UC RIVERSIDE Academic Senate

May 5, 2021

Dear Colleagues,

Since Spring 2018 and at the request of then Interim Vice Provost for Undergraduate Education Richard Cardullo, a Senate ad hoc committee has met to evaluate the use of the student teaching evaluations. I want to take this opportunity to express gratitude to Chair David Nelson and the entire committee who completed this work partially through a pandemic. Their attached report and recommendations represent thoughtful and thorough examination of the use of student teaching evaluations and articulate a vision for solutions that begin to address biases in the system.

Senate and faculty executive committees reviewed the report and have provided additional context, suggestions, and support for the propositions and substance of the report. Importantly, as was raised in the Senate review, maintaining these evaluations was seen as an important aspect of giving students a voice in the evaluation of the instruction they receive, but that biases in evaluation which unequally impact our faculty need to be mitigated in how these evaluations are used in career advancement reviews.

Our next steps will be to plan changes informed by the report through coordination with the Provost, Associate Provost, Vice Provost and Dean for Undergraduate Education, Vice Provost for Academic Personnel, and Office of Assessment. I will continue to communicate proposed plans of change to the Student Evaluations and seek input from Senate committees as the plan is crystalized.

Jason

Attached: Final Report from the Ad Hoc Committee on Evaluation of Teaching and related committees' feedback

UC RIVERSIDE Academic Senate

AD HOC COMMITTEE ON EVALUATION OF TEACHING

January 20, 2021

To: Jason Stajich, Chair
Riverside Division

From: David Nelson, Chair
Ad Hoc Committee on Evaluation of Teaching

Re: Final Report of the Ad Hoc Committee on Evaluation of Faculty Teaching

Dear Senate Colleagues:

The Ad Hoc Committee on Evaluation of Teaching was convened in February 2018. The committee was charged with reviewing the current policies, procedures, and mechanisms for how UCR faculty members are evaluated for their teaching and was tasked with the following:

1. Establish policies and procedures for accurate and reliable evaluation of faculty teaching in undergraduate, graduate, and professional courses. Recommendations should take into account the Academic Personnel requirements outlined in APM 210-1-d.1 that requires more than one kind of evidence of teaching evaluation in each faculty review file.

2. Review iEval, used as the primary student evaluation instrument of faculty members of teaching, for its efficacy in all courses. If appropriate, recommend changes that maximize student participation and provide faculty members with constructive feedback to improve their teaching. Any recommendations on evaluation instruments should take into account appropriateness of course level as well as differences in pedagogical approaches (i.e., traditional lectures, online offerings, hybrid classes, etc.).

In 2020, in coordination with the Director of Evaluation and Assessment, Omar Safie, we conducted a survey of the experiences and opinions of UCR faculty with the current iEval system. This survey indicated widespread dissatisfaction with the current system and its use as a measure for teaching effectiveness (see Appendices B-E of Subcommittee A report).

Here we present the recommendations that we have developed in response to the Senate's charges, having considered the concerns of UCR faculty and well-documented problems with student evaluations of teaching. Our recommendations are detailed in the attached summary reports from two subcommittees. Subcommittee A was tasked with proposing revisions to iEval that could make it a more fair and effective instrument. Subcommittee B was tasked with providing recommendations of how the campus should interpret and responsibly use teaching evaluations for merit and promotion decisions.

While our work was underway, similar reviews of problems and best practices for student evaluations of teaching were undertaken by the Academic Council Teaching Evaluation Task Force and the UC Centers for Teaching and Learning. As detailed in our last communication with the Senate, we have taken their reports under consideration and note that in general we have come to similar conclusions. We highlight several particularly relevant recommendations from the 2019 report from the UC Centers for Teaching and Learning report in Appendix A of the Subcommittee A report.

On the following page we propose a two-phased approach to improve our current system of teaching evaluation that draws from our Subcommittees' reports. Phase I consists of several stopgap measures that we anticipate could be implemented as soon as this current winter quarter. Phase I recommendations are not intended to be long-term solutions in themselves, and Phase II will be required to bring about meaningful change.

Sincerely,
The Ad Hoc Committee on Evaluation of Teaching

Chair, David C Nelson, Botany and Plant Sciences
Michelle E. Bloom, Comparative Literature and Languages
Kevin Costello, Mathematics
Long Gao, School of Business
Jean Helwege, School of Business
John M. Heraty, Entomology
Jennifer Hughes, History
John Levin, Graduate School of Education
Suveen Mathaudhu, Mechanical Engineering
John Stamp, University Writing Program
Ex Officio Jennifer Lynn Brown, Undergraduate Education (Vice Provost for Undergraduate Education)
Ex Officio Richard A. Cardullo, Evolution, Ecology, and Organismal Biology
Ex Officio Daniel Jeske, Statistics (Vice Provost of Academic Personnel)
Ex Officio Mariam B. Lam, Comparative Literature and Languages (Vice Chancellor of Diversity, Equity, and Inclusion)
Ex Officio Josh Bright, Interim Associate Vice Chancellor Chief Information Officer

Phase I

- Add a preamble about bias in student evaluations of teaching (SETs) at the beginning of the iEval form.
- Replace iEval question #5 ("I normally spent at least two hours preparing for each hour of class"), which was considered one of the least informative iEval questions by the committee, with "I have considered the possibility that personal biases around race, gender identity, age and sexual orientation may impact my responses to this survey."
- Stop including department and campus comparative statistics (percentiles, mean, median, SD) in the numerical iEval report.
- Communicate to UCR faculty and P&T evaluating bodies the importance of additional forms of evidence of teaching effectiveness that supplement the iEval SET. This should include a reminder of the requirement in APM 210-1-d.1 that "More than one kind of evidence [of teaching effectiveness] shall accompany each review file." Although "opinions of students" is only one of the significant types of evidence listed, it dominates our current P&T evaluation process. Specify different forms of evidence of teaching effectiveness that will be considered acceptable by evaluating bodies.
- In coordination with the previous recommendation, modify eFile so that faculty can easily include different types of evidence of teaching effectiveness. That is, add new categories of evidence to eFile that can be selected and uploaded. The current designation of "Other Teaching Info" implies that alternative forms of evidence, which must further be categorized as "Other" Activities, do not carry as much weight as SETs listed under Teaching Records. More explicit instructions on what constitutes useful information in a self-reflective Teaching Statement could also be provided.

Phase II

The next phase will entail the re-design of the current student evaluation of teaching to produce a more equitable and useful tool for evaluation of teaching effectiveness and pedagogical improvement. In the attached reports from Subcommittees A and B, we have made a set of recommendations that provide guidance on issues that should be addressed in an iEval revision and in the continued use of student evaluations of teaching for promotion and tenure decisions. These recommendations are largely congruent with the 2019 UC Centers for Teaching and Learning Report, which should also be taken into consideration. Although we have provided examples of how the evaluation could be modified, this committee recommends that the exact wording and selection of questions should be designed by experts in pedagogy based upon current research.

SUBCOMMITTEE A (revision of iEval student evaluation of teaching)

Charge

Propose revisions to iEval to become "*an instrument that is more closely tied to student learning, rather than just satisfaction, and that is tailored to the specific class and its format, teaching practices, and learning goals. … to develop a course evaluation that provides better information to faculty, encourage students to reflect more thoughtfully about their educational experiences and enhances the learning partnership*" (CEC report, Stanford 2013)

Background and Process

The format of UCR's Student Evaluation of Teaching has been unchanged for almost fifteen years, since 2006, with the exception that in the fall of 2009 UCR Academic Senate approved that assessment would be moved online. The Senate formed an Ad Hoc committee to examine Student Evaluation of Teaching in 2015-16. That committee's recommendations were rejected by Senate Academic Council in Winter of 2016 as not being far-reaching or substantive enough in their proposed revision. A new Ad Hoc committee was formed in the spring of 2018 and has been meeting ever since. In 2019 the committee was divided into two subcommittees, with Subcommittee A dedicated to reviewing UCR's iEval instrument with a mandate to make suggestions so that it might become more closely tied to student learning.

Subcommittee A compared teaching evaluations from multiple campuses, studied the 2020 UC Teaching and Learning Centers Report (UCTLCR), and conducted a survey of faculty for opinions on our current iEval system (**see Appendices A-E**). We debated which questions were most helpful for providing formative feedback to instructors and evaluating teaching effectiveness, and weighed the balance of evaluating student satisfaction with the professor versus evaluating the learning experience (e.g. gains in knowledge, course content, and instruction).

Subcommittee A spent considerable time discussing the problem of bias against women faculty and faculty of color in student evaluations of teaching, both in the instrument itself and in the way that evaluations are deployed in merit and promotion.[1] We identified the problem of bias as among the most serious flaws in our current SET and its uses on our campus. UCR has the lowest percentage of women faculty in the UC system (at around 38%) and the percentage of women faculty at the rank of Full Professor is about 25% below the national average (the national average is only 25% and UCR is 25% below this). We discussed various mechanisms for minimizing bias.

With this in mind, we support the **UCTLCR** conclusions on bias (reproduced in modified version here)[2]: The research suggests that bias is pervasive and that women faculty and faculty of color are significantly disadvantaged in academic personnel reviews by these surveys, particularly when institutions rely heavily on quantitative

[1] https://www.insidehighered.com/news/2016/01/11/new-analysis-offers-more-evidence-against-student-evaluations-teaching
https://www.scienceopen.com/document/vid/818d8ec0-5908-47d8-86b4-5dc38f04b23e

[2] https://senate.universityofcalifornia.edu/_files/reports/kkb-divs-teaching-evaluation-task-force-report.pdf

measures to assess faculty performance. 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; Spooren 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).

We suggest the following immediate reforms:

- remove comparative statistics for Departments and Campus
- include a bias statement to be included at the beginning of evaluations

Overall Recommendations

Continue use of Student Evaluation of Learning

Our subcommittee agrees that student feedback is essential to teaching and learning on our campus and that there must be a mechanism for students to communicate their experience. This is particularly true at UCR where many of our students are first-generation, minority, and/or from socioeconomically disadvantaged backgrounds. These factors may shape the student learning experience in ways that are unfamiliar to many faculty. SETs can provide a bridge to communicate those challenges and experiences. We believe our students have an essential role to play not as consumers but as active learners who are engaged in their education. Their reported experience in the classroom is valuable information that can help support UCR's mission by influencing instructional approaches.

We therefore recommend the continued use of a substantively revised Student Evaluation of Learning survey instrument, while acknowledging considerable faculty frustrations with our SET to date. We believe that the SET can be improved to become more fair and useful, but acknowledge it will always be an imperfect instrument. Therefore, to reduce our campus' current overreliance on SETs for the purposes of merit and promotion, we also recommend that additional forms of evidence of teaching effectiveness be widely adopted.

Maintain numerical scales

Our subcommittee agrees that given the size of UCR's student body that our revised survey instrument must continue to include some scale-based questions, but we support the recommendations of the UC Centers for Teaching and Learning that these be changed to frequency scales (often/sometimes/rarely) or scales of agreement

(UCTLCR p6). We also recommend that our campus maintain the five-point scale, as evidence continues to suggest this is the most effective.[3]

However, in keeping with the **UCTLCR 2020 Report** we advise against using this data to create campus and department medians against which individual faculty are compared. UCTLCR recommends that campuses should not compare one instructor's scores to departmental averages, for example (p4). We refer to the guidance of Subcommittee B on this matter.

Substantially revise the current iEval survey

Our subcommittee recommends the substantive revision of our iEval questions both to make them a better measure of student learning outcomes and a more useful instrument to faculty as they strive to improve their teaching. Rather than simply removing problematic questions, each question adopted in our new system should reflect the best and most recent research. We endorse the guidelines in developing SET that appear on pp 2-3 of the UCTLCR Report.

Below we make nine specific recommendations for an iEval revision. Among these suggestions we provide several examples to illustrate what a modified iEval question might look like. However, we do not feel that our Ad Hoc committee has the research expertise to determine which questions are most effective measures and least prone to bias. We recommend the campus convene a group of faculty whose research is in the field of teaching, learning, education, and pedagogy to select, create, and/or identify the most effective questions drawing on the most current research.

1. Our campus should fully prepare and equip our students to participate in the evaluation of teaching as described in **UCTLCR 2020 Report** (p6)**.** This preparation should include providing information and resources to students to learn about the importance of student feedback, providing examples of helpful and unhelpful feedback, and creating multiple opportunities for students to learn about the importance of evaluations We add here that students need to learn about the problem of bias in SET.

2. There should be fewer and more precisely crafted questions. This may reduce evaluation fatigue, leading to higher participation rates and more meaningful responses from students. If a question does not ***often*** provide useful information to faculty or P&T evaluation committees, it should be removed to promote more careful consideration of questions with higher assessment value.

3. Those questions most prone to bias should be excluded. The current iEval questions #10, #11, #12, #13, and possibly others, fall into this category. Question #13 regarding the "overall ranking of the instructor" is especially of concern since many departments and P&T evaluation bodies rely on it centrally.

[3] Revilla, Melanie A., et al. "Choosing the Number of Categories in Agree–Disagree Scales." *Sociological Methods & Research*, vol. 43, no. 1, 2013, pp. 73–97., doi:10.1177/0049124113509605.

4. iEval questions should focus on student learning rather than instructor performance. **UCTLCR 2020 Report** provides examples of student experience centered questions that can help assess the effectiveness of a class. We note that a "Student Evaluation of Teaching" implies a one-way relationship in which the instructor delivers a product and the student evaluates their satisfaction as a consumer. We propose a transition to a "Student Evaluation of **Learning**" that aims to rebalance this relationship by acknowledging that learning is a two-party process. Some committee members felt that iEval questions #6-#13 fall into the category of instructor performance and should be excluded. Other committee members felt strongly that it was important to preserve some evaluation of instructor performance.

5. Consider including the instructor's stated learning outcomes or learning objectives for the course as a component of the student evaluation. For example:

   For each of the following course learning objectives (if provided), please evaluate the gains that you have made this quarter.
   Minimal=1 Moderate=3 Strong=5

   Course Objective 1: <list instructor-provided course objective here> 1 2 3 4 5

   Course Objective 2: <list instructor-provided course objective here> 1 2 3 4 5

   Course Objective 3: <list instructor-provided course objective here> 1 2 3 4 5

   Course Objective 4: <list instructor-provided course objective here> 1 2 3 4 5

6. Narrative comment fields should be retained, but the iEval survey should provide carefully crafted prompts that guide the scope of student comments rather than simply having an open-ended Comments field. This shifts the tone of the evaluation from a consumer satisfaction survey to a more formative evaluation, and may provide useful information for faculty to understand how students engage with the course. For example:

   What approaches or materials did you find most effective or successful in aiding your learning in this course? Please describe approaches that you used as well as those of the instructor.
   <short answer box>

   What changes could you have made and/or the instructor have made to improve your ability to achieve the learning goals of this course?
   <short answer box>

7. A statement or “preamble” about bias should be presented to students at the beginning of the survey (see example below). Evidence suggests that such statements are partially effective in mitigating bias. This change could be implemented easily and immediately as a stopgap measure. We recommend that additional strategies are developed to educate both students and faculty about the prevalence of bias in SETs.

   > *“Student evaluations play an important role in ongoing instructional improvement and institutional reviews of faculty teaching effectiveness. The University of California, Riverside recognizes that student evaluations of teaching may be influenced by unconscious and unintentional biases about the gender and race of the instructor. Women faculty and faculty of color are systematically rated lower in their teaching evaluations than white male faculty, even when there are no actual differences in instruction or what students have learned.*
   >
   > *As you fill out the course evaluation, please focus your opinions on the content of the course, the course materials, and the quality of your learning.”*

8. Consider using the Student Evaluation of Learning to gather information about other obstacles to learning our students face that are beyond the control of the instructor. This information would be shared with administration to guide campus development and resource allocation. It may also be used by P&T evaluation committees to identify factors that may unduly influence student perceptions of faculty teaching effectiveness in a course. For example:

   > **Was the university-provided infrastructure (for example, classroom facilities and equipment, external student resources, class size, number of teaching assistants) appropriate to promote positive learning outcomes in this course?**
   > <short answer box>

9. The placement of prompts for written comments relative to scale-based questions may influence numerical ratings. It is possible that providing written comments first, particularly if the student is invited to reflect on their role in the learning process, may provoke a more thoughtful and less biased consideration of scores on scale-based questions. Alternatively, specific criteria that are raised by scale-based questions may provide a useful framework for composing written comments. We recommend that a campus study be performed to evaluate this hypothesis.

Members of the Subcommittee A Task Force

John Heraty, Subcommittee A Chair, Professor, Entomology
Michelle Bloom, Professor, Comparative Literature and Languages
Kevin Costello, Associate Professor of Teaching, Math
Jennifer Scheper Hughes, Associate Professor, History
Suveen N. Mathaudhu, Associate Professor, Physics
Dave Nelson, Committee Chair, Associate Professor, Botany and Plant Sciences
David Gracey, Associate Vice Chancellor, Information and Technology Services
Stefano Lonardi, Professor, Computer Science & Engineering (2018-19)
Traise Yamamoto, Associate Professor, English (2018-19)
Teri Eckman, Interim Executive Director, Student Information Systems
Israel Fletes, Director of Academic Technology (2018-19)

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

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research.*

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.

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:https://www.scienceopen.com/document/vid/42e6aae5-246b-4900-8015-dc99b467b6e 4?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 A - Most Relevant Recommendations from the 2019 UC Teaching and Learning Centers Report

R1 Campuses should rely on multiple methods to evaluate teaching effectiveness, balancing student evaluation of teaching (SET) scores with other measures (i.e. peer observation, self-assessment).

- self-assessment should provide a reflection on their teaching development (focusing on growth rather than attainment).
- peer observations should be separated into assessment (internal) versus evaluation (external merit review). Note that peer observation is also subject to the same sort of unconscious bias that troubles SET.
- Encourage use of teaching portfolios.

R2 Campuses should adopt the broadest language possible to more fully capture the kinds of bias that may be systemic in teaching evaluations.

- Eliminate heavily gendered criteria such as “helpfulness” and “warmth” from survey questions.
- Change name to Student Experience of Teaching Surveys
- Bias training and education about/on the importance of the evals are important at all levels.
- Institute a process for faculty who wish to respond to discriminatory comments.

R3 Campuses should adopt a more developmental approach to teaching (i.e., create opportunities for faculty to experiment and innovate in their courses without fear of “failing,” rewards teaching professional development.

R4 Campuses need to be willing to spend more time evaluating teaching effectiveness and expect instructors to spend more of their (finite) time improving their teaching.

R5 Campuses should clarify what aspects of teaching SET the questions are attempting to evaluate.

- Rewrite SET questions to reflect student experiences in the classroom (“the instructor defined important concepts”)
- Shift from numerical scales to frequency scales.

R6 Specific questions that prompt students to consider different aspects of a course and its instruction should precede more global questions.

R7 SETs should include open-ended questions (comments) that ask students to describe their experience in the course, what aspects facilitated their learning, and how it could be improved

R8 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.

R9 Provide professional development opportunities for faculty to improve their teaching.

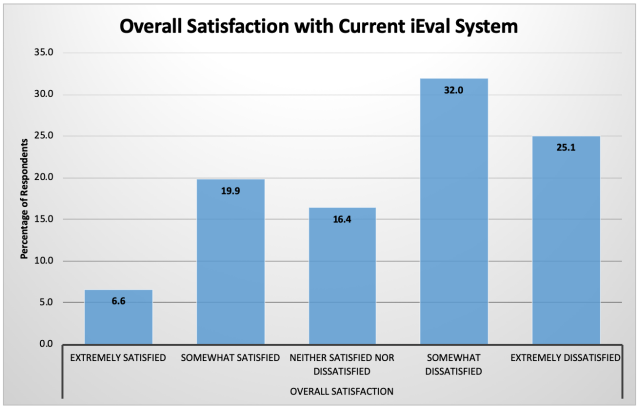
APPENDIX B - Summary of 2020 UCR campus survey on iEval

The iEval Faculty Opinion Survey was conducted by the Ad Hoc Committee on Evaluation of Teaching with the assistance of Dr. Omar Safie, Director of Evaluation & Assessment (*ex officio* member). The survey was distributed to all instructors at UCR, and was left open from May 18, 2020 to June 18, 2020, with two reminders sent to faculty who had not yet completed the survey. Respondents completed surveys anonymously and their responses will be kept confidential within the committee. A total of 351 respondents completed all or part of the survey.

The majority of faculty surveyed were from CHASS, followed by CNAS. CHASS and CNAS have the largest populations of majors and faculty at UCR. Combined, faculty from CHASS and CNAS comprise nearly 80% of survey participants. BCOE, GSE, SOB accounted for 16.3% of the responses. The faculty completing the survey were nearly evenly split between Man and Woman cisgender with a significant percentage of respondents declining to state.

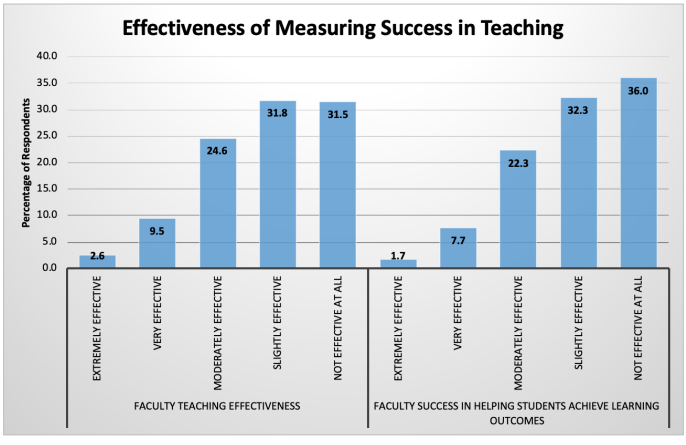
Based on IPEDS definitions of ethnicity and race, the majority of participants were not of Hispanic/Latino ethnicity and were majority White. However, it is important to note that a quarter of faculty completing this survey declined to state their ethnicity or race.

The majority of faculty indicated dissatisfaction with the current iEval system.



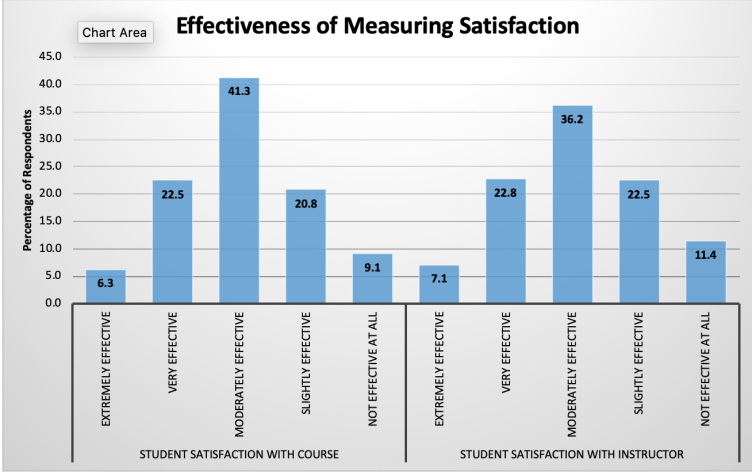
Effectiveness of Measuring Success in Teaching

In addition to their overall dissatisfaction with the current iEval system, a majority of faculty indicated that the current iEval system **was not very effective at measuring teaching effectiveness or faculty success** in helping students achieve learning outcomes. This indicates a need to refocus iEval on factors that measure teaching effectiveness and student success, such as student feedback on teaching methodologies and a focus on achievement of outcomes.



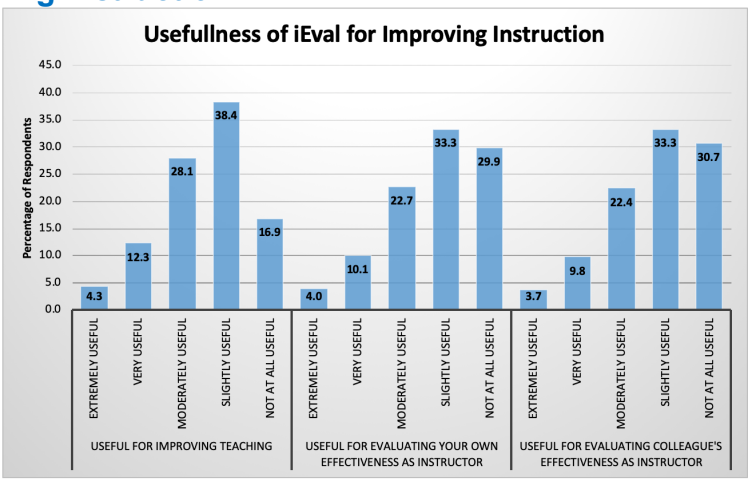
Effectiveness of Measuring Satisfaction

The largest percentage of faculty surveyed landed right in the middle, indicating the current iEval system was moderately effective at measuring satisfaction with the course or with the instructor



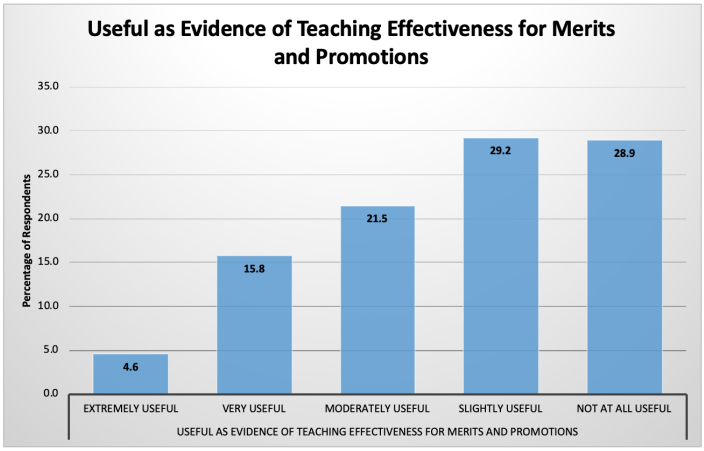
Usefulness of iEval for Improving Instruction

A majority of faculty did not find the current iEval system useful, or only slightly useful, in evaluating their own or a colleague's effectiveness as an instructor. While more faculty indicated that the current iEval system is moderately to extremely useful for improving teaching, a general review of comments indicated that this is dependent on the open-ended comments from students



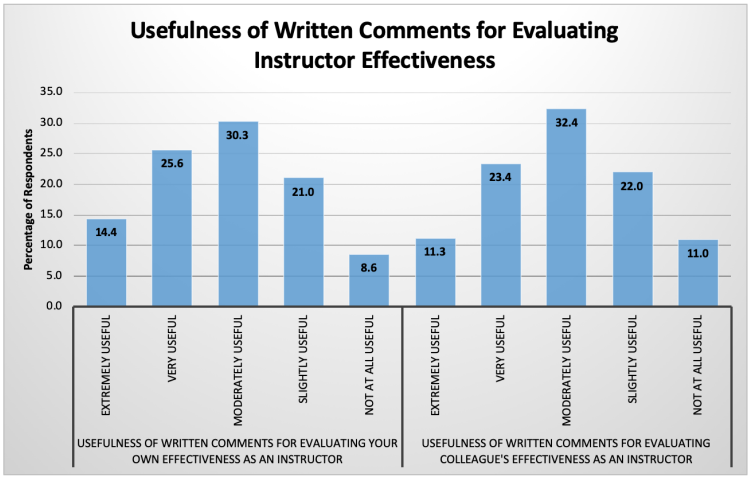
Usefulness as Evidence of teaching Effectiveness for Merits and Promotions

The majority of faculty surveyed indicated that the current iEval system was only moderately useful, at best, as evidence of teaching effectiveness for merits and promotions.



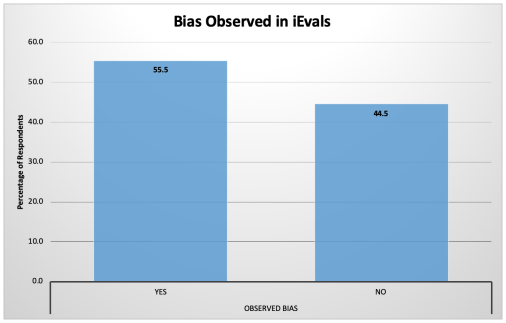
Usefulness of Written Comments for Evaluating Instructor Effectiveness

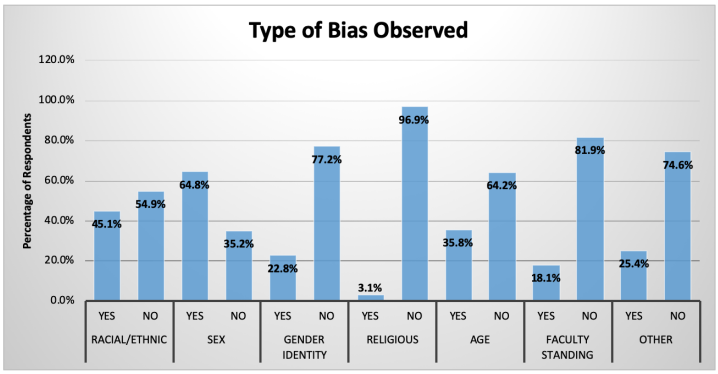
the majority of faculty indicated that they were moderately to extremely useful for evaluating their own effectiveness as an instructor. However, a nearly evenly split percentage of faculty indicated that the written comments were useful for evaluating a colleague's effectiveness as an instructor

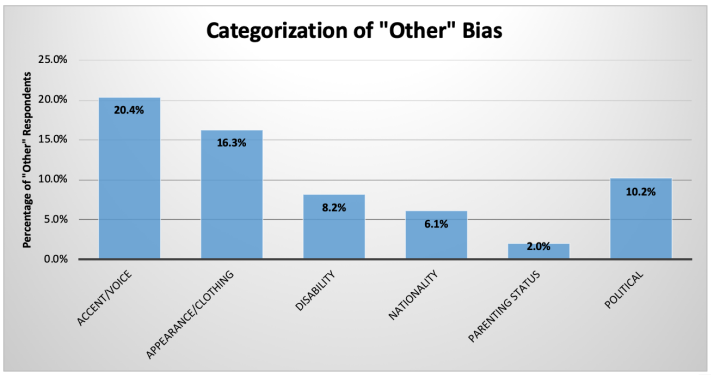


Bias in Current iEval

The majority of faculty indicated that they observed bias in the current iEval system.

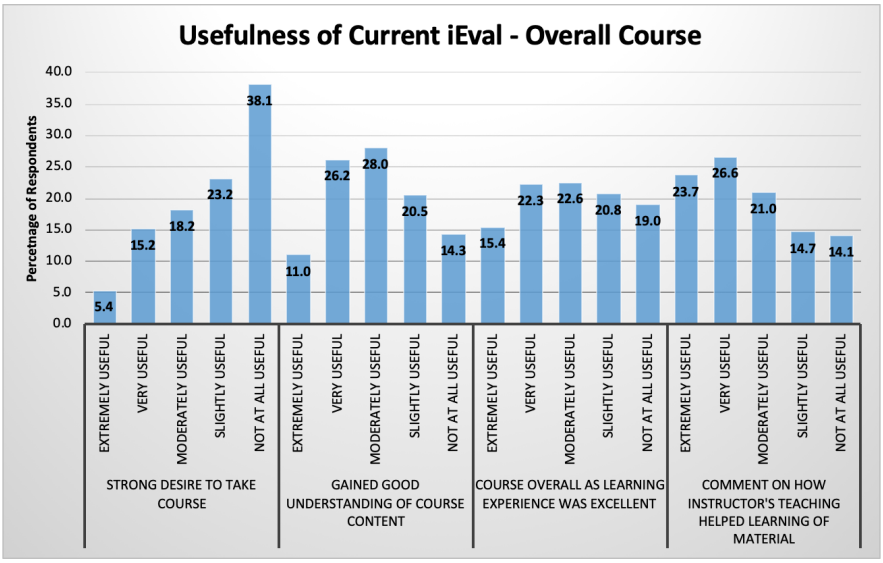






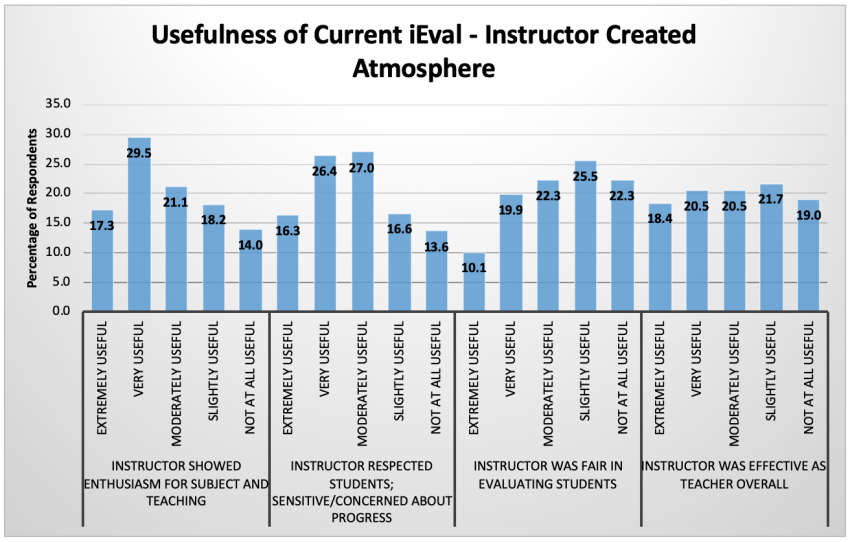
Usefulness of Current iEval – Overall Course Questions

The majority of faculty indicated that asking students if they had a strong desire to take a course is not at all useful or only slightly useful. Given that some courses are required for majors, such a question does not seem to be useful for faculty seeking information to improve their practice. In addition, asking students if a course was an excellent learning experience also may not be useful to faculty. However, more faculty found useful the questions focused on whether the student gained a good understanding of course content and the comments on how an instructor's teaching helped the learning of material. This lends support to a need to shift from satisfaction, general questions, and comparative rankings, to more focused questions on learning and outcome achievement that provide useful information for faculty to improve their instruction.



Instructor Created Atmosphere

Faculty surveyed indicated that faculty demonstrating enthusiasm for the subject and teaching was useful with nearly 30% indicating very useful. In addition, faculty surveyed also indicated that instructor respecting and being sensitive to and concerned about student progress were also important. However, nearly 50% of faculty surveyed indicated that the question related to instructor being fair in evaluating students was only slightly or not at all useful.



Overall Suggestions Based on Survey Results

The suggestions below are preliminary and are based on a combination of findings from this survey and the UC Centers of Teaching and Learning November 2019 report on Recommendations for Evaluating Teaching Effectiveness. These suggestions are meant to move the conversation forward. They are not conclusive nor are they prescriptive.

1. **iEval Content**
    a. Exclude questions that focus on satisfaction.
    b. Exclude questions that focus on fairness.
    c. Include questions that focus on instructor implementation.
    d. Include questions that focus on usefulness/applicability of course materials, but excludes exams.
    e. Include questions that focus on achievement of course outcomes.
    f. Include questions that focus on faculty created atmosphere.
    g. Include open-ended questions that are focused on improving instruction and not left completely open.

2. **iEval Use**
    a. Consider how iEval findings are used. Specifically, how should iEval findings be used as a way to evaluate teaching effectiveness and for merits and promotions?
        i. One solution is to ensure that iEval Content is focused on teaching effectiveness and not satisfaction.
        ii. Another is to allow faculty the opportunity to reflect on and speak to how they used iEval reports to improve their practice as instructors. This allows for faculty to share not only the context within which the course was implemented and how they managed to “close the loop.”
    b. Consider removing comparative rankings in the iEval report. Given the vast differences between courses across the campus and even within the same discipline, the usefulness of this aspect of the report is limited.

3. **iEval Implementation**

Consider providing a core set of questions for all faculty, as well as a set of questions that individual faculty or departments can “add” for their iEvals. This would allow for flexibility for faculty and programs to not a get a core set of results from iEval, as well as results that may be more relevant/useful to specific cases.

Appendix C - List of current iEval questions for context

1. I had a strong desire to take this course
2. I attended class regularly
3. I put considerable effort into this course
4. I gained a good understanding of the course content
5. I normally spent at least two hours preparing for each hour of class
6. Instructor was prepared and organized
7. Instructor used class time effectively
8. Instructor was clear and understandable
9. Instructor exhibited enthusiasm for subject and teaching
10. Instructor respected students; sensitive to and concerned with their progress
11. Instructor was available and helpful
12. Instructor was fair in evaluating students
13. Instructor was effective as a teacher overall
14. The syllabus clearly explained the structure of the courses
15. The examinations reflected the materials covered during the course
16. The required readings contributed to my learning
17. The assignments contributed to my learning
18. Supplementary materials were informative (e.g. films, slides, videos, demonstrations, guest lectures, iLearn, web pages, etc)
19. The course overall as a learning experience was excellent

Comment:

Appendix D - Some potential questions suggested from the faculty survey

- Describe aspect(s) of the course that you found beneficial to your learning.
- Describe one thing YOU (the student) can do to improve your learning experience in the course. Describe one thing I (the instructor) can do to improve your learning experience in the course.
- Did this course lead you to want to take more courses in this field of study?
- Have you grown in your knowledge of the course subject matter?
- How would you rate the overall course experience?
- How would you rate your own contribution to the course? Do you come on time, pay attention, stay off your phone, actively contribute to you own learning? Do you treat the learning environment with respect and professionalism?
- The instructor appeared to reflect and tried to adapt their teaching when things didn't go as planned or when students were struggling.
- What's Been most Helpful to Your Learning? Please select the teaching element that has been most helpful to your learning, and then provide a detailed written comment about what worked well and why. 1. The inclusiveness of this course. 2. Support from the instructor. 3. Feedback from the instructor. 4. The level of challenge of this course. 5. The quality of course materials. 6. The clarity of instructions for assignments and grading. 7.The use of active learning practices. 8. Interactions between students in this course. 9. Instructor communication in this course. 10. The organization of this course. 11. The relevance of the course content. 11. The assignments or projects in this course. 12. The accessibility of this course. 13. None of the elements above are helpful to my learning.
- Do you feel the course was a valuable experience in helping to develop your own thinking/direction/approach on the subject?
- I have considered the possibility that personal biases around race, gender identity, age and sexual orientation may impact my responses to this survey.
- The instructor explained concepts clearly.

Appendix E - Examples of negative comments provided in the faculty survey

- Abandon evaluations as other institutions have. They do not work.
- Allow ONLY positive student comments to be included in merit letters; no quoting of negative comments.
- Are you aware that faculty who are women and/or persons of color are rated lower across the board in faculty evaluations, nationwide?
- Get rid of evaluations. They are largely useless. And, since they are optional, they only take a small sampling from the class, which gives very unrealistic data.
- I do not believe our teaching should be numerically evaluated — every study indicates that these formats are particularly vulnerable to bias.
- I wouldn't recommend anything since the SET approach is not valid and has come under extreme review with all kinds of empirical support that the internal validity of the instrument is poor at best and deeply misleading and misused at worst. They suffer from significant gender discrimination.
- If it is not mandatory that all students complete ievail then they are of no use. It does not give an accurate portrait.
- Student evaluations should never be used alone to evaluate a professor. I would highly recommend that an unbiased committee of peers also be involved in evaluations. And why must the University evaluate a professor every single quarter. It leads to grade inflation as studies have shown us for a very long time. Everyone knows who the good professors are and to evaluate each quarter is folly. If there is a problem evaluations are necessary but otherwise why? We all know professors who get great evaluations and also noticed they give mostly A's and few C's and little else. This is not always the case of course but it is a dirty little secret that to get better evaluations the trick is to lower your standards. We all know that and the University knows that as well as this topic has been debated for many years. If you really wish to know about the teaching abilities of a professor look at their syllabus and the quality of their exams. Come to their classes and listen to their lectures. This would benefit both the professors and especially the students. But I do not expect any changes as I said above that it has long been known and researched again and again that teaching evaluations by students alone have unintended consequences. My question is: why are you bothering with a survey? Let me end by saying that some of my best teachers during my school years were not appreciated by me until years later. They made me work hard, they challenged my convictions and the like. And I feel so sad that I didn't appreciate them in my youth. In fact, they are responsible in large part for my success as they forced (a rather lazy student) to work hard and earn my grades and to love the world of ideas.
- Wrong approach! The problem is the inherent biases of ieval, not the content of individual questions.

SUBCOMMITTEE B (best practices to evaluate teaching)

Charge

Provide recommendations to the campus on how to interpret and use teaching evaluations responsibly for merit/promotions. Address the following questions:

- *What biases (or other factors) should department chairs or ad hoc merit/promotions committees keep in mind when using teaching evaluations?*
- *How should the data be presented so they are more intuitive to interpret, without encouraging problematic uses or abuse?*
- *How to incorporate for class size, level (lower/upper division, grad/undergrad, required/elective), type (lecture, on-line, hybrid, flipped), etc. in the interpretation of teaching evaluations?*
- *How to implement the instructions (APM 210-1-d) to Review Committees that Advise on Actions Concerning Appointees in the Professor and Corresponding Series when considering the teaching record of the candidate's performance.*
- *How to incorporate more comprehensive models of teaching evaluations (e.g., teaching portfolio)?*

Item 1: *What biases (or other factors) should department chairs or ad hoc merit/promotions committees keep in mind when using teaching evaluations?*

Overall, transparency and fairness should guide all forms of personnel evaluation. Furthermore, evaluators should reflect upon their personal assumptions and biases in their evaluative role. In light of this, we have three primary areas for recommendations:

Academic Personnel Office (VPAP & APO) and Committee on Academic Personnel (CAP) Direction

At the central administrative and Academic Senate level, evaluators need to communicate their role in academic personnel evaluation to Colleges and Departments, so that all academic personnel who undergo review are clear about these roles. Part of this communication, consistent with transparency and fairness, should indicate (a) the separate and distinct roles of the VPAP and CAP; (b) the goals and possible biases of these entities (e.g. 1, that the VPAP reports to the Provost and Chancellor and has a legal responsibility to the university, and thus has some common level of interest with the university's administration; e.g. 2, that CAP members ascribe to some shared sense of academic merit and productivity that may retain some measure of disciplinary biases); and (c) the emphases each party places upon a faculty member's overall performance (e.g., research productivity, resource acquisition, teaching and mentorship, and the quantity and quality of university and professional service).

Policy Guidance for Department Chairs, Associate/Divisional Deans, and Deans

At this level, there is a need for guidance on evaluation practices that constitute both intentional and unintentional bias. This guidance could include definitions of bias, forms of biases, and familiar biases evidenced across teaching evaluations (including student evaluations of teaching). For example, forms of bias in evaluation include weighted (in)appropriate disciplinary expertise, identity category bias/social stereotypes, teaching course assignment biases, and quantitative/qualitative research biases.

Interpretative directions/guidance for individual Faculty members and Department Chairs in deliberation meetings

For department chairs, there is a need for guidance with respect to the evaluation of teaching, particularly with reference to the Call and APM, so that matters outside of evaluation policy do not enter into deliberations or review decisions. For example, there is great inconsistency across departments about whether it is permissible to discuss family leave in personnel deliberations; this should be clarified for both candidates and department faculty members long before such meetings. Furthermore, there is a need for department chairs, and those chairing departmental review committees, to inform departmental members about what constitutes conscious and implicit bias in teaching and mentoring. Department chairs and departmental members need to become cognizant of relevant sections of the APM that can be loaded with bias: for example, "through *objective…* appraisal" (APM 210-1a) and "ability to organize material and to present it with *force* and logic" (APM 210-d1).

Item 2: *How should the data be presented so they are more intuitive to interpret, without encouraging problematic uses or abuse?*

Currently, UCR's numeric SET reports consist of 19 items, each containing 17 columns of data: the distribution of student responses, followed by the mean, median, and standard deviation at a course, department, and campus level, as well as department and campus-wide percentile rankings. In practice, many instructors and evaluation committees ignore most of those numbers and simply review the columns that are easiest to take in at a glance: the mean and the percentiles. Unfortunately, these are not good representations of student responses. At the same time, the volume of data on UCR's SET reports, and the increased precision to which the data are being calculated, invite a degree of comparative analysis that may not be warranted.

Some changes to how the data are presented in SET reports—and what data are presented—can guide reviewers to the most useful information, reign in the temptation to overanalyze, and avoid what Philip B. Stark calls "quantifauxcation."

Item Means

Prior to Fall 2016, the item means on SET reports were rounded to one decimal place. Since then, the means have been rounded to two decimal places. However, student responses do not support this level of precision. A mean of 4.3 is little different from a 4.4, and variations in the mean up to ±0.3 or ±0.4 might not be significant (Cashin rec 23, Linse 6.3). If item-by-item means are to be presented in SET reports, they should be rounded no further than one decimal place (Cashin rec 22).

Percentiles

Few campuses include percentile rankings in SET reports. For instance, among eight UC campuses surveyed so far, only UCR includes percentile rankings, and UCR has had difficulty making them work well.

Prior to Fall 2016, percentile ranks were assigned based on an item's mean rounded to one decimal place. Since then, they have been assigned based on an unrounded mean. This was done to reduce "clumps" in the percentile rankings and increased the number of ranks from roughly 25 to roughly 500. This is problematic, given the potential variation in means discussed above, but there are other considerations that further discourage the use of percentiles:

- Because student responses tend to be highly positive—typically in the 4-5 range—the percentile ranks re-cast a narrow range of mean values onto a 100 point scale. This tends to magnify insignificant differences. For instance, a 0.1 shift in an item's mean may translate to a jump of 10 or more percentile points. While the former is typically perceived as inconsequential, the latter can have the appearance of a significant "real" change in overall student response.
- The percentiles are frequently misinterpreted. As the 7 Sep 2016 memo "Changes to iEval Effective Fall 2016" notes, an item can have largely positive ratings (e.g. 85% positive) and still have a low percentile rank (35th percentile).

Yet in practice, when the percentile rank is low, reviewers have frequently interpreted the positive ratings as a sign of "poor performance."

- Given the potential variation in the mean value of student responses, confusion about what SETs measure, and concerns about the degree to which gender, ethnicity, age, course subject, class size, teaching innovation, etc. can affect student responses, it is counterproductive to rank SET results in a department or across campus.

In short, the percentile rankings, whether intended or not, invite reviewers to do a kind of analysis that they should not be doing via SETs (Linse 6.6). The percentile rankings should be discontinued.

Department and Campus Statistics

SET reports at a number of universities, such as UCI, UCSC, and UCSD, show responses only for the course itself. Those at other universities, such as UCB and UCSB, add departmental statistics, usually in the form of item means. Fewer, such as at UCSB, add campus-wide statistics. This lack of consensus about what to report is reflected in the literature. Some, such as Cashin, argue that one cannot meaningfully interpret the responses without comparison. Others, such as Freishtat, Linse, and Stark argue that SETs are so subjective and confound so many variables that it makes little sense to compare them.

The disagreement may arise from the fact that SET reports have two audiences: instructors and evaluation committees. An instructor might wish to see the department and campus means as a rough way to contextualize responses for a course, but an evaluation committee might not, given the low correlation between ratings and teaching effectiveness. In the committee this was summed up in one question: How relevant are these numbers? This merits further discussion.

Open-Ended Comments

There have been concerns about open-ended comments, particularly with regard to their variability, relevance, and appropriateness. While they can often yield interesting information, the general consensus is to look for trends in responses—with the usual caveat that open-ended responses should also be taken with a grain of salt. One question in the committee was whether it would be worthwhile to run the comments through sentiment analysis software, as one might do in marketing research. The thought was that this might counterbalance the tendency of reviewers to latch onto individual comments, but there was also concern about how well this would work in practice. The collection of individual responses should be made available in any case.

Visualization

A Likert survey, such as UCR's SET, is a collection of items that as a group provide information about an underlying phenomenon of interest; items should not be interpreted independently. It may be helpful, then, if SET reports explicitly grouped related items to encourage a holistic interpretation of student responses.

SET reports should also be presented in a way that foregrounds the distribution of responses (Freishtat J.1 & J.2). A table is the most frequent way to do this. Another option would be to graph the responses. A grouped diverging bar chart can fit all of the items onto a single page, which may encourage reviewers to interpret responses holistically. It offers a quick, intuitive way to see the response distribution. And it visually emphasizes the overall positivity or negativity of responses. That may be all that one would reasonably wish to glean from numeric SETs. While it may be tempting to analyze the SETs in greater detail and complexity than this, one should instead look to other evidence of teaching effectiveness as recommended in APM 210-1-d.

A sample table and chart based on UCR's current SET are included below. Each gives a readable summary that foregrounds the distribution of responses without inviting over-analysis.

Summary of Recommendations

1. Round item means to one decimal place, if they are to be included.
2. Remove the department and campus percentiles from SET reports.
3. Consider whether or not the remaining department and campus statistics are relevant in SETs.
4. Consider using a grouped diverging bar chart as the primary method of presenting responses in SET reports, or alternatively, a table that foregrounds the distribution of responses.

Instructor Name
Faculty Evaluation Responses: 2019S, ANON 0000-0
(22 enrolled, 17 responses)

Student Profile

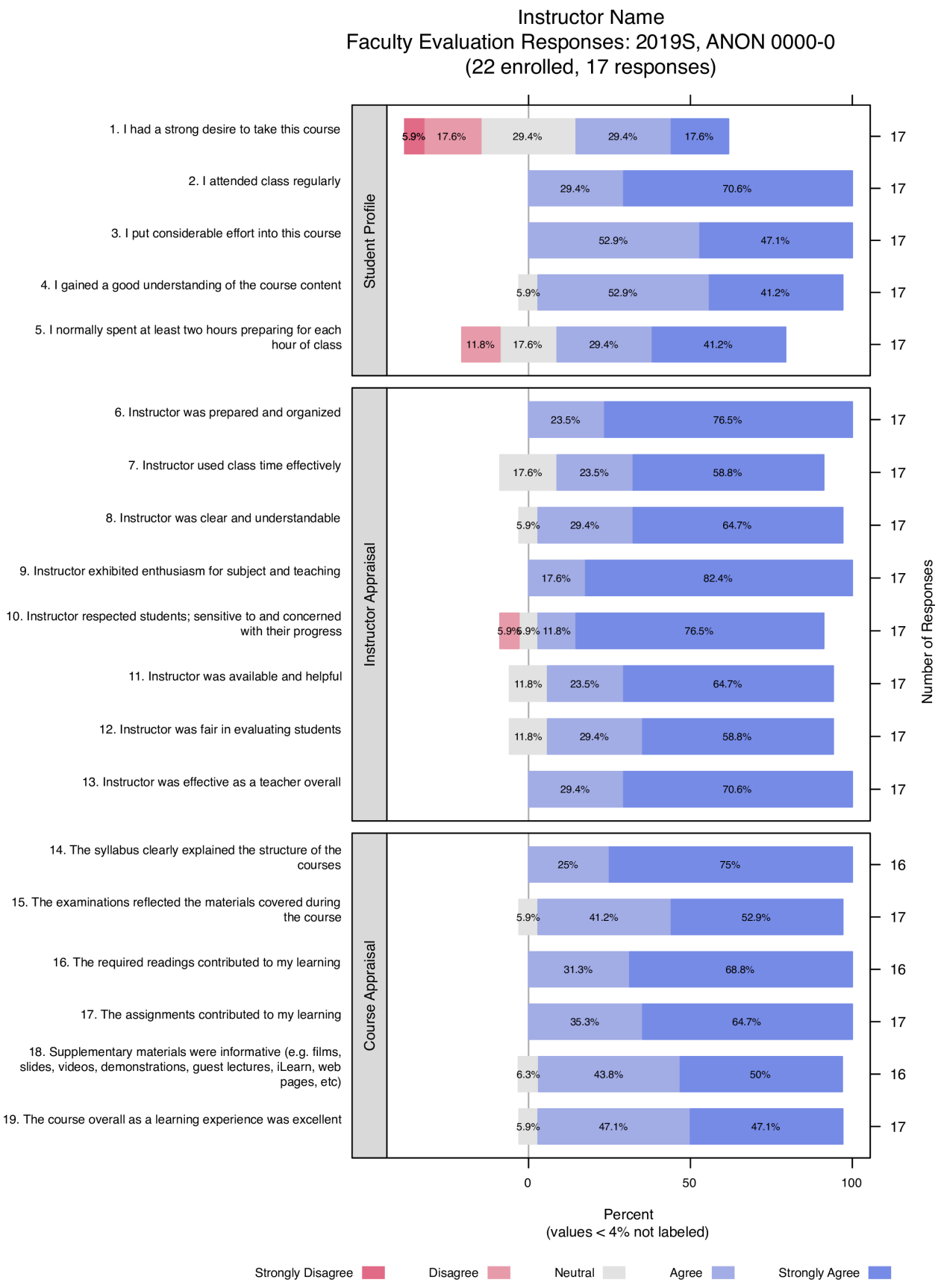
	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)	**Total**	Mean	Median
1. I had a strong desire to take this course	1 (5.9%)	3 (17.6%)	5 (29.4%)	5 (29.4%)	3 (17.6%)	17	3.4	3.0
2. I attended class regularly	0 (0%)	0 (0%)	0 (0%)	5 (29.4%)	12 (70.6%)	17	4.7	5.0
3. I put considerable effort into this course	0 (0%)	0 (0%)	0 (0%)	9 (52.9%)	8 (47.1%)	17	4.5	4.0
4. I gained a good understanding of the course content	0 (0%)	0 (0%)	1 (5.9%)	9 (52.9%)	7 (41.2%)	17	4.4	4.0
5. I normally spent at least two hours preparing for each hour of class	0 (0%)	2 (11.8%)	3 (17.6%)	5 (29.4%)	7 (41.2%)	17	4.0	4.0

Instructor Appraisal

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)	**Total**	Mean	Median
6. Instructor was prepared and organized	0 (0%)	0 (0%)	0 (0%)	4 (23.5%)	13 (76.5%)	17	4.8	5.0
7. Instructor used class time effectively	0 (0%)	0 (0%)	3 (17.6%)	4 (23.5%)	10 (58.8%)	17	4.4	5.0
8. Instructor was clear and understandable	0 (0%)	0 (0%)	1 (5.9%)	5 (29.4%)	11 (64.7%)	17	4.6	5.0
9. Instructor exhibited enthusiasm for subject and teaching	0 (0%)	0 (0%)	0 (0%)	3 (17.6%)	14 (82.4%)	17	4.8	5.0
10. Instructor respected students; sensitive to and concerned with their progress	0 (0%)	1 (5.9%)	1 (5.9%)	2 (11.8%)	13 (76.5%)	17	4.6	5.0
11. Instructor was available and helpful	0 (0%)	0 (0%)	2 (11.8%)	4 (23.5%)	11 (64.7%)	17	4.5	5.0
12. Instructor was fair in evaluating students	0 (0%)	0 (0%)	2 (11.8%)	5 (29.4%)	10 (58.8%)	17	4.5	5.0
13. Instructor was effective as a teacher overall	0 (0%)	0 (0%)	0 (0%)	5 (29.4%)	12 (70.6%)	17	4.7	5.0

Course Appraisal

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)	**Total**	Mean	Median
14. The syllabus clearly explained the structure of the courses	0 (0%)	0 (0%)	0 (0%)	4 (25%)	12 (75%)	16	4.8	5.0
15. The examinations reflected the materials covered during the course	0 (0%)	0 (0%)	1 (5.9%)	7 (41.2%)	9 (52.9%)	17	4.5	5.0
16. The required readings contributed to my learning	0 (0%)	0 (0%)	0 (0%)	5 (31.3%)	11 (68.8%)	16	4.7	5.0
17. The assignments contributed to my learning	0 (0%)	0 (0%)	0 (0%)	6 (35.3%)	11 (64.7%)	17	4.6	5.0
18. Supplementary materials were informative (e.g. films, slides, videos, demonstrations, guest lectures, iLearn, web pages, etc)	0 (0%)	0 (0%)	1 (6.3%)	7 (43.8%)	8 (50%)	16	4.4	4.5
19. The course overall as a learning experience was excellent	0 (0%)	0 (0%)	1 (5.9%)	8 (47.1%)	8 (47.1%)	17	4.4	4.0



Subcommittee B report for Ad Hoc Committee on Evaluation of Teaching

Item 3: *How to incorporate for class size, level (lower/upper division, grad/undergrad, required/elective), type (lecture, on-line, hybrid, flipped), etc. in the interpretation of teaching evaluations?*

The question of how to evaluate different categories of courses has been addressed at a number of different institutions. In our initial analysis, we have concluded that perhaps the best solution may be to differentiate mostly on class type, referenced as standard (i.e. lecture-based) and non-standard (utilizing technology and not requiring face-to-face interactions) formats. This has resulted in two recommendations for these two formats with the first recommendation giving faculty members some flexibility in the questions that students use based on their specific standard course. The second recommendation builds on the first recommendation but replaces student self-evaluation questions with a series of questions that address the student's experience in the virtual realm.

Recommendation 1: For standard (i.e., face-to-face) classes, a set of questions should be common across all courses within categories that allow for equitable evaluation regardless of class size or level. Within each category, instructors will be given the choice of choosing a default question or choosing alternate questions that are tailored for their particular course. All questions will be pre-approved by a faculty committee to ensure consistency and accuracy between instructors and courses. The course evaluation question bank will be administered and maintained by the UCR Center for Teaching and Learning.

Background: UCR has employed one student evaluation form for over a decade and the current form (iEval) does not differentiate between class size or level. Allowing more flexibility in the evaluation instrument provides instructors more meaningful feedback to improve their teaching while allowing evaluators (peers, departments, colleges, etc.) an opportunity to assess the overall quality of instruction in different courses.

UC Berkeley employs a course evaluation question bank that gives instructors flexibility in designing their own evaluations that can be tailored to specific courses and formats (https://teaching.berkeley.edu/course-evaluations-question-bank). The question bank is updated by faculty members as course formats evolve and are maintained by the Berkeley Center for Teaching & Learning. Although instructors can choose from a bank of questions, the evaluation will always have four different categories that must be addressed through the evaluation process:

1. Instructor-specific question themes
2. Course-specific question themes
3. Student self-evaluation questions
4. Open-ended questions

Each category has a number of default questions that were based on faculty feedback, but faculty are free to use other questions that are available in a course evaluation question bank. The default questions for the first two categories (required) are:

CATEGORY 1: Instructor-specific question themes

- Presentation of Content: The instructor presented content in an organized manner
- Clarity of Expectations or Directions: The instructor explained concepts clearly
- Helpfulness/Availability: The instructor was helpful when I had difficulties or questions
- Useful/clear feedback on performance: The instructor provided clear constructive feedback
- Encouraging of participation/discussion: The instructor encouraged student questions and participation
- Overall Teaching Effectiveness (MANDATORY): Considering both the limitations and possibilities of the subject matter and the course, how would you rate the overall effective of this instructor?

CATEGORY 2: Course-specific question themes

- Course Content: The course was effectively organized
- Application & Skill Development: The course developed my abilities and skills for the subject
- Theory/Content Knowledge: The course developed my ability to think critically about the subject
- Course Overall: Considering both the limitations and possibilities of the subject matter and course, how would you rate the overall effectiveness of this course?

CATEGORY 3: Student Self-Evaluation Questions

1. How many class sessions did you attend?
2. On average, how many hours per week have you spent on this course (or section), including attending classes, doing readings, reviewing notes, writing papers, and any other course-related work?
3. How satisfied were you with your effort in this course?

CATEGORY 4: Open-Ended Questions

1. Please identify what you consider to be the strengths of the course.
2. Please identify area where you think the course (or section) could be improved.
3. Feedback for other students: What advice would you give to another student who is considering taking this course?

<u>Recommendation 2:</u> For non-standard courses (online, hybrid, flipped) that require a virtual interface, the same form that is used for standard courses can be used for categories 1, 2, and 4. Questions in category 3 should be replaced to reflect those aspects of the course that directly impact student learning and the use of technology and degree of engagement in the absence of personal contact.

Background: Many institutions of higher education are still struggling with how to effectively evaluate online formats for courses. Some professional organizations that offer online courses have tackled this issue and, in addition to questions that address overall instructor and course effectiveness, ask students a number of questions that are unique to online and hybrid courses. The American Historical Association is one of those organizations and has students answer the following questions in a course on Native American History *(*https://www.historians.org/teaching-and-learning/teaching-resources-for-historians/teaching-and-learning-in-the-digital-age/the-history-of-the-americas/online-course-in-american-indian-history/student-evaluation-of-online-courses):

1. Course-related resources were available for me
2. I was successful in this course as I am in similar courses taught in the classroom
3. I participated in this course as much as I would for similar courses taught in the classroom
4. I felt that I was a part of a learning community
5. I found it easy to remain motivated to complete the course requirements
6. The technology interfered with my learning
7. After taking this class, I am more comfortable using technology
8. Technical help was available when I needed it.
9. I would take another online course.
10. Prior to the start of this class I had completed:
    - No web-based classes
    - 1-2 web-based classes
    - 3-5 web-based classes
    - *More than 5 web-based classes*

Items 4 & 5: *Subcommittee comments on a comprehensive model of teaching effectiveness, and how to implement the instructions (APM 210-1-d) to Review Committees when considering the teaching record of the candidate's performance.*

Notes:

- Teaching effectiveness does not refer only to in-class teaching but includes all other aspects of educating students such as, but not limited to, general guidance, mentoring, supervising, and advising at all levels.
- The document pertains only to evaluation of teaching effectiveness by reviewing committees, not as a tool for continuous teaching improvement.

The sections of APM 210-1-d and the CALL that are relevant to these issues are included in Appendix 1.

APM 210-1-d spells out what evaluation committees should consider when judging teaching effectiveness (comprehensive model: e.g., command of the subject; ability to organize material and to present it with force and logic; capacity to awaken in students an awareness of the relationship of the subject to other fields of knowledge; fostering of student independence and capability to reason, etc.). Further, APM 210-1-d lists the most common types of evidence of teaching effectiveness that can be used by the evaluation committees (e.g., opinions of other faculty members; through student evaluations (SET); development of new and effective techniques of instruction, etc.). APM 210-1-d emphasizes that more than one kind of evidence should accompany each review file and that the committee should judge the total performance of the candidate with proper reference to assigned teaching responsibilities.

Our observation is that at UCR, for the most part, evaluation committees rely disproportionally (and often exclusively) on SET in evaluating teaching effectiveness, contradicting the requirement and the spirit of APM 210-1-d. An informal survey of the practices at other UC campuses reveals similar state (see Appendix 2). This is a problematic practice because recent research suggests that SET may not accurately reflect the quality of teaching due to biases and other factors (see for example, here, here, here and here), thus the need not to overemphasize SETs and certainly not to use them as the sole source for teaching evaluation.

Our recommendations below are based on secondary information gathering of practices at other universities (Appendix 3) and our own understanding and insights.

Key recommendation

In accordance with APM 210-1-d, the University should require that more than one kind of evidence of teaching effectiveness should accompany each review file. To implement this recommendation the following steps should be taken:

Education

The AP 210-1-d requirement should be widely disseminated to all faculty. It should be emphasized that it is the responsibility of the candidate to provide more than one kind of evidence of teaching effectiveness. The requirement should be emphasized in all orientation sessions to new faculty, department chairs, and CAP members.

The Academic Personnel Office's current Department Chair Checklist, Career Development section, item 1 states that department chairs should "Have annual reminder conversations with all faculty about The CALL and APM-210." As part of this conversation, we recommend that each department discuss what is the most appropriate evidence of teaching effectiveness (other than SET) in their own discipline/department and encourage faculty to provide such evidence when preparing their own evaluation file. A "one size fits all" notion of good teaching is not possible and, as such, disciplinary variations of the dimensions of effective teaching and the type of corroborating evidence are expected.

In support of this effort, the Center for Teaching & Learning should prepare a document (available online) that outlines common types of evidence of teaching effectiveness that are not SET. See, for example, the following sites:
UCLA
The University of Kansas
University of Washington
University of Michigan and here

The Center for Teaching & Learning in preparing the document on the common types of evidence of teaching effectiveness that are not SET should be informed by the discussion in the various departments.

We also recommend that the "Teaching Statements" in the eFile (the Department's Teaching Statement) should include a statement listing the most appropriate evidence of teaching effectiveness (other than SET) in their own discipline.

Taken together, action on these recommendations can provide faculty with much needed clarity about how teaching effectiveness is evaluated and the range of evidence that can be brought to bear on it.

Facilitation
To facilitate the process and to encourage candidates to provide multiple kinds of evidence of teaching effectiveness, the eFile collection system should be expanded to include additional explicit categories (under the Teaching Information) such as "peer-review", "lecture notes", "evaluation/assessment instruments", "measures of student achievement," etc. These categories should include an option to upload relevant documents. Further, the input category for "Syllabus" that is currently under "Other Information" should be moved to under the "Teaching Information" category.

For a specific recommendation on revising the structure of the eFile see Appendix 4.

"Enforcement"
Possibly, after transition period, after education and input facilitation would make providing more than one kind of evidence of teaching effectiveness routine, the eFile will not be considered complete unless the candidate provides more than one kind of evidence. At this point, CAP should return any departmental letter that does not discuss more than one kind of evidence of teaching effectiveness. We recommend that the "enforcement" stage would not commerce without further endorsement by the Senate.

Members of the Subcommittee B Task Force

John Stamp, Subcommittee B Chair, University Writing Program
Jean Helwege, Professor, Area of Finance
John Levin, Distinguished Professor, Graduate School of Education
Rami Zwick, Professor, Area of Marketing
Jennifer Brown (ex-officio), Vice Provost for Undergraduate Education
Rich Cardullo (ex-officio), Professor, Evolution, Ecology, and Organismal Biology
Daniel Jeske (ex-officio), Vice Provost of Academic Personnel
Mariam Lam (ex-officio), Vice Chancellor of Diversity, Equity, and Inclusion
Ameae Walker (ex-officio), Vice Provost of Academic Personnel

APPENDIX 1 – APM 210-1 and the relevant sections from the CALL

APM 210-1

210-1. Instructions to Review Committees That Advise on Actions Concerning Appointees in the Professor and Corresponding Series

…….

210-1d. Criteria for Appointment, Promotion, and Appraisal

210-1d1. Teaching - Clearly demonstrated evidence of high quality in teaching is an essential criterion for appointment, advancement, or promotion. Under no circumstances will a tenure commitment be made unless there is clear documentation of ability and diligence in the teaching role.

In judging the effectiveness of a candidate's teaching, the committee should consider such points as the following (*bullet points are not in the original*)

- the candidate's command of the subject
- continuous growth in the subject field
- ability to organize material and to present it with force and logic
- capacity to awaken in students an awareness of the relationship of the subject to other fields of knowledge
- fostering of student independence and capability to reason
- spirit and enthusiasm which vitalize the candidate's learning and teaching
- ability to arouse curiosity in beginning students, to encourage high standards, and to stimulate advanced students to creative work
- personal attributes as they affect teaching and students
- extent and skill of the candidate's participation in the general guidance, mentoring, and advising of students
- effectiveness in creating an academic environment that is open and encouraging to all students, including development of particularly effective strategies for the educational advancement of students in various underrepresented groups

The committee should pay due attention to the variety of demands placed on instructors by the types of teaching called for in various disciplines and at various levels, and should judge the total performance of the candidate with proper reference to assigned teaching responsibilities.

The committee should clearly indicate the sources of evidence on which its appraisal of teaching competence has been based. In those exceptional cases when no such evidence is available, the candidate's potentialities as a teacher may be indicated in closely analogous activities.

In preparing its recommendation, the review committee should keep in mind that a redacted copy of its report may be an important means of informing the candidate of the evaluation of his or her teaching and of the basis for that evaluation.

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.

More than one kind of evidence shall accompany each review file.

Among significant types of evidence of teaching effectiveness are the following:

(a) opinions of other faculty members knowledgeable in the candidate's field, particularly if based on class visitations, on attendance at public lectures or lectures before professional societies given by the candidate, or on the performance of students in courses taught by the candidate that are prerequisite to those of the informant;
(b) opinions of students;
(c) opinions of graduates who have achieved notable professional success since leaving the University;
(d) number and caliber of students guided in research by the candidate and of those attracted to the campus by the candidate's repute as a teacher; and
(e) development of new and effective techniques of instruction, including techniques that meet the needs of students from groups that are underrepresented in the field of instruction.

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;
(b) a quarter-by-quarter or semester-by-semester enumeration of the number and types of courses and tutorials taught since the candidate's last review;
(c) their level;
(d) their enrollments;
(e) the percentage of students represented by student course evaluations for each course;
(f) brief explanations for abnormal course loads;
(g) identification of any new courses taught or of old courses when there was substantial reorganization of approach or content;
(h) notice of any awards or formal mentions for distinguished teaching;
(i) when the faculty member under review wishes, a self- evaluation of his or her teaching;
(j) evaluation by other faculty members of teaching effectiveness.

When any of the information specified in this paragraph is not provided, the department chair will include an explanation for that omission in the candidate's dossier.

If such information is not included with the letter of recommendation and its absence is not adequately accounted for, it is the review committee chair's responsibility to request it through the Chancellor.

The CALL

……..

J. Departmental Recommendation Letter

J1b Evaluation of Teaching

Where possible and applicable, the departmental letter should comment on items such as the following:

i. The role of the candidate in the graduate and undergraduate instructional program including such items as the amount relative to the department norms variety and difficulty of the teaching assignments and the preparation and attention given by the candidate to his/her teaching responsibilities. Make reference to teaching information form.

ii. Out-of-class teaching and advising at both the graduate and undergraduate levels (careful thought should be given to the advising role of each candidate): directed research, special studies, help given to students, office hours with students, contributions to the teaching of other faculty, etc.

iii. Graduate student supervision and advising: PhDs, Masters, committees, post-doctoral, and Graduate Research Assistant supervision. Attention may be given to the role of the candidate in attracting high caliber graduate students to the campus.

iv. Development of new and effective techniques of instruction; writing of teaching materials, manuals, textbooks.

v. Evaluation of teaching as judged by departmental colleagues. Guidelines dealing with the evaluation of teaching are contained in APM 210-1 and should be consulted by Chairs on behalf of their departments. Among other elements of teaching, faculty colleagues are particularly well qualified to make thoughtful and substantial assessments of the candidate's command of subject matter and continuous growth in his/her field. Faculty perceptions derived from direct observation and information should be shared with colleagues at the departmental personnel meeting concerning the candidate and incorporated into the file in an appropriate manner.

vi. Evaluation of teaching by students. Materials submitted by students should be discussed by the department in its meeting and summarized and evaluated in the departmental letter. All teaching evaluations performed during the review period should be assessed and commented on. Hearsay is not acceptable for use in teaching evaluations.

APPENDIX 2 - UC practices

April 18, 2019

Vice Provost for Academic Personnel (or Faculty) at the UC campuses were asked to comment on the following questions:

1. Implementing APM 210-d-1 that stipulate that there be more than one form of evaluation of teaching
2. Do you incentivize SET?
3. Do you use teaching portfolios to evaluate teaching?

The following is a summary of the replies:

UCSD
Our Center for Innovations in Teaching and Learning (CITL) has recently complied a webpage of resources on documenting teaching effectiveness, which includes a guide for evidence for personnel reviews.

1) Everyone uses course evaluations, and almost everyone discusses graduate mentoring, and UCSD count those as two sources of evidence. A few departments have a culture of including syllabi, and a few get letters from students.

2) UCSD is struggling with incentivization, but are on hold worrying about it, as UCSD is likely to transition to a new online system very soon.

3) Seldom use teaching portfolio in a personnel file, but CITL has started to talk about them.

UC Davis
UC Davis requires peer review of teaching for the two promotion steps; at any other step it is optional and not often done. The peer review typically involves one or more letters familiar with the candidate's subject matter material and involves attending at least one lecture and reviewing course materials. Dossiers without these will be sent back to the departments.

Student evaluation is not incentivized except in the School of Veterinary Medicine, which has local control over its entire curriculum. Students cannot register for the next semester's courses until they complete their course and instructor evaluations from the previous semester. UC Davis recognizes the limitations of SET both in undergraduate and graduate courses.

UC Davis Dean for Undergraduate Education would like to try an experiment with having a small team of trained faculty within a college do more exhaustive teaching evaluations for all faculty. So far, no college is willing to participate in such an experiment.

Policy. The Step Plus system for merit and promotion enables rewards for instructors who make outstanding contributions to excellence in teaching. Also, peer evaluation of teaching is necessary for promotion.
http://academicaffairs.ucdavis.edu/policies/step-plus/guidelines-for-advancement-senate.html

UCSF

#1 – there is variability by department and type of teaching (e.g., large lecture in prof. schools; small graduate seminars; individual lab-based instruction). Most dossiers contain structured evaluations of teaching from on-line evaluations (primarily professional school instruction) and individual letter of support for teaching contributions. UCSF does not send back files that contain only one type of teaching evaluation.

#2 – some professional schools do not release student grades until a certain percentage response rate (to on-line evaluations) is reached.

#3 – a minority of our faculty use teaching portfolios (primarily those associated with our campus' Academy of Medical Educators; but others as well).

UCSB

The UCSB Senate also has a task force investigating the statistical course evaluations.

1) in regard to APM 210-d-1, Departments generally provide both the statistical evaluation and the written student comments (there is a trial in some Departments for online evaluations, but the return response is low)

2) UCSB does not use any incentives

3) Faculty often provide a self-assessment of their teaching performance.

UCSB would like to have better statistical analysis of SET (and especially to educate the students on what these evaluations really are).

UCLA

In regard to #1, UCLA typically use student evals (numbers and comments, typically comments are much more informative), but also letters and/or forms from peers and departments. Procedures however are not standardized and different departments and schools use different approaches.

For #2 UCLA experience the lower rates with on-line evals. Using class time helps increase percentages of students evaluating. UCLA new Center for Advancement of Teaching [CAT, used to be Office of Instructional Development (OID)] is experimenting

with new evaluation programs and procedures. Also some of UCLA departments are adding procedures to evaluate special programs for faculty to participate in to help improve their teaching which will substitute for peer evaluation. This is still in an experimental and trial phase.

Other Forms of Evaluating and Reflecting on Teaching
Center for Education Innovation and Learning in the Sciences
https://ceils.ucla.edu/resources/teaching-guides/other-forms/

Given the many concerns with focusing only on student ratings of instruction, what are other ways that you and your institution can assess teaching effectiveness and student learning?
The Association of American Universities (AAU) created a matrix of various campus strategies in the evaluation of faculty teaching (e.g., promotion and tenure). The following is from the AAU matrix at https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/P%26T-Matrix.pdf

UC Merced
For 1., course evaluations ratings are still the 'meat and potatoes' of assessment.

For 2., no incentives right now.

For 3., at review time, faculty will occasionally submit a set of auxiliary materials, such syllabi, lesson plans, new material they have developed, etc. Often these are individuals about to undergo tenure review, and need to show how they have addressed areas of weakness identified in the MCA.

In addition,
UC MERCED try to make teaching stand out as important during onboarding. Encourage mentors to discuss effective teaching practices with pre-tenured faculty. The Center for Excellence in Teaching (CETL) gives presentations to all new faculty. CETL can also help informally assess and address weaknesses, when a faculty member seeks this help. Some departments encourage weaker instructors to go CETL, and many do not. CETL was approached (last year?) to help develop a more robust formal assessment approach, but CETL is not there yet.

UC Berkeley
CAP or the VP for faculty will send back threshold cases that limit the assessment of teaching to just reporting the means of student evaluations.

UC Berkeley still use a mix of online and paper forms. Online response rates are poor; although they can be boosted somewhat if the instructor asks students to bring their laptops to the last class and do the online evaluations then. UC Berkeley encourages portfolios.

The ninth meeting of the Berkeley Division of the Academic Senate (DIVCO) was held on Monday, February 11, 2019. One item on the agenda was a discussion of new proposed policies for the evaluation of teaching
https://academic-senate.berkeley.edu/sites/default/files/the_fortnightly_february_12_2019.pdf

DIVCO discussed proposed revisions to the policy on teaching evaluation. These revisions were promoted in part by Philip Stark and Richard Freishtat's research on the merits of the existing evaluations in assessing learning outcomes and teaching effectiveness. The Committee on Teaching had proposed a teaching dossier several years ago, and a new set of questions were proposed by the Committee on Budget and Interdepartmental Relations (Budget Committee) last year, and circulated to other Senate committees for feedback and commentary. Following this process, Vice Provost for the Faculty Ben Hermalin formulated an official revised policy. While there was consensus that the revised policy is an improvement on the current policy (which dates to 2006), DIVCO members raised concerns about increased workload both for staff compiling and entering data into APBears and for department chairs conducting peer reviews of teaching.

The VP document is at "Teaching_Evaluation_Memorandum_v2.pdf"

UCI
UCI requires a second piece of evidence for teaching. This is enforced because they use an electronic review platform and the file does not move forward unless the second piece is uploaded.

After a pilot of accepting any form of second evidence (3 years ago) UCI provides the guidance as this URL.http://dtei.uci.edu/second-piece-of-evidence-of-teaching-effectiveness/

The majority of faculty submit a reflective teaching statement as their second piece of evidence and this has been quite useful not only for review but for spreading and highlighting innovative things lots of faculty are doing when discussed at each level of review. However, UCI do have some that submit teaching portfolios, peer evaluations, and a very small number that continue to just submit syllabi (which was common first year but these are generally not helpful and UCI has discouraged this practice in absence of accompanying reflective teaching statement).

Policy. Evaluation System for P/T. Making changes toward using a three-bucket system to evaluate accomplishments in research, teaching and service separately.

While published UC policy indicates that at least two types of evidence should support evaluation of teaching (see the Supplemental Material or visit www.ucop.edu/academic-personnel-programs/_files/apm/apm-210.pdf, p. 5), in practice, student evaluations are often the only evidence used. For the 2016 review cycle, UC Irvine has required

individuals to upload at least one additional type of evidence to evaluate teaching (e.g., reflective teaching self-statement, syllabus, peer evaluation, or measure of student achievement). This change is a first step toward conducting a more thorough evaluation of the contributions to teaching. It also broadens the discussion of teaching by everyone involved in the review process and thus has the potential to increase awareness of the innovative and effective teaching practices taking place on campus.
(from CBE—Life Sciences Education Vol. 16, No. 4 Essay, Aligning Practice to Policies: Changing the Culture to Recognize and Reward Teaching at Research Universities. Published Online :15 Mar 2018 https://doi.org/10.1187/cbe.17-02-0032

APPENDIX 3 – Survey of some non-UC practices

NSF Award Search: Award#1725959 - Collaborative Research: Transforming the Evaluation of Teaching: A Study of Institutional Change to Advance STEM Undergraduate Education
2017-2022
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1725959&HistoricalAwards=false
In this project, the University of Kansas, the University of Colorado at Boulder and the University of Massachusetts Amherst will approach change in ways best suited to their campus culture. Each, though, will implement an evaluation framework that is based on two decades of scholarship on scholarly teaching and its evaluation. The framework draws on multiple sources of evidence, including students, peers, instructors, and that speaks to multiple dimensions of teaching and learning. The work on each campus will center on the development and use of a teaching evaluation rubric that provides a richer, more complete view of teaching practice, and the evidence that speaks to it.

See also at https://www.aau.edu/sites/default/files/Blind-Links/Teaching-Eval-AAU-Chairs-Oct2018.pdf

The University of Kansas
KU developing new, more insightful method of evaluating teaching than traditional student forms
03/18/2019
http://today.ku.edu/2019/02/28/ku-developing-new-more-insightful-method-evaluating-teaching-traditional-student-forms

BENCHMARKS FOR TEACHING EFFECTIVENESS
https://cte.ku.edu/sites/cte.ku.edu/files/docs/Branding/Benchmarks/BenchmarkswRubricwCCUpdated.pdf
see at “KU BenchmarkswRubricwCCUpdated.pdf”

USC
Excellence in Teaching - USC Provost
March 27, 2018
https://www.provost.usc.edu/excellence-in-teaching/
Academic Senate Task Force on Teaching Evaluations
The Academic Senate convened a Faculty Evaluation Task Force in 2013. The resulting white paper was reviewed by the full Academic Senate in December, 2013. The final paper was sent to the Provost’s office in February, 2014. Teaching is covered on pages 3 and 4.

Teaching Evaluations Update | Academic Senate | USC
https://academicsenate.usc.edu/teaching-evaluations-update/

September 20, 2018

Recommendations of the Faculty Evaluation Task Force (2013) of the Academic Senate
https://docs.google.com/document/d/13YvYZpm_HWAuI8LYB5vJhBJdLuSvcZiZQPN6bTGlSyA/edit
APPENDIX V. EXCERPTS FROM 2013 UCAPT MANUAL (p. 14)

University of Washington | Seattle, WA
A guide to best practice for evaluating teaching | Center for Teaching and Learning
https://www.washington.edu/teaching/teaching-resources/assessing-and-improving-teaching/evaluation/

- Self-assessment
- Peer review
- Student evaluations

University of Michigan
Evaluation of Teaching | CRLT
http://www.crlt.umich.edu/resources/evaluation-teaching

Evaluation of teaching can have many purposes, including collecting feedback for teaching improvement, developing a portfolio for job applications, or gathering data as part of personnel decisions, such as reappointment or promotion and tenure. Most of the methods described below can be used for all of these functions. What follows are multiple methods for collecting information about instructors' activities, accomplishments, and effectiveness in teaching, in the classroom and beyond. While this list includes best practices for using student ratings, it also offers suggestions for ensuring that student ratings are not the only source of evidence used to assess instructional effectiveness, an approach consistent with research. In addition, detailed resources are available on the topics of **student ratings of instruction**, **peer review of teaching**, and **teaching portfolios and course portfolios**.

Guidelines for Evaluating Teaching | CRLT
http://www.crlt.umich.edu/tstrategies/guidelines

University of Oregon
Revising UO's Teaching Evaluations | Office of the Provost
https://provost.uoregon.edu/revising-uos-teaching-evaluations

Revising UO's Teaching Evaluations

The Office of the Provost and the University Senate are currently working together to critique and revise our entire teaching evaluation system. Recent research suggests that student ratings may not accurately reflect the quality of teaching due to biases and other factors [3, 4, 5, 6]. The University of Oregon's own assessment of student course evaluation ratings have corroborated these findings [2].

The Association of American Universities (AAU) and other universities around the globe from University of Colorado, Boulder to University College London, England have argued that it is time for universities' practices regarding teaching excellence and evaluation to align with their policies [1, 7, 8]. As such, the University of Oregon seeks to develop a holistic new teaching evaluation system that does more than simply replace problematic evaluation instruments so that we can help the UO community more effectively define, develop, evaluate, and reward teaching excellence. More specifically, we want to increase equity and transparency in the use and efficacy of teaching evaluations for merit, contract renewal, promotion and tenure, while simultaneously providing instructors with tools for continual course improvement.

Draft April 29 2018
Prepared by Senate Task Force for Teaching Evaluation
Proposed UO Peer Review of Teaching Framework
https://provost.uoregon.edu/files/proposed_peer_review_framework_aug2018.pdf

The University of Texas at Austin
Peer Observation of Teaching | Faculty Innovation Center
https://facultyinnovate.utexas.edu/peer-observation

University of Wisconsin System
Student Evaluation of Instruction | Board of Regents
https://www.wisconsin.edu/regents/policies/student-evaluation-of-instruction/

In assessing teaching effectiveness, the Regents believe that student evaluations are an important and useful source of evidence that should be explicitly considered in reaching judgments. The Regents assert, however, that student evaluation must not be a substitute for direct peer judgment of teaching effectiveness through a variety of means such as observation of teaching, assessment of syllabi, examinations and other course materials, and evaluation of contributions to development and strengthening of departmental curriculum. Moreover, effective peer judgment of teaching effectiveness necessarily includes both examination of the faculty member's current level of performance, and also his or her potential for growth.

Purdue University
Task force to recommend alternatives for evaluating teaching
November 29, 2017
https://www.purdue.edu/newsroom/purduetoday/releases/2017/Q4/task-force-to-recommend-alternatives-for-evaluating-teaching.html

Boston University
Report and Findings of the Task Force on Evaluating Teaching » Office of the Provost
February 7th, 2019
https://www.bu.edu/provost/2019/02/07/report-and-findings-of-the-task-force-on-evaluating-teaching/

Report of the Task Force on Evaluating Teaching
October 2018
https://www.bu.edu/provost/files/2019/02/Report-of-the-Task-Force-on-Evaluating-Teaching.pdf

Executive Summary
In May 2017, the Task Force on Evaluating Teaching was charged with gathering information about the language and content of student course evaluations at Boston University and considering how this content aligns with current research and national conversations about how best to evaluate university teaching. The Task Force was further charged with drafting a University-wide course feedback form with both common content and questions and flexible options for schools, colleges, programs and departments to tailor the surveys to their specific needs. Additionally, the Task Force was asked to make recommendations about a move to collect student course feedback online. Finally, the Task Force was invited to comment on the place of student course feedback in the overall process of evaluating teaching on campus, considering the role of other measures, such as peer evaluation of teaching, self-reflection, and review of portfolios of teaching materials, all with the goal of creating a set of standard guidelines and practices for evaluating teaching on the Charles River and Medical Campuses.

Key recommendations
3. The University should require multiple measures in order to responsibly evaluate teaching, including peer evaluation and teaching portfolios, for all personnel decisions, including contract renewal.

APPENDIX 4 – Proposed changes to the eFile structure

eFile Categories

CV Overview

- Publications
- Creative Activities
- Patents
- Professional Service
- University and Public Service
- **Teaching Information**
    - Teaching Information and Evidence of Teaching Effectiveness
        - Teaching Records and Student Evaluation of Teaching
        - Teaching Releases
        - Departmental Teaching Statement
        - Faculty's Self Statement of Teaching Objectives and Performance
        - Pedagogical Activities (*allow upload of documents*)
            - Enter "Description"
                - Peer-Review of Teaching Effectiveness
                - Curriculum Developments
                    - Program Development
                    - Course Development
                    - Course Revision
                - Development of New and Effective Techniques of Instruction
                - Representative Teaching Materials
                    - Syllabus
                    - Teaching Notes
                    - Examples for Assessment Used
                    - Videos of Lectures
                    - Other
                - Letters From Former Students Who Have Achieved Notable Professional Success (Non-Confidential)
                - Extension Teaching
                - Mentoring
                - Advising
                - Post Doc
                - Summer Internship
                - Senior Thesis Advisor
                - Mentor-No UCR Student
                - Dissertation Reading-Non UCR Student
                - Visiting Researcher Visiting Scholar
                - Thesis Director Non UCR Student
                - Non UC-Teaching
                - Other

The category of "Student Instruction and Sponsorship" should not be a new category but should be included under the general category of "Pedagogical Activities".

- **Student Instruction and Sponsorship**
  - Student Instruction
    - Committee (required)
      - PhD Dissertation
      - Qualifying Exam
      - Prospectus
      - Masters Thesis
      - Oral Exam
      - Advisory Committee
      - Analytical Project
    - Role
      - Major Professor
      - Co-Major Professor
      - Chair
      - Co-Chair
      - Member
      - Thesis Advisor
      - Other
  - Student Sponsorship
    - Type of Support (required)
      - Training And Grant
      - Student Grant
      - Non-UC Fellowship
      - UC Fellowship
      - Research Grant
      - Intramural Funds
      - In Kind Sponsorship
      - Other
- Fellowship, Grant, and Gift Activities
- Memberships
- Honors and Awards
- Employment History
- Education
- Self-Statements
- Other Information
  - Non-Confidential Document (Type of document - required)
    - Comments from Extramural Grant Reviewers Contract
    - Invitation Letters
    - News Articles
    - Public Acceptance Letter Reader's Reports
    - Thank-you Letter
    - Unsolicited Letter (Nan-Confidential)
    - Other
  - Letter from Other Departments/Programs, Institutes and Centers (Proposed action – required)

UC RIVERSIDE UNIVERSITY OF CALIFORNIA
Marlan and Rosemary Bourns College of Engineering

Marlan and Rosemary Bourns College of Engineering
446 Winston Chung Hall
900 University Avenue
Riverside, CA 92521

April 1, 2021

TO: Jason Stajich, Chair
Academic Senate

FROM: Philip Brisk, Chair
BCOE Executive Committee

RE: Ad Hoc Committee Report on Evaluation of Teaching

Dear Jason,

The BCOE Executive Committee reviewed the Final Report of the Ad Hoc Committee on Evaluation of Faculty Teaching. While the Executive Committee appreciates the efforts of the Ad-Hoc Committee, the Report fails to substantially address the fact that student evaluations of teaching (SETs) are irrevocably flawed, and their use as evidence of teaching in the merit and promotion process incentivizes grade inflation and fan service over assessment-driven improvements to pedagogy.

I will begin this letter by quoting directly from an article entitled "Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching" by Professor Troy Heffernan at La Trobe University, Melbourne, Australia, published on March 6, 2021, in the journal Assessment & Evaluation in Higher Education (DOI: 10.1080/02602938.2021.1888075). Citing prior work, which I will not reproduce here, the paper notes that (1) trends in abusive comments in SETs are increasing; and (2) prejudices in SET results and the way that they are used is leading to increasing issues relating to academic mental health and wellbeing. I will quote the conclusion directly, noting that the primary concern of this paper is not the ineffectiveness of SETs in general, but their impact in marginalized groups from a DEI perspective:

> This paper has shown that no university, and indeed the higher education sector as a whole, can declare to be a gender equal employer or have an interest in growing a safe, inclusive and diverse workforce if they continue using SETs to evaluate course and teacher quality.
>
> This paper provides an evidence base which can be used as part of the growing material and argument against the practice of collecting SET data. When SET data is known to be highly prejudiced against many groups, methods must be changed, and using SET data as a component of hiring, promotion and/or firing decisions must be seen as the blatantly discriminatory practice that it is.
>
> The need for immediate policy changes is clear. Women and marginalised groups are losing jobs, failing to achieve promotion, and are being negatively impacted at every step where SETs are concerned, and will continue to be discriminated against every time SET data is collected until the practice is stopped. These practices not only harm the sector's women and most underrepresented and vulnerable, it cannot be denied that SETs also actively contribute to further marginalising the groups universities declare to protect and value in their workforces.

The current practice of willfully including student perspectives in the evaluation of teaching using a flawed and biased measure does just as much disservice to the students as it does to the faculty. The BCOE Executive Committee values student feedback on teaching and strongly encourages the Senate to consider effective and unbiased mechanisms to incorporate student feedback into the Merit and Promotion process.

Having read the Ad Hoc Committee's Report in detail, the BCOE Executive Committee feels that the majority of its suggestion represent incremental improvements in comparison to the status quo; to conserve space, the remainder of this letter will set forward some specific BCOE concerns about the adverse of impact of how SETs are presently used today, provide specific feedback on some of the suggestions in the Report, add a few unordered comments from various Committee members, and summarize some ideas put forward by Kathryn Hammar, the Student Representative on the BCOE Executive Committee.

The BCOE Executive Committee's biggest overall concern has to do with the fact that many BCOE majors take the vast majority of their first- and second-year courses outside of BCOE in subjects such as English, Mathematics, Physics, Chemistry, and Biology. The SETs for the instructors who teach those courses are reviewed by members of those Departments and the respective Deans in CHASS and CNAS, and one BCOE Representative on CAP (who maintains confidentiality and provides feedback to no one). From the BCOE perspective, student "satisfaction" with their experience in these courses (as measured by SETs) is far less important than whether or not these courses have adequately prepared the students for success in upper-division engineering coursework for which these courses are prerequisites. Without loss of generality, an instructor who teaches an introductory course in calculus or physics who has low SETs via iEval is incentivized to improve those evaluations to optimize their likelihood of successful merit and promotion; this incentivization is completely decoupled from BCOE's interests, which is to ensure that students are properly prepared for rigorous upper-division coursework. The fact that the merit and promotion process at UCR is in part tied to student opinion of lower-division preparatory coursework, but is wholly decoupled from any actual measure of the effectiveness of that preparation, is fundamentally flawed.

While Subcommittee A's overall recommendation states in earnest that "there must be a mechanism for students to communicate their experience," that fact that UCR exclusively measures student satisfaction but not learning outcomes and academic preparation, and bases merit and promotion decisions on this factor alone with respect to classroom teaching, irreparably degrades the institution. The reality of this farce is that BCOE students struggle in their courses *and* BCOE instructors suffer from lower SETs due to frustrated students who were poorly prepared by non-BCOE instructors who were incentivized to optimize their SET scores; in turn, BCOE instructors are incentivized to play the same game. SETs incentivize the wrong objectives, and the result is that they fail to serve both students and faculty well; moreover, the question of whether or not SETs serve the interests of prospective employers of UCR graduates is not considered.

The problems with SETs are specific examples of a much more general problem with these types of assessments:

> Campbell's Law: The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.
>
> Goodhart's Law: Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.
>
> Goodhart's Law (as restated by Strathern): When a measure becomes a target, it ceases to be a good measure.

The current merit and promotion process uses SETs as a measure of teaching excellence; however, it is clearly a target: "How can I improve my SET score?" is not the same question as "How can I improve my teaching effectiveness?" The dichotomy is even greater for underrepresented faculty: the gap between "How can improve my SET score in the presence of bias?" and "How can I improve my teaching effectiveness?" may be insurmountable. Informing students about bias upfront and replacing

Another form of bias elucidated by Heffernan's paper was the difference in evaluation scores between instructors who teach qualitative and quantitative subjects. The paper reported that faculty who teach quantitative subjects were 3.3 times more likely to score in a lower evaluation bracket than those being evaluated by students in qualitative subjects. Those teaching quantitative subjects were also 1.88 times more likely to "fail" their evaluations. This is not surprising, as quantitative subjects have a reputation for requiring more effort than qualitative subjects, and do require strong grounding in prerequisite for students to do well in exams. When students are unwilling or unable to work to overcome inadequate preparation (which may not be their fault) they blame the instructor. SETs are particularly trenchant if students have managed to obtain good grades in their prerequisites due to grade inflation without developing a thorough understanding of the material. Once again, the incentives do not align with optimizing teaching effectiveness.

Another concern is accusations of academic misconduct. The policy of SCAIP requires instructors to report misconduct within 30 days of discovering it; given that students accused of misconduct are unlikely to give positive SETs, this disincentivizes reporting misconduct and punishes instructors who do so, except when misconduct is discovered late in the quarter. This issue becomes more salient with the emergence of websites like Chegg and CourseHero and the fact that many students have established their own communities using services such as Discord outside of UCR's purview. If academic misconduct is to be taken seriously, the merit and promotion system must protect faculty from adverse impacts when misconduct is reported, rather than creating a situation where they can explicitly be punished for doing their job properly.

Making matters worse, the merit and promotion process gives faculty little opportunity to respond to student reviews. There is insufficient space in the reasonably short self-statements that are used for merits; this is less of a concern for promotions, where the self-statements have no page limit. The BCOE Executive Committee supports efforts to institutionalize the use of self-reflective teaching statements, as suggested by the Academic Council Teaching Evaluation Task Force and the UC Centers for Teaching and Learning and suggested as a Phase I stopgap measure in the Ad Hoc Committee's report.

With respect to the contents of the Task Force Report, the BCOE Executive Committee makes the following comments.

Subcommittee A's overall recommendations:

- Subcommittee A's general suggestions to incrementally modify the iEval process is insufficient and disappointing, given the outcomes of the 2020 UCR campus survey on iEval reported in Appendix B. Changing the questions in iEval is a superficial change, as it does not adequately address student learning, student preparation for subsequent classes in prerequisite chains, etc.

- The BCOE Executive Committee views favorably the suggestion that "additional forms of evidence of teaching evidence be widely adopted." The report did not precisely clarify what other forms of evidence should be considered, although some ideas can be taken from the Academic Council Teaching Evaluation Task Force, but hopes that at the very least self-reflective teaching statements be included.

- The suggestion to maintain the five-point numerical scale but without comparison to Department and Campus means and medians is sensible; however, it is unclear what numerical value represents a threshold for acceptable teaching quality.

- The BCOE Executive Committee supports the nine specific recommendations for the iEval revision; however, there is some concern about Recommendation 5, which asks about Course Objectives. At present, Course Objectives are a best-practice, but are not institutionalized (although in BCOE, objectives are effectively required for ABET accreditation); however, the process of communicating

them to students has not been institutionalized. To the best of our understanding, it is recommended that instructors include the objectives on syllabi, but that may be insufficient when students are rapidly completing an iEval survey. It is also unclear if the learning objectives are specific to the course, or if different instructors can alter the learning objectives. To proceed in this fashion, some solution is needed to institutionalize them -- the most aggressive form of which might be imposing a Senate approval process (this is not an endorsement of the suggestion to do so).

- Subcommittee A's report incorrectly labeled Associate Professor Suveen Mathaudhu as being a member of the Physics Department; he is a member of the Mechanical Engineering Department.

- Among the potential iEval questions suggested by Faculty that are listed in Appendix D, the BCOE Executive Committee strongly opposes the following:

    - "Did this course lead you to want to take more courses in this field of study?" This question makes sense for students who are admitted to Colleges, but not directly to majors, and take many general education courses in the first two years while choosing their major. It does not make sense for students in BCOE who, depending on their major, take the major of their courses in the Calculus, Physics, Chemistry, and/or Biology sequences, plus English 001A and 001B; and only take courses within their majors in upper division. Likewise, it is not a good question for upper-division courses that are not prerequisite for other courses in the major, and are likely to be taken by students in their Junior and Senior years.

    - "The instructor appeared to reflect and tried to adapt their teaching when things didn't go as planned or when students were struggling." This question assumes that problems are the norm; it is unfair to ask the question if things do in fact go as planned when a course is offered; it also assumes that students are qualified to assess when things do or do not go as planned, and whether or not they and their peers struggle. It also places responsibility for struggling students on the instructor, which may or may not be appropriate, depending on the context.

    - "Do you feel the course was a valuable experience in helping to develop your own thinking/direction/approach on the subject?" This type of question might be appropriate for a course in the social sciences or humanities, but would not make sense for many courses that BCOE students take (e.g., Calculus).

With respect to Subcommittee B's report, the BCOE Executive Committee sees the potential for improvements in the current iEval process, but is not convinced that the fundamental flaws will be overcome. The UC Berkeley option to allow instructors to configure the evaluation questions is perhaps the most substantive improvement. Additionally, many of the specific items are quite thoughtful, such as Item 3: "How to incorporate for class size, level (lower/upper division, grad/undergrad,required/elective), type (lecture, on-line, hybrid, flipped), etc. in the interpretation of teaching evaluations?"

One concern is that Subcommittee B listed the following question as mandatory: "Overall Teaching Effectiveness (MANDATORY): Considering both the limitations and possibilities of the subject matter and the course, how would you rate the overall effective of this instructor?" This is similar to Question #13 on the current iEval, which Subcommittee A identified as being among "those questions most prone to bias [that] should be excluded."

Other potentially problematic questions include:

- **CATEGORY 1: Instructor-specific question themes:**
    - "Encouraging of participation/discussion: The instructor encouraged student questions and participation." While this question may seem innocuous, it isn't reasonable to expect participation and discussion in courses with 100+ students. Another concern is that many students do not attend due to various reasons, including jobs and perhaps misplaced priorities. The students actually do NOT want in-class participation to be incentivized or required, as their preference is not to attend. This question invites them to attack the instructor for what is generally considered to be good teaching practices.
    - "Useful/clear feedback on performance: The instructor provided clear constructive feedback." UCR has grown its undergraduate student population rapidly over the past decade, and class sizes have increased dramatically. It is not possible for instructors to provide substantive feedback throughout the quarter when class sizes grow above 20-30 students. Reductions in certain types of teaching quality should be expected when class sizes grow and adequate resources, such as funding for TAs and readers, are scarce.
- **Category II Course Specific Themes:**
    - "Application & Skill Development: The course developed my abilities and skills for the subject." The problem here is one that I alluded to earlier in this letter. A student may not know if they adequately developed their skills and abilities until a later course in the prerequisite chain. A student may feel confident in their abilities and skills at the completion of a Calculus course, but may find themselves woefully unprepared a few quarters later in an upper-division course on an engineering topic such as heat transfer.
    - "Theory/Content Knowledge: The course developed my ability to think critically about the subject." This question is biased toward classes in the social sciences or humanities. Critical thinking in subjects like mathematics may not really come until graduate school. Engineering disciplines require students to learn things like "how to compute the integral of common types of functions" or "how to write a for-loop in Python," not how to think critically about the greater ramifications (or lack thereof) when computing a Fourier Transform.

Subcommittee B's report failed to consider many factors that contribute to the student experience in a class, and may impact SET scores, but are beyond the control of the instructor. Example is facilities (e.g., classroom, laboratories, performance spaces, etc.), which are briefly mentioned in Subcommittee A's report, or issues like scheduling -- for example, students may be unhappy with a course that meets at 8am, without recognizing that this was not the instructor's choice. Examples of questions that could lead to better evaluation of these factors could include:

- Did the classroom environment affect learning?
- Did the scheduling affect learning?
- Adequacy of laboratories?

Somewhat more generally, the BCOE Executive Committee recommends that the iEval questions be vetted by faculty who regularly use surveys in their research.

The BCOE Executive Committee held a rather open-ended discussion on SETs that went far beyond the scope of the Ad-Hoc Committee Report. The Committee feels that many of these comments may be useful, they are reproduced here, in no specific order:

- iEval response rates were much higher when students were required to submit evaluations before seeing their grade. Now that they are optional, response rates are lower, and, similar to social media, tend to reflect the most polarized views. Opportunities for legitimate feedback are lost. There is general concern that they capture extreme opinion, not mean or median opinion, and that instructors are adjusting their teaching in responses to extremes. There could be severe adverse outcomes on student learning that are not measured as a consequence.

- A counterbalance to iEval, which are very short-term in terms of what they assess, could be continued engagement with alumni. Sometimes, the value of a class is not properly understood until years later, either in the workforce, graduate school, etc.

- BCOE is required to collect a monumental amount of assessment data as part of the ABET accreditation process. Moreover, BCOE programs are required to analyze and act upon that data. Instructors who assess and modify their courses through this continuous improvement process should be empowered to present evidence that they have done so in their merit and promotion files.

- Many students do not like some things that ABET requires, and may inadvertently blame the instructor on their SETs.

- It could make sense to drop the point/scoring system and focus exclusively on the written comments; coherent writing might yield a more holistic picture of an instructor.

- It is very challenging to tease out lies in the written comments (and presumably poor ratings given by the same students). This is especially true in cases where students have been accused of academic conduct violations, within 30 days of discovery, as required, but prior to the closing of the iEval period.

- Instructors could benefit from an iEval-based formative survey that is carried out midway through the quarter. This way instructors can show adjustment and can demonstrate impact with the present summative survey at the end of the class. These questions should focus both on the instructor and the student, for example:
    - What can the student do to make the class better?
    - What can the instructor do to make the class better?
    - A question about how the responses to the mid-quarter survey are used to adjust teaching could be asked at the end-of-quarter survey?

Lastly, the BCOE Executive Committee would like to highlight comments made by Kathryn Hammar, the Student Representative. Kathryn's comments both highlight the weaknesses of the current approach to SETs, but also put forward some alternative mechanisms to incorporate student feedback and opinion in a more constructive manner.

- Many students do not know why they are asked to fill out SET; they know nothing of their institutional purpose.

- Many students do not care about improving the course unless they get a bad grade, and they just want to take it out on the professor.

- Making SETs optional creates bias in terms of which students choose to fill them out.

- Numerical responses that don't clearly correlate to written comments that explain them are not particularly useful.

- Retention of learned information is challenging from one course to the next.

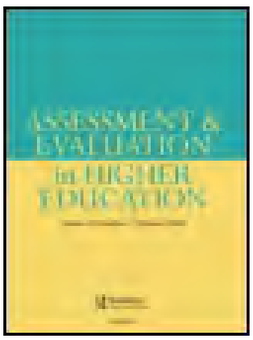
- Students will provide better qualified feedback if there is greater awareness and information about the role of iEval on campus. Student participation will increase if they believe that their feedback will improve future offerings of a course.

- Student leaders (e.g., Presidents and Vice Presidents of student organizations) can be an effective vehicle to promote change in the culture of how SETs are administered and interpreted. Student leaders could be trained to communicate the importance of critical feedback through SETs to their classmates. Examples where they can show evidence of change would increase student confidence that their opinions are taken seriously.

- Continuous feedback through the quarter, rather than just waiting until the end, would be beneficial.

- It is important to separate evaluation of teaching from the course and infrastructure, especially for majors that have many laboratory classes. For example, students may have a strong opinion about a laboratory manual, but may be unaware that it was not written by the instructor.

The BCOE Executive Committee feels that Kathryn's comments about engaging student leaders in the process of evaluating teaching is a step in the right direction, although the precise form of engagement, time commitment, potential compensation, etc. are open questions that we did not discuss.



Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/caeh20

Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching

Troy Heffernan

To cite this article: Troy Heffernan (2021): Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2021.1888075

To link to this article: https://doi.org/10.1080/02602938.2021.1888075

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
https://doi.org/10.1080/02602938.2021.1888075

Routledge
Taylor & Francis Group

Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching

Troy Heffernan

La Trobe University, Melbourne, Australia

ABSTRACT

This paper analyses the current research regarding student evaluations of courses and teaching. The article argues that student evaluations are influenced by racist, sexist and homophobic prejudices, and are biased against discipline and subject area. This paper's findings are relevant to policymakers and academics as student evaluations are undertaken in over 16,000 higher education institutions at the end of each teaching period. The article's purpose is to demonstrate to the higher education sector that the data informing student surveys is flawed and prejudiced against those being assessed. Evaluations have been shown to be heavily influenced by student demographics, the teaching academic's culture and identity, and other aspects not associated with course quality or teaching effectiveness. Evaluations also include increasingly abusive comments which are mostly directed towards women and those from marginalised groups, and subsequently make student surveys a growing cause of stress and anxiety for these academics. Yet, student evaluations are used as a measure of performance and play a role in hiring, firing and promotional decisions. Student evaluations are openly prejudiced against the sector's most underrepresented academics and they contribute to further marginalising the same groups universities declare to protect, value and are aiming to increase in their workforces.

KEYWORDS

Student evaluations; academic equity; wellbeing; prejudice; abuse; higher education

Introduction

This paper is part of a research project that investigates questions concerning student evaluations of courses and teaching (SETs). The project examines the extent to which SET results provide objective assessments of the course, the academic teaching the course, how biased the information may be, and what groups are potentially disadvantaged by these processes. It also investigates what adjustments can be made to create a more equitable system. Ample evidence indicates that course evaluations (that is, evaluations of a course in terms of its content and outcomes) and