of Diversity and Inclusion and other sources of expertise on diversity and bias, include standard language reminding peer evaluators of the possibilities of implicit biases when completing observations and examinations of course materials.
6. Treat peer evaluation/observation as a facet of department/university service.

Self-assessment of teaching

1. Encourage instructors to adopt the practice of making teaching portfolios/dossiers to more comprehensively document their course design, materials, syllabi, etc.
2. Remind instructors of CEE's available services for teaching reflective practices for effective instruction, for consulting on course design and delivery, etc.
3. Encourage instructors to use mid-quarter SET (ratings or open-ended questions) to learn about students' experiences of the course in time to make instructional adjustments to that course.

APPENDIX 2 - UC DAVIS

4. Remind instructors to document and report efforts made to enhance course delivery and improve their own teaching effectiveness, including any professional-development activities, use of new teaching methods or technologies, etc.

APPENDIX 2 - UC DAVIS

RECOMMENDATIONS FOR ACTIONS THAT MAY BE TAKEN OVER THE LONGER TERM

General

1. Continue working on culture change from teaching evaluation as *summative* for the purposes of merit and promotion to *also* being *formative* for instructors' development of professional skills and *pedagogical* for enhancing teaching effectiveness and students' learning outcomes.
2. Encourage academic units to develop their local standards, expectancies and guidelines for effective teaching evaluation within their unit, for example, identifying the specific questions that are most relevant to include in SET for that discipline, the instructor competencies and course features that should be the focus of peer evaluators, and the particular contents of teaching portfolios and statements that instructors should prepare.
3. Adapt MIV to accommodate more comprehensive teaching portfolios for instructors to document multiple aspects of their courses (syllabi, instructional materials, links to Canvas, etc.).

Student evaluations of teaching (SET)

1. Change the wording of the AS policies and regulations such that, rather than "evaluation", SET are characterized as "Student *experience* of teaching", "Student *experience* of the course" or similar.
2. No source of information about teaching is without potential bias, but concern was strongest about the validity and usefulness of SET for summative purposes. Some SCET members advocated for AS removing SET from merit and promotion decisions by FPC and CAP (e.g., Oregon) and basing the evaluation of teaching for summative purposes on other sources of information (peer evaluation, self-assessment). Other members of SCET supported retaining the use of SET, with carefully implemented improvements, as one component of a broader assessment of teaching. This topic warrants further study and consultation with vested university constituents prior to deciding on a course of action.
3. Include information on the class mean grade and grade distribution, and departmental norms and ranges for SET and grades, to accompany candidate's SET when FPC and CAP engage in teaching evaluations.
4. Transfer responsibility for running online SET from CAES to Center for Educational Effectiveness (CEE), which already runs scantron SET, in order to increase potential for flexibility and research. Develop mechanisms for CEE to be able to match de-identified (name and SID removed) information on student performance/grades and characteristics (e.g., major, transfer student status, etc.) with SET responses. This would support CEE's ability to do research on relations between SET and learning outcomes, and on instructional effectiveness across different campus constituencies. This will involve defining a minimum class enrollment size for matching to reduce risks of identifying individuals through responses.
There are perceptions of potentially systematic differences across fields and disciplines with respect to students' norms for completing SET (i.e., students in some units give

APPENDIX 2 - UC DAVIS

lower ratings than students in other units that are not due to differences in instructional quality). This may be an appropriate question for CEE to research.

5. Have CEE partner with Affirmative Action and Diversity Committee, STEAD, Vice-Chancellor of Diversity and Inclusion and other sources of expertise on diversity and bias, to examine the specific wording of items and content areas in SET
6. If actions to improve the validity of SET are put in place, consider the development of a platform for sharing SET for courses (by course, not by instructor) with students (e.g., Yale). Currently students share their opinions about courses informally and on platforms of dubious quality and accuracy (e.g. RateMyProfessor). Sharing accurate SET data may help to combat such inaccurate information.
7. Implement mid-quarter SET for instructors to learn about students' experiences of the course in time to make adjustments. These should be formative and for instructors' ability to enhance the course, not summative for the purposes of merit/promotion evaluations, and hence shared only with the instructors (and, possibly, the students).
8. Avoid individual carrot-or stick approaches (e.g., student gets an extra point for completing SET, or student does not get grade released until SET completed). Study whether establishing collective reward procedures (e.g., all students get extra point if $\geq 85\%$ of class completes SET) may be appropriate as part of an effort to advance the cultural shift toward effective and inclusive teaching evaluation.
9. Outliers may introduce statistical bias in the interpretation of SET. Consider implementing methods to attenuate this, for example, removing the extreme tails of the distributions (trim the top and bottom of the student responses symmetrically, such that a percentage of the lowest scores and the same percentage of the highest scores are excluded from the average score).

Peer evaluations of teaching

1. Establish protocol for observer(s) to meet with instructor prior to observation to define expectations and instructor's goals for the class, and to meet with instructor after the observation to discuss the class experience.
2. Develop training and support structure for conducting more effective peer observations. This could include an in-person workshop or an online training module for peer observers to complete prior to meeting with instructors, guidelines about expectations for effective observation and feedback, a rubric of rating scales and open-ended prompts on specific aspects of class instruction for observers to complete as part of their evaluation, criteria for evaluating course syllabi and other materials, etc.
3. Having more than one observer and/or observing more than one in-class instruction should become the norm in order to have more comprehensive and accurate, and likely less individually biased, feedback and evaluation by peer observers.
4. Consider observer teams including both a within-department or within-discipline member (for content) and a member from outside the department (for greater objectivity).
5. Improving the effectiveness of peer evaluations for formative, pedagogical and summative purposes is likely to increase the time involved in conducting peer evaluations. AS needs to conduct a cost/benefit analysis of this, and determine appropriate methods of compensating peer observers for their time (e.g., establish this as formal university service).

APPENDIX 2 - UC DAVIS

Self-assessment of teaching

1. Develop a rubric for guiding self-evaluation; for example, a template that parallels the content areas of the rubric for peer observations may be effective.
2. Provide examples of self-assessment strategies that instructors can use to evaluate their own performance, including use of mid-quarter SET to obtain student feedback on experience of course and instruction in time to make adjustments.
3. Create a platform with examples of effective teaching portfolios/dossiers and teaching statements with guidelines for what should/could be included in these, including description of self-assessment procedures in the teaching assessment (Vanderbilt, USC).

APPENDIX 2 - UC DAVIS

REFERENCES

- Benton, S. L., & Young, S. (2018). Best Practices in the Evaluation of Teaching. *IDEA Paper*, 69.
- Clarke, J. L., & Boud, D. (2018). Refocusing portfolio assessment: Curating for feedback and portrayal. *Innovations in Education and Teaching International*, 55(4), 479-486.
- Clayson, D. E. (2018). Student evaluation of teaching and matters of reliability. *Assessment & Evaluation in Higher Education*, 43(4), 666-681.
- Fletcher, J. A. (2018). Peer observation of teaching: a practical tool in higher education. *The Journal of Faculty Development*, 32(1), 51-64.
- Hattie, J. (2015). The applicability of Visible Learning to higher education. *Scholarship of Teaching and Learning in Psychology*, 1(1), 79.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56-64.
- Kane, T., Kerr, K., & Pianta, R. (2014). *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. John Wiley & Sons.
- Linse, A.R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Springer, Dordrecht.
- McClain, L., Gulbis, A., & Hays, D. (2018). Honesty on student evaluations of teaching: effectiveness, purpose, and timing matter!. *Assessment & Evaluation in Higher Education*, 43(3), 369-385.
- Siddiqui, Jonas-Dwyer and Carr. *Medical Teacher* 2007; 29: 297-300.
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42.

APPENDIX 2 - UC DAVIS

Membership on the Special Committee on Evaluation of Teaching

Paul Hastings, Chair and Professor of Psychology

David Begun, College of Biological Sciences Representative and Professor of Evolution and Ecology

Christyann Darwent, Committee on Faculty Welfare Representative and Associate Professor of Anthropology

Gina Dokko, Committee on Affirmative Action and Diversity Representative and Associate Professor of Management and Organizations

Debbie Niemeier, College of Engineering and Professor of Civil and Environmental Engineering

John Payne, School of Medicine Representative and Professor of Physiology and Membrane Biology

Sara Proctor, Academic Federation Representative and University Extension Teacher in Continuing and Professional Education

Michelle Rossi, Graduate Student Association Representative and Ph.D. Student in Sociology

Hollis Skaife, Graduate School of Management Representative and Professor of Accounting

Alexander Soshnikov, Committee on Academic Personnel – Oversight Representative and Professor of Mathematics

Brian Soucek, School of Law Representative and Professor of Law

Jan Szaif, Committee on Courses of Instruction Representative and Professor of Philosophy

Amelia Triest, Academic Federation Representative and Lecturer in Music

Cassandra Tucker, College of Agricultural and Environmental Sciences Representative and Professor of Animal Science

William Vernau, School of Veterinary Medicine Representative and Professor of Pathology, Microbiology and Immunology

Theresa Costa, Committee Analyst

APPENDIX 2 - UC DAVIS

Davis Division of the Academic Senate Special Committee on Evaluation of Teaching Approved by Executive Council on January 10, 2019

Overview:

The Davis Division Executive Council discussed Student Evaluations of Teaching (SET) in December 2018. This discussion highlighted faculty concerns about the current practices at UC Davis in obtaining SET (e.g., wording of items; low response rates; harmful and biased comments) and the overreliance on SET to assess teaching quality for merits and promotions. It is clear that more study and discussion is needed. A special committee should be created to bring forward recommendations to Senate leadership, Executive Council, and the Administration. The special committee's report may be submitted as near-term recommendations (potential minor revisions to current practices for evaluating teaching) as well as long-term recommendations (potential major revisions to current practices for evaluating teaching) for consideration.

Charge:

The Special Committee on Evaluation of Teaching will be to (1) evaluate the current practices for evaluation of teaching for merits and promotions at UC Davis; (2) consider practices for evaluation of teaching at peer universities (e.g., wording of items used in SET; methods to improve response rates for SET; alternative ways to evaluate teaching outside of SET such as peer review); and (3) consider research literature on best practices for evaluation of teaching in higher education. The final report should recommend possible near-term minor revisions to the current practices as well as potential long-term major revisions.

Membership:

The Committee on Committees will select the chair of the Special Committee on Evaluation of Teaching. The Senate Chair will seek a representative from: Committee on Academic Personnel (CAP), Faculty Welfare (FW), Affirmative Action and Diversity (AA&D), Committee on Courses of Instruction (COCI), and each School and College Faculty Executive Committee. The Academic Federation will be invited to name two representatives. There should be one student representative appointed by each of the ASUCD and GSA. The following administrative units will be consulted as needed: Provost's Office, Student Affairs, Undergraduate Education, Graduate Studies, and owners of the Automated Course Evaluation (ACE) system.

Duration:

The Special Committee on Evaluation of Teaching will be appointed by February 1, 2019 and will submit their report by June 3, 2019. The special committee will meet as necessary to organize work efforts and meet the committee's deadline.

Staff Support:

Staff support will be provided by Theresa Costa with assistance from Liz Lopez.

APPENDIX 3 - UC SAN DIEGO

ACADEMIC SENATE: SAN DIEGO DIVISION, 0002
UCSD, LA JOLLA, CA 92093-0002
(858) 534-3640
FAX (858) 534-4528

December 19, 2019

ELIZABETH H. SIMMONS
Executive Vice Chancellor

SUBJECT: Review of the Senate Administration Workgroup Report on Holistic Teaching Evaluations

Dear EVC Simmons,

Senate Council reviewed the Senate Administration Workgroup Report on Holistic Teaching Evaluations at its meeting on December 16, 2019. The reviewers commend the Workgroup for its thorough effort to identify and make available multiple existing tools for teaching evaluation, to establish a campus culture where assessment of teaching and learning is a standard practice and to institute faculty development programs. The reviewers endorsed the Workgroup's recommendations. Additional comments are summarized below.

- Reviewers proposed that a faculty oversight committee be established to help departments and colleges understand the portfolio model and how it can be used for assessment, pedagogical improvement and evaluation.
- Reviewers recommend that the overhaul to the existing CAPE questionnaires include the opportunity for student feedback.
- Reviewers asked for clarity regarding how peer review reports, CAPES, and teaching portfolios will be weighted and emphasized the need for Deans and Department Chairs to receive support and instruction on how to compile teaching portfolios.
- Reviewers expressed that The Teaching + Learning Commons should be given additional resources to implement the Workgroup report's core recommendations.
- Reviewers recommend that at the graduate student level best practices be developed to evaluate the faculty's mentorship activities and provide mentees with the opportunity to give constructive feedback to their mentors.

We look forward to further discussion of the Workgroup Report at the January SAC meeting.

Sincerely,



Maripat Corr, Chair
San Diego Divisional Academic Senate

Enclosure

Cc: Stephen Constable, Vice Chair, San Diego Divisional Academic Senate
Ray Rodriguez, Director, San Diego Divisional Academic Senate Office

APPENDIX 3 - UC SAN DIEGO

Senate-Administration Workgroup on Holistic Teaching Evaluation – Final Report August 28, 2019

1. Introduction

The Senate-Administration Workgroup on Holistic Teaching Evaluation was convened on January 11, 2019 to provide recommendations that will allow the university to:

- Identify and make available multiple existing tools for teaching evaluation
- Establish a campus culture where both formative and summative assessment of teaching and learning is a standard practice
- Institute or augment faculty development programs

In the process, the workgroup was asked to base its recommendations on literature, best practices, and existing UC San Diego resources. The workgroup also was asked to identify new resources/programs, review incentives, identify leaders in this area, and consider how to communicate the results to the campus communities. The full charge is included in Appendix A.

This is a wide-ranging charge; over the course of 12 meetings during Winter and Spring 2019, the workgroup made significant headway and is now able to make several recommendations. We hope that these can form the basis of further conversations and that the campus consider implementing the core recommendations:

- Maintain a clear distinction between assessment (formative) and evaluation (summative) to both encourage pedagogical awareness and facilitate the evaluation of teaching effectiveness in academic files (section 2).
- A thorough overhaul of CAPE questionnaires, taking into account best practices with respect to student feedback on teaching (section 3)
- The adoption of a holistic teaching portfolio and accompanying instructor self-reflection as the primary means of evaluating teaching in academic files (section 4)
- Recommendations regarding training and oversight (section 5)

As noted in the charge, our campus has relied primarily on Course and Professor Evaluations (CAPEs) for evaluating teaching effectiveness in academic files. Problems with such student evaluations are well-documented (see section 3). Furthermore, while the Academic Personnel Manual (APM) mandates at least two types of teaching effectiveness evaluation; this has not been the norm over many years and attempts to solicit additional input has led to *ad hoc* mechanisms that have not always been helpful. Our portfolio recommendation provides a guide to more comprehensive and holistic evaluation. However, adopting a portfolio model requires considerable guidance and support – both for faculty and campus reviewers; section 5 recommends ways this might be achieved, as well as how the transition might be monitored.

Each section will begin with a statement of the primary challenge and the committee's recommendation; the recommendations – sometimes augmented – are reiterated at the end of each section.

2. Assessment versus Evaluation

Challenge: Faculty often come to their teaching roles with little pedagogical training and typically get minimal feedback on how to improve teaching and support student learning.

APPENDIX 3 - UC SAN DIEGO

Recommendation: Develop separate formative (assessment) and summative (evaluation) processes that allow faculty to hone their teaching skills and demonstrate their commitment to teaching effectiveness.

A fundamental distinction – well-documented in the literature – is what we call ASSESSMENT versus EVALUATION (referred to as ‘formative’ versus ‘summative’ in the workgroup charge and in Hoyt and Pallett 1999). The primary distinction between these practices has to do with their intended purpose:

Assessment:

- Assessment is oriented towards the improvement of teaching and is part of faculty instructional mentorship and development. It is supportive and confidential; it does not become part of a faculty member’s file. The purpose of teaching assessment is to provide constructive feedback to instructors so that they might improve pedagogical practices and better enable student learning.

Evaluation:

- The purpose of teaching evaluation is to inform personnel decisions and judge the results of mentorship and development. Instructors are accountable to high standards of teaching effectiveness, as detailed in APM 210-1-D: “Clearly demonstrated evidence of high quality in teaching is an essential criterion for appointment, advancement, or promotion.”
- Evaluation begins at the departmental level and becomes part of the basis for academic personnel recommendations by campus reviewers. Departmental evaluations should take discipline and departmental standards into account and rely on multiple data points.

The availability of teaching assessment is a crucial resource in the service of our campus commitment to student-centeredness. Given that faculty often come to their teaching roles with little pedagogical training, it is incumbent upon the institution to make this a part of faculty professional development. Our campus has invested heavily in the Teaching+Learning Commons to provide this type of support to faculty.

Because teaching effectiveness is a criterion for faculty appointments and promotions, there needs to be a fair and transparent mechanism for teaching evaluation. Just as campus reviewers evaluate research/creative activity and university (and other) service when considering faculty files, there need to be relatively objective mechanisms for evaluating teaching effectiveness.

Assessment and evaluation can be difficult to tease apart and they can interact in interesting ways; nevertheless, understanding the difference is crucial for achieving their goals.

One source of confusion between assessment and evaluation comes from the fact that the same type of vehicle may be used in either mode. For example, peer observations might form the basis for assessment (where the peer observer provides constructive feedback to the instructor) or evaluation (where the observation report is included in an academic file). Faculty, departments, or the campus might set guidelines regarding what should be used for assessment versus evaluation, but it is easy to imagine this subject to variation across units and faculty. Furthermore, as we will recommend below, faculty might elect to include assessment materials and comment on how they have responded to them as part of their evaluation portfolio. Such self-reflection and engagement in assessment can demonstrate active interest in teaching excellence.

Despite the potential ambiguity between assessment and evaluation, it is crucial that they remain distinct and that faculty have confidence that assessments will not automatically form part of their evaluation. The Teaching+Learning Commons is adamant that their services be viewed as providing formative

APPENDIX 3 - UC SAN DIEGO

assessments and that, regardless of their recommendations, the results will not be used in academic files without the instructor's permission. Otherwise, there could be a powerful disincentive for faculty to seek support and formative feedback to engage in assessment.

Ideally, we want faculty to engage in assessment that results in improved teaching and student learning. The evaluation of teaching effectiveness should examine multiple sources of information; as we propose below, this might take the form of a teaching portfolio and faculty self-reflection. Faculty might elect to include assessment experiences in their self-reflections. However, such discussions should be voluntary and their omission should not negatively affect file evaluation.

Finally, it is important to stress that both assessment and evaluation should be in the service of encouraging active engagement in pedagogy with the goal of improved student learning.

Recommendation: To improve pedagogy, student learning, and faculty accountability, clearly delineate assessment from evaluation and identify which vehicles are used for which.

3. Student Feedback

Challenge: While student feedback has been the primary vehicle for evaluating teaching effectiveness, the literature has demonstrated the potential for significant bias.

Recommendation: Revise student feedback questionnaires include targeted questions regarding pedagogical practice (as opposed to instructor characteristics). Base the questions on best practices, as reported in the literature.

As mentioned above, UC San Diego has – almost exclusively – relied on Course and Professor Evaluations (CAPEs) in the evaluation of teaching effectiveness.¹ CAPEs began as a grassroots student project in 1972 and ran with significant student engagement for several decades. The CAPE evaluations included (and still include) multi-valent answers to a series of questions about the course and instruction; in addition, students are invited to write comments.² While several questions pertain to specific aspects of the course and instruction (e.g. “Exams are representative of the course material,” “Instructor is well prepared for classes”), two summative yes-no questions are the ones most often referenced in evaluations: “Do you recommend this course overall?” and “Do you recommend this instructor overall?” CAPEs were published in a book format, which several student staff members edited and produced. Furthermore, because evaluations were filled out in class on paper, the CAPE organization employed students (‘runners’) to visit each class to distribute and collect CAPE forms. Because of the large staff, CAPE represented a sizable and enthusiastic student organization.

¹ In this report, we will use ‘student feedback’ instead of ‘student evaluation’. We wish to retain the technical use of ‘evaluation’ as discussed in section 2 – the summative evaluation employed in academic personnel contexts. The use of ‘evaluation’ in the CAPE acronym is historical and does not represent the new definition.

² For many years the CAPE book produced facsimiles of particularly outrageous student comments in a popular “Off the Wall” section. While intended to be humorous, many of these comments were inappropriate, leading to concern that the existence of this feature encouraged disrespectful and discriminatory comments, degrading the quality of the overall evaluation process. The section was eliminated in 1999 because of a formal complaint from the Chancellor’s Diversity Council. It was reinstated in 2003, but ended after 2004, which saw the last published CAPE book.

APPENDIX 3 - UC SAN DIEGO

In 2010, CAPE moved to an online format; students now fill out the surveys online and the results are posted on a website (cape.ucsd.edu).³ This has had several consequences:

- Response rates have declined.
- Student comments – previously hand-written – became available electronically (although only available to the instructor and campus reviewers).
- Without a published book or CAPE runners, the need for a large student staff has disappeared.

The lower response rates have been cause for concern. While the response rate has dropped, it has not done so as dramatically as one might think; e.g., 56.2% in 2009 vs 46.9% in 2010. Furthermore, the response rate has increased in recent years and was at 52.3% 2018-19 (F,W,SP). In addition, many faculty worry that participation will be skewed towards disgruntled students. CAPE conducted a study comparing paper-based evaluations (from Fall 2002-2009) with online evaluations (Winter 2010-Spring 2011). The results show that 83.4% of the online evaluations have scores within 5 points of the paper evaluations (for the “Do you recommend this instructor?” question); 92.1% are within 10 points (<http://cape.ucsd.edu/files/ComparisonofOnlinevsPaper.pdf>). Nonetheless, the perception that online CAPEs put faculty at a significant disadvantage remains. Finally, the CAPE website mentions several ways instructors can encourage more participation (<http://cape.ucsd.edu/faculty/tips.html>); given recent increases – approaching the response rate of the pre-online format – these may have had an impact.

Electronic student comments have greatly facilitated the ability of campus reviewers to augment numeric scores with additional context. Members of the Committee on Academic personnel (CAP) often note that student comments allow for more nuanced interpretation of CAPE results. Previous to the on-line collection of comments, files either included no comments (leading to CAPE scores as the only evidence of teaching effectiveness) or photocopied handwritten comments (often with poor legibility).

The reduced student staff has led to a situation where there is very little actual student engagement in the CAPE organization. One part-time student worker, supervised by the Office of Undergraduate Education, is paid for about 10 hours a week. The supervising staff reports that applicants for the position often have little idea of what CAPE is and no concept of its history as a student-run organization. While students are certainly aware of CAPEs (since over half of students fill out evaluations), it is not widely known that it is a student organization; nor do students generally realize that evaluations are used in faculty files. Conversely, faculty, while well aware that CAPEs go into their files, do not often realize that the instrument was developed by students and is nominally a student organization. There is a faculty advisory committee, but it meets once a year and participation has been variable; for example in Spring 2018, no faculty members showed up for the single scheduled meeting.

In sum, the evaluation of teaching effectiveness has, for many years, been achieved through an instrument that was not designed to be evaluative. It was intended to provide students a way of sharing their impressions regarding courses and faculty – much like RateMyProfessors.com. Furthermore, while there was once significant student engagement in the process, that has all but disappeared, following the move to an on-line format.

³ The last CAPE book was published in 2004; between 2005 and 2009, students continued to complete paper evaluations, but the results were reported online.

APPENDIX 3 - UC SAN DIEGO

There is an extensive literature on student feedback on teaching (sometimes referred to as 'student evaluation of teaching' or SETs; see footnote 1). Much of the work in this area documents several limitations, including:

- Potential for bias
- Small sample sizes
- Questions that may elicit answers based on factors other than teaching effectiveness
- Problems with averaging student scores

While some studies failed to find evidence of bias in student feedback (e.g., Benton and Cashin 2011; Benton and Ryalls 2016; Winer, DiGenova, Costopoulos, and Cardoso 2016). several others have found evidence of gender-, age-, and ethnic-based bias (Smith 2007; Stark and Freishtat 2014, MacNell, Driscoll, and Hunt 2015; Boring, Ottoboni, and Stark 2016; Mengel, Sauermann, and Zölitz 2018). Indeed, a recent *Inside Higher Ed* opinion piece points out that the evidence for bias is significant and that institutions that use SETs in academic files may face legal challenges (Owens 2019). Stark and Freishtat 2014 discuss the issue of sample size; with response rates lower than 50%, extrapolation is problematic. This same work warns against reporting averages of student scores, as these obscure the range of responses. Finally, many of the above works note that certain questions may elicit answers that do not always bear on teaching effectiveness. For example, student comments on instructors' command of material and personal characteristics may be based on a variety of factors, as may be their responses to general questions regarding whether they recommend the course and/or instructor.

While the literature is not unanimous in its assessment of student feedback, there is certainly cause for caution. Nonetheless, student input does provide a potentially valuable source of feedback on teaching effectiveness; after considerable discussion, the workgroup agreed that it should be included as one of several data points.

Bearing the above in mind, the workgroup reviewed questions used in other universities and developed a bank of potential questions for student feedback on teaching; these are presented in Appendix B. The workgroup envisions that units might tailor questions from this bank for their own use; in particular, some questions are flagged as optional because they may not be applicable to all courses. In addition, each question is recommended as being for assessment versus evaluative purposes. That is, some questions are included for background material and/or for instructors to use as feedback; they would not necessarily be included in academic files (although instructors could elect to include them). The evaluative questions would be included in files. The bank does not include general "do you recommend ..." questions. The literature suggests that these questions are particularly subject to bias. Omitting them will also guard against the tendency to look for a teaching bottom line, based on two (often one) average scores.

The bank is divided into four categories:

Student participation: These questions probe student participation in terms of attendance, office hour uses, etc. All of the questions are for assessment, as this is useful feedback for instructors, but does not directly measure teaching effectiveness.

Practice: This section pertains to the execution of the course – they focus on the instructional activities. All of these are for evaluative purposes, but they are also all optional because the instructional activities will vary from course to course.

APPENDIX 3 - UC SAN DIEGO

Student Learning: These questions – all evaluative - they ask students to evaluate how the course helped them learn and engage in the material.

Structure/Inclusiveness: The questions in this section ask about ways the instruction is structured and how students are included in the learning process. These are evaluative.

Additionally, the workgroup suggests several open-ended prompts to elicit student comments. As mentioned above, campus reviewers have found these comments particularly useful in providing context for numeric scores. Again, it is important that these prompts focus on pedagogical practices and avoid soliciting general comments that might be based on extra-pedagogical factors. The workgroup recommends that these comments be evaluative and be included in faculty files.

Also included in Appendix B are several questions regarding the effectiveness of instructional assistants, including open-ended prompts for comments. This type of feedback is crucial for improving the student experience and learning through the use of graduate and undergraduate student instructional assistants. Graduate students often rely on this feedback when going on the job market.

Finally, the group suggests that the campus develop a standard set of best practices to increase student participation, including, perhaps, opportunities to complete questionnaires in-class.

One way to mitigate bias in student feedback is to provide guidance for students, emphasizing the important role students play in commenting on and improving instruction. Several universities provide materials for students, including videos, instructions, etc. A succinct example, from the University of Michigan, is included in Appendix C. The workgroup recommends adapting this type of material and publicizing the importance of student feedback. Again, it is likely that few students realize that CAPEs are currently used in academic files. McGill University has a policy that allows student forms, if they contain hateful or discriminatory comments, to be removed from the instructors' portfolio (<https://mcgill.ca/mercury/about/equity>). This issue has been discussed with respect to graduate student instructional assistant evaluations; the workgroup recommends exploring such a policy for all student feedback. In particular, the workgroup recommends establishing a procedure whereby instructors can identify discriminatory comments and request removal.

The literature also discusses the problematic nature of averaged scores in student evaluations. Averaging potentially obscures the nature of variation in student responses. Therefore, the workgroup recommends that, instead of average scores, distributive histograms should be reported for each question.

As discussed at the beginning of this section, our campus – for historical reasons – has relegated the evaluation of teaching effectiveness to CAPEs. This is a student-developed vehicle that was never intended to be used in this manner. While student feedback does provide a valuable data point, it should be used in conjunction with others means of evaluation (see section 4, below). Furthermore, its limitations must be acknowledged, and steps should be taken to mitigate potential bias. The workgroup recommends replacing CAPEs with a feedback mechanism along the lines of the question bank in Appendix B. In doing so, this raises the question of what to do with CAPEs. It would be cumbersome and counter-productive to simultaneously run two student feedback vehicles. However, the importance of CAPE, as a student-run organization should not be diminished (although the student voice in the organization is now considerably reduced). One solution would be to simply replace the current CAPE survey with the new questions, maintaining the student worker and the faculty advisory committee.

APPENDIX 3 - UC SAN DIEGO

Recommendations:

- *Replace the existing CAPE questionnaires with the recommended questions in Appendix B; including questions regarding the instructor, instructional assistants, and prompts for student comments.*
- *Include questions used for both assessment and evaluation.*
- *Charge campus constituents to work towards practices that increase student participation (see section 5).*
- *Provide instructions to students that are designed to reduce bias and underscore the importance of thoughtful student evaluations of teaching.*
- *Provide a mechanism for instructors to report biased comments and for biased comments to be removed.*
- *Report the results of student evaluations of teaching with histograms.*

4. Holistic Evaluation

Challenge: *While the Academic Personnel Manual (APM) requires more than one measure of teaching effectiveness in academic files, departments lack guidance for providing meaningful evaluation. The result is an over-reliance on student feedback.*

Recommendation: *Establish that teaching effectiveness evaluation be holistically based on a teaching portfolio, an instructor's self-reflection, and the chair's summary.*

We have seen that assessment of teaching effectiveness is crucial for ensuring student learning and creating a student-centered university. The university is mandated to evaluate teaching effectiveness as part of the academic review process. Indeed, APM 210-1-D notes that "Clearly demonstrated evidence of high quality in teaching is an essential criterion for appointment, advancement, or promotion." This section of the APM goes into considerable detail about what constitutes effective teaching and how departmental recommendations should present teaching evidence; both the criteria set forth in the APM and instructions provide by the Committee on Academic Personnel align closely with the recommendations presented in this section.

The discussion of student evaluations of teaching noted the need to embed student feedback in a larger context – one that considers several aspects of pedagogical practice and one that allows faculty to demonstrate they are engaged in active pedagogy. Much of the literature emphasizes the importance of this multi-faceted approach to evaluating teaching effectiveness and advocates the use of teaching portfolios that detail the various means of engagement. However, an unstructured list of pedagogical practices will be cumbersome and difficult to interpret; campus reviewers need to have a guide. Hence, evidence for engagement in pedagogy and effective teaching should include both a teaching portfolio and a self-evaluation. The portfolio consists of documentation of pedagogical activities (see below); the self-evaluation provides a guide to the portfolio and a discussion of how the faculty member has engaged in pedagogy during the review period. Just as academic files are evaluated for engagement in research or creative activity, a fundamental criterion in evaluating teaching effectiveness is evidence of engagement in pedagogy. Thus, just as a well-written research statement contextualizes research or creative activity, a well-written teaching self-evaluation provides a guide to a teaching portfolio.

4.1. The teaching Portfolio

APPENDIX 3 - UC SAN DIEGO

Rather than relying almost exclusively on student feedback, the teaching portfolio provides an opportunity to document pedagogical activity in a more holistic manner. While student feedback is included, it provides only one set of data points among others. To better frame portfolio materials, we recommend grouping material according to the framework detailed in Natasha, et. al. 2018, which organizes the materials according to several categories. Note that not all portfolios will necessarily contain elements from each category. Nevertheless, APM 210-1-D requires more than one kind of evidence for teaching effectiveness and it is expected that a successful portfolio will contain several of the elements below. Some – e.g., course information and student feedback – will be routinely included, but candidates are encouraged to assemble substantial portfolios. Finally, it is important to emphasize the need for departments to establish standards on how to evaluate the quality of artifacts included in the portfolio.

A. Teaching and Supportive Learning

Because teaching effectiveness develops over time and is an on-going process, this section of the portfolio documents teaching activities and feedback. Elements may include:

- Course information (course number, enrollment, graduate/upper/lower division, general education course, enrollment numbers.)
- Syllabi
- Teaching awards
- Student feedback – only evaluative questions are required, but others may be included
- Reports or descriptions based on peer observations of instruction
- Examples of assignments/projects/examinations

B. Professional Learning and Development

Evidence of participation in professional development activities demonstrates engagement and self-reflection. This section of the portfolio may include:

- Pedagogy workshop attendance (e.g., the Commons, off-campus, and others)
- Pedagogy conference attendance
- Participation in professional development opportunities to perform peer or pedagogical expert review of teaching
- Participation in professional development opportunities on constructing teaching portfolios

C. Mentorship

Mentorship comes in many forms, including:

- Undergraduate student mentorship
- Graduate student mentorship
- Postdoctoral scholar mentorship
- Faculty mentorship, related to pedagogy
- Writing letters of recommendation written for all of the above
- Engaging in peer observation of instruction

APPENDIX 3 - UC SAN DIEGO

- Advising student organizations (as it relates to student learning)

D. Research, Scholarship and Inquiry

- Research on pedagogy – both general educational research and Scholarship of Teaching and Learning; this is particularly relevant for teaching professor files. Examples:
 - Using concept inventories or other tools (e.g., student portfolios) to get a sense of students' development of mastery during a course
 - Activities whereby an instructor compares different methods of teaching during one quarter or over several, to get a sense of which is more effective for their students
 - Formal pedagogical research done within the context of one's teaching of one's own courses
 - Formal pedagogical research done across at larger scale – might involve one's own classes, other faculty member's classes, other universities, collaborators...

E. Educational Leadership

This category includes activities that help advance the educational mission of the institution. While some activities overlap with university service, the following are examples of educational leadership that demonstrate engagement in teaching:

- Participation in assessment of program learning outcomes
- Teaching in colleges and inter-disciplinary programs
- Committee work that focuses on student-centered learning (e.g. CAMSEE)
- Developing a new course or initiative that addresses an identified need to better support student learning

4.2. Self-Reflection

The teaching portfolio documents activities related to pedagogy and student learning. The purpose of the self-reflection is to provide a guide to the portfolio and put the activities in context for campus reviewers. The point is to create a narrative that demonstrates engagement in teaching and teaching effectiveness. It is also an opportunity for faculty to discuss how they have used feedback and what steps they have taken to improve teaching. Finally, the self-reflection provides an opportunity to situate pedagogical activity in the context of a teaching philosophy. The self-reflection should be both specific and succinct – about 2-3 pages, although in the case of Teaching Professors, it may be more detailed.

Using the same model as above, a self-evaluation might address the following:⁴

A. Teaching and Supportive Learning

- What have you learned from student evaluations and how has student input affected your approach?

⁴ Note that not all categories will be relevant to all portfolios. The goal is for instructors to reflect on their teaching – what is going well and what can (or has improved).

APPENDIX 3 - UC SAN DIEGO

- How have your course structures evolved over the review period (use teaching load data and syllabi for support)?
- How has peer input affected your approach?
- Discussion of exceptional teaching loads (either high or low)
- Discussion of new or substantially reorganized courses; particularly when engaging inclusive practices
- Evidence of informal developmental activities (e.g., discussions with colleagues, independent reading ...)

B. Professional Learning and Development

- How has participation on pedagogy workshops, conferences, and other such activities influenced your approach to teaching?

C. Mentorship

- Describe the ways you have mentored students, instructional assistants, post-doctoral scholars, and faculty to improve student learning.
- Describe other ways you have supported students and others (e.g. letters of recommendation).

D. Research, Scholarship and Inquiry

- Discuss any pedagogical research/inquiry you have engaged in. What is its impact? How has it influenced your teaching?

E. Educational Leadership

- How have you supported the student-centered mission of the University (e.g., in assessment of learning outcomes, pedagogically related committee work, teaching in colleges and/or programs, etc.)?

4.3. The role of the department chair

APM 210-1-D notes the role of the department chair in evaluating teaching effectiveness:

“It is the responsibility of the department chair to submit meaningful statements, accompanied by evidence, of the candidate’s teaching effectiveness at lower-division, upper-division, and graduate levels of instruction.”

Both the portfolio and the candidate’s self-reflection provide the basis for a departmental discussion of this aspect of the file and the chair’s subsequent description. Thus, it will be critical for department chairs to provide evaluative summaries of faculty portfolio, aligned with departmental standards.

The portfolio, self-reflection, and chair’s letter provide three levels of evidence that feed into the review process. The portfolio materials and the self-reflection allow faculty to make use of assessment materials

APPENDIX 3 - UC SAN DIEGO

to create a narrative around their engagement in pedagogy. The chair's letter, which should reflect the departmental discussion, provides additional context (e.g., departmental standards). While campus reviewers will have access to all three levels, the guiding context will allow them to look beyond single numerical scores found in student feedback. This has the potential to allow for more holistic evaluation.

The portfolio model presents an opportunity to incentivize both student-centered teaching practices and various campus-wide pedagogical priorities. Currently, many faculty feel they are at the mercy of a single CAPE percentage (recommend instructor) and have little recourse to address low CAPE scores. This model will encourage faculty to view teaching holistically and to participate in professional development. It is also conceivable that even if student feedback remains somewhat negative, faculty may be able to mitigate this by showing engagement in other ways. Finally, the model provides incentives to participate in a variety of teaching-related activities – e.g., teaching in college courses and participating in assessment of learning outcomes; these are emerging campus priorities, but have not, to date, been sufficiently incentivized. The portfolio also provides faculty an opportunity to gain insight into their colleagues' course content and teaching practices.

Finally, the portfolio model will be particularly useful in evaluating files in the Teaching Professor series. This series places particular emphasis on teaching excellence. While many faculty in this series routinely receive strong CAPE scores, some, particularly those who teach large required (and sometimes unpopular) courses, do not. Nonetheless, these faculty are often dedicated to improving instruction, student learning, and are often at the forefront of pedagogical initiatives on campus. The portfolio model provides a means of rewarding pedagogical engagement in a manner that is much more nuanced and holistic than a single CAPE percentage.

Recommendation: Establish the teaching portfolio, self-reflection, and chair's summary, as described in this section, as the primary means of evaluating teaching effectiveness.

5. Guidance and oversight – next steps

Challenge: A shift to holistic teaching evaluation and sustained pedagogical assessment requires cultural change, guidance, and oversight.

Recommendation: Facilitate collaboration between departments, campus reviewers, and the Teaching+Learning Commons to provide guidance and support for assessment and holistic evaluation. Appoint a standing committee to provide oversight over the evaluation process.

The workgroup was charged with discussing training, incentives, and cultural practices. There was only cursory discussion of these topics in the time allotted. Therefore, these will await future conversations. Nevertheless, we make a few recommendations and observations.

The Teaching+Learning Commons is already providing considerable support in a wide array of pedagogical practices and is the source of considerable expertise in the area of the assessment of teaching effectiveness. If the portfolio model were to be adopted, the Commons would be a likely partner in guiding faculty, chairs, and reviewers in the mechanics of assembling, describing, and evaluating portfolios. One way to accomplish this would be to bring the Commons, Academic Personnel Services, and the Faculty Director for Faculty & Leadership Development together to create training materials and events. Faculty, chairs, and campus reviewers would all benefit from a systematic presentation of the portfolio model and how it can be used in both assessment, pedagogical improvement, and evaluation. Campus reviewers

APPENDIX 3 - UC SAN DIEGO

include members of the Committee on Academic Personnel; this committee should include training in portfolio evaluation as part of its on-boarding process.

The Commons has already worked with faculty to develop peer observation protocols and tools; the adoption of the portfolio model provides synergy to these efforts.

Students should be informed about their important role in teaching assessment and evaluation. The Commons is well-positioned to develop materials describing responsible student evaluations of teaching, including discussions of bias. If the administrative structure of CAPE remains the same (a student director advised by a faculty committee and the Office of Undergraduate Education), these advisory bodies should play a more active role in publicizing the importance of student evaluations and implement ways to increase student participation.

Adopting a portfolio model will require new practices and cultural change. It is important that there be sufficient oversight to the process to ensure a smooth transition and to identify any unintended consequences. We recommend paneling a senate-administration committee to monitor the assessment and evaluation of teaching effectiveness. It is possible that this group might merge with the CAPE Advisory Committee and form a standing committee on teaching effectiveness.

Recommendations:

- *Create a collaboration between the Commons, Academic Personnel Services, and the Faculty Director for Faculty & Leadership Development to oversee guidance and support for the portfolio model.*
- *Use the CAPE advisory structure to train students on the importance of student evaluations of teaching.*
- *Set up a committee to oversee pedagogical assessment and evaluation. This committee might subsume the functions of the CAPE Advisory Committee.*

References

Benton, Stephen L. and William E. Cashin. 2011. IDEA Paper No. 50: Student ratings of teaching: A summary of research and literature. Manhattan, KS: The IDEA Center.

Benton, Stephen L. and Kenneth R. Ryalls. 2016. IDEA Paper No. 58: Challenging misconceptions about student ratings of instruction. Manhattan, KS: The IDEA Center.

Boring, Anne, Kellie Ottoboni, and Philip B. Stark. 2016. Student evaluations of teaching (mostly) do not measure teaching. *ScienceOpen Research*.

Donald P. Hoyt and William H. Pallett. 1999. IDEA Paper No. 36, Appraising teaching effectiveness: Beyond student ratings. Manhattan, KS: The IDEA Center.

MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt. 2015. What's in a name: exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40, 291-303.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2018. Gender Bias in Teaching Evaluations. *Journal of the European Economic Association*, 17.2, 535-566.

Natacha, Kenny, Carol Berenso, Cheryl Jeffs, Lorelli Nowell, and Kimberley Grant. 2018. *Teaching Philosophies and Teaching Dossiers Guide*. University of Calgary Taylor Institute for Teaching and Learning.

APPENDIX 3 - UC SAN DIEGO

Owens, Ann. 2019. The next lawsuits to hit higher education. *Inside Higher Ed*, June 24, 2019.

Smith, Bettye P. 2007. Student ratings of teaching effectiveness: An analysis of end-of-course faculty evaluations. *College Student Journal*, 41, 788-800.

Philip B. Stark and Richard Freishtat. 2014. An evaluation of course evaluations. Published in *ScienceOpen Research*.

Winer, Laura, Lina DiGenova, Andre Costopoulos, and Kristen Cardoso. 2016. Addressing common concerns about online student ratings of instruction: a research-informed approach. *Canadian Journal of Higher Education*, 46.4, 115-131.

Membership

- John Moore, Dean of Undergraduate Education, Workgroup Co-Chair
- Gail Heyman, Professor, Psychology, Co-Chair
- Janet Becker, Teaching Professor, SIO & MAE
- Karen Christman, Associate Dean for Students, Jacobs School of Engineering
- Becky Petitt, Vice Chancellor for Equity, Diversity & Inclusion
- Samuel Rickless, Professor, Philosophy
- Carolyn Sandoval, Associate Director, Teaching+Learning Commons
- Kuiyi Shen, Professor, Visual Arts
- Haim Weizman, Teaching Professor, Chemistry & Biochemistry
- Gabriele Wienhausen, Faculty Director, Teaching+Learning Commons
- Paul Yu, Interim Dean, Graduate Division

APPENDIX 3 - UC SAN DIEGO

Appendix A - Senate-Administration Workgroup on Holistic Teaching Evaluation Charge May 2018

In accordance with Academic Senate policies this is a joint Senate and Administration committee, which will be co-chaired by an administrator and a Senate member.

In recent discussions between the Executive Vice Chancellor (EVC) and the Committee on Academic Personnel (CAP), and between the EVC and divisional deans and department chairs, several observations about our current teaching evaluation practices were raised:

- Teaching evaluation on our campus presently appears to rely primarily on the Course and Professor Evaluations (CAPE) scores/comments. The CAPE instrument was developed by students for use by students, is not a research-validated instrument like the Student Assessment of their Learning Gains (SALG), and so is not a satisfactory tool.
- Although the APM requires that at least two methods be used to evaluate teaching in personnel reviews, this is not uniformly observed or undertaken in sufficient depth to be informative to CAP reviewers.
- Individuals appointed in the Security of Employment/Teaching Professor are primarily evaluated on the basis of their teaching performance, and the number of individuals employed in this series is increasing. We need to ensure that our campus employs an appropriate suite of methods for evaluating their work.
- Proposed [changes to APM 285](#) related to the Security of Employment/Teaching Professor rank/step, would, if implemented, require more robust ways of evaluating teaching and learning for the 2018-19 evaluation cycle.
- The WASC reaccreditation visit in 2019 will expect us to demonstrate that we are assessing teaching, learning, and program impact in compelling ways.

In light of these challenges, we have determined that there is an opportunity now to rethink our approach to teaching evaluation. The overarching goal is to develop a holistic, transparent, and flexible evaluation process and to provide the tools and support for faculty and CAP reviewers that will lead to more effective evaluations. The Workgroup is charged to undertake a study and provide a set of recommendations that will enable our university to:

- Identify and make available multiple reliable existing tools for teaching evaluation and also provide training on how to use and interpret them. It is conceivable that different units may deploy different combinations of tools, depending on what is appropriate to the discipline, curriculum, teaching method, etc.
- Establish a campus culture where both formative (critique to help one improve one's teaching on an ongoing basis) and summative (evaluative) assessment of teaching and learning are standard practice. This would, for instance, encourage faculty statements

APPENDIX 3 - UC SAN DIEGO

in CAP files to comment on what the faculty learned from the formative assessments, and how they were adjusting their teaching practice to keep improving student learning. An additional benefit is that scholarly approaches to teaching and judicious experimentation would be seen as valuable and as safe to undertake, without the current worries about the potential impact on CAPE scores that tend to stifle innovation.

- Institute or augment faculty development programs that prepare both new and more experienced faculty to create high-quality teaching materials and learning plans. Use of rubrics and templates to help with this will ensure evaluation of teaching is based around those rubrics, for consistency and alignment.

We encourage the Workgroup to think expansively. In particular, we ask that it:

- Examine the literature and consider best practices from peer institutions;
- Review current UC San Diego resources and practices; explore the expertise within the Teaching & Learning Commons, Rady School of Management and Jacobs School of Engineering;
- Determine what new resources, tools, and/or training programs are needed and how to coordinate them with each other and with what already exists;
- Review incentives (e.g., use of teaching in merit reviews, campus awards, external grant proposals) and see what others might be valuable, and whether existing incentives are aligned with our goals;
- Consider cultural factors: which units and/or faculty groups are already leaders in this area? Which can help others make progress? Which are fearful or resistant to discussing these issues?
- Consider a communications strategy: How can we best make case that change benefits students, faculty and the institution as a whole?

The Workgroup should consult broadly, seeking input from Student Affairs, Information Technology Services (ITS) and faculty and staff colleagues with experience in this area. To assist the group's work, attached is a list of resources that may prove helpful. The Workgroup's efforts will inform the campus' WASC reaccreditation self-study.

MEMBERSHIP

Administration

- John Moore, Dean of Undergraduate Education, Workgroup Co-Chair
- Karen Christman, Associate Dean for Students, Jacobs School of Engineering
- BeckyPetitt, Vice Chancellor for Equity, Diversity & Inclusion
- Kit Pogliano, Dean, Graduate Division
- Gabriele Wienhausen, Faculty Director, Teaching & Learning Commons [or her designee]

APPENDIX 3 - UC SAN DIEGO

Senate

Five members, including a member with CAP experience, to be nominated by the Committee on Committees and appointment by the Senate Chair.

TIMELINE

- Spring Quarter 2018: Workgroup meets
- Summer 2018: Workgroup report drafted
- Fall 2018: Final report submitted October 1, 2018

Attachments: List of Holistic Teaching Evaluation Resources

IDEA Center Paper # 36 “Appraising Teacher Effectiveness: Beyond Student Ratings”

IDEA Center Paper #50 “Student Ratings of Teaching: A Summary of Research and Literature”

APPENDIX 3 - UC SAN DIEGO

Appendix B - Student Evaluation Question Bank

1. Student Participation

How often did you attend scheduled classes?	(Assessment)
How often did you engage in podcasts?	(Assessment, Optional)
How often did you engage with other course materials (e.g., readings, lecture notes, course videos, ...)?	(Assessment)
How often did you attend the instructor's office hours?	(Assessment)
Hours per week of work outside of class: 1-5, 6-10, 11-15, more	(Assessment)

2. Practice (e.g., course materials, exams, presentations, lectures)

How well did lectures help you understand the course substance?	(Evaluative, Optional)
How well did the in class activities help you understand the course substance?	(Evaluative, Optional)
How well did the assigned reading help you understand the course substance?	(Evaluative, Optional)
How well did the assignments help you understand the course substance?	(Evaluative, Optional)
Students were encouraged to participate in class discussions.	(Evaluative, Optional)
Do exams emphasize understanding?	(Evaluative, Optional)

3. Student Learning

I found feedback helpful to understand how to improve.	(Evaluative)
Expected grade in the course	(Evaluative)
The course developed my abilities and skills for the subject.	(Evaluative)
My interest in the subject has increased as a consequence of this course.	(Evaluative)
I have found the course intellectually stimulating and challenging.	(Evaluative)
The course material was presented to include relevant applications/current problems in this field.	(Evaluative, Optional)
The instructor helped me develop a sense of how knowledge/new ideas are developed and incorporated into the field.	(Evaluative, Optional)

4. Structure/Inclusion

The instructor effectively used communication tools and on-line course management systems to facilitate student learning.	(Evaluative, Optional)
The instructor used multiple teaching methods to engage me in the material (e.g., class discussions, in class problem solving, small group work, clickers/polling software, lectures, ...).	(Evaluative)

APPENDIX 3 - UC SAN DIEGO

The instructor provided a syllabus detailing course goals, structure and expectations. (Evaluative)

I felt there were ways to get help, if needed. (Evaluative)

This instructor provided opportunities for me to learn with and from other students in the course. (Evaluative)

The instructor created a learning environment that welcomed and supported me. (Evaluative)

The instructor's teaching strategies acknowledged and valued differences, including differences of opinion. (Evaluative)

5. Open-ended prompts for student comments

Please describe any specific teaching practices that your instructor used that particularly helped you to learn the material and/or develop your own critical perspectives on the material.

Please describe any specific teaching practices that your instructor used that helped you to feel engaged with the course material or that encouraged you to feel that you could succeed in the course.

Please describe any specific teaching practices that were less helpful for your learning, or offer constructive suggestions that might improve their effectiveness.

6. Instructional assistant evaluation

How often did you attend sections (please name the instructional assistant)?

How well did section discussions/activities help you understand the course substance?

I found instructional assistant's feedback (in class, on assignments, exams, term papers, presentations, etc.) helpful to understand how to improve.

The instructional assistant communicated effectively in section.

The instructional assistant communicated effectively in office hours.

Open-ended prompts to elicit comments:

Please describe any specific teaching practices that your instructional assistant used that particularly helped you to learn the material and/or develop your own critical perspectives on the material.

Please describe any specific teaching practices that your instructional assistant used that helped you to feel engaged with the course material or that encouraged you to feel that you could succeed in the course.

Please describe any of your instructional assistant's teaching practices that were less helpful for your learning, or offer constructive suggestions that might improve their effectiveness.

APPENDIX 3 - UC SAN DIEGO

Appendix C - Course Evaluations: Providing Helpful Feedback to Your Instructors

Instructors often find students' written comments the most valuable element of course evaluations. To help your teachers get the most out of your end-of-term feedback, please keep the following in mind:

- Remember that you are writing to your instructor. Your feedback can valuably influence the ways they teach this course and others in the future. (Unlike an online review site like "Rate My Professor," this is not a forum for saying whether or not you recommend a course to other students.)
- Specific constructive suggestions that focus on your learning are far more useful than general praise or critiques. See below for examples of ways you can provide feedback that helps instructors understand how their instructional choices facilitated or hindered your learning. Both positive and negative feedback is most helpful when very specific.
- Comments that are not related to your learning diminish the value of your feedback. For example, it is not helpful to comment upon an instructor's appearance or to include personal insults in your feedback.

Some examples of constructive feedback:

Less helpful: Vague critique or praise	More helpful: Specific suggestions that could improve your learning, or explanations of why the course helped you learn
"The professor just lectures."	"The professor just lectures... "...and a short break would help me pay attention for the full lecture." "...but we need more time for student questions during lectures." "...and I would learn more if I got more hands-on practice."
"The readings were redundant."	"The readings were redundant..." "...I didn't understand why we read so many different articles on the same topic." "...Could you offer more guidance on what we're supposed to look for in the readings?"

APPENDIX 3 - UC SAN DIEGO

“Discussions were awesome!”	“Discussions were awesome!...” “...I loved how the prof created an environment where students were willing to share perspectives and disagree.” “...It was really helpful that you kept notes on the board during our discussions.”
-----------------------------	---

University of Michigan Center for Research on Learning and Teaching (CRLT)

APPENDIX 4 - UCAADE

UNIVERSITY OF CALIFORNIA, ACADEMIC SENATE



BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY COMMITTEE ON AFFIRMATIVE ACTION, DIVERSITY,
AND EQUITY (UCAADE)
Mona Lynch, Chair
lynchm@uci.edu

ACADEMIC SENATE
University of California
1111 Franklin Street, 12th Floor
Oakland, California 94607-5200

May 7, 2020

Dear Teaching Evaluation Task Force:

I am writing on behalf of the University Committee on Affirmative Action, Diversity, and Equity (UCAADE) to share our comments on the draft “Recommendations for Evaluating Teaching Effectiveness” from the UC Teaching and Learning Group (Nov. 1, 2019) and provide initial input on the work of the Teaching Evaluation Task Force. We look forward to continuing to provide feedback and partnership on the critically important issue of how faculty are evaluated in their teaching assignment.

We were impressed with the scope and depth of the report, and generally endorse the recommendations. We have a few additional suggestions we would like to make, and several points of emphasis related to UCAADE concerns, delineated below.

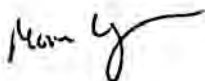
1. In concurrence with the report, we recognize and are concerned with patterns of bias in student evaluations as a function of instructor gender, race, ethnicity, nationality, and other identities and statuses. We very much appreciate the efforts to mitigate these biases through the design of the evaluations; the language used to characterize the evaluation process and goals; the development of procedures to respond to biased evaluations; and the recommendation to educate students about bias in evaluations. To that end, we suggest that students receive some general training about implicit bias and other forms of bias and that this training assignment does not fall on instructors themselves who are being evaluated.
2. We also wanted to voice support for the role of well-designed student evaluations, especially because they can provide students with an avenue to express their concerns about potential instructor bias or insensitivity to issues of equity, diversity and inclusion in the teaching context. Therefore, we do not recommend eliminating student evaluations as one form of teaching assessment.
3. We recognize the value of having high response rates for a complete and accurate assessment of the student experience, and that mitigates the problem of bi-modal extremes, and were happy to see that addressed in the report. We would encourage further development of strategies to improve response rates on student evaluations. Some strategies, such as rewards for completing evaluations, may have negative impacts, so consideration of possible costs to strategies would be warranted to ensure not only higher response rates but meaningful evaluations.

APPENDIX 4 - UCAADE

4. In regard to the nature of bias, we recommend that the report broaden the language used to more fully capture the kinds of bias that may be systemic. For instance, gender bias is not limited to that experienced by those identifying as women; there is also gender-based bias that can happen as a function of other gender and/or affectional orientation identities. Similarly, bias may happen over other identities, including national and religious identities, as well as statuses such as age, that are not identified in the report.
5. We wish to underscore, as well, the variation that can occur in student evaluations due to the nature of the course, and recommend that course type be an explicit context in which evaluations are considered in review processes. For instance, some very challenging courses and some core requirement courses generate lower scores than more specialized, substantive-area courses. In regard to our committee's purview, we note that courses that include topically challenging material, including those that grapple with issues around race, ethnicity, gender, and other such topics, can and do produce some extreme bi-modal scores, and/or lower overall scores. Such courses, which often fill GE breadth requirements, are also disproportionately likely to be taught by faculty from underrepresented groups, so the consequences of the lower scores often disproportionately fall on those faculty.
6. More generally, we would like to push the Teaching Evaluation Task Force to develop more specific recommendations to campus CAPs about how to use SET data in the merit and promotion process, including within a given case, as well as across cases. We strongly endorse the recommendation that CAPs do not compare across instructors, and would like to see more advice about how that might be achieved, particularly when units continue to report department- and school-level means.

Thank you for the opportunity to comment on this report.

Sincerely,



Mona Lynch
Chair, UCAADE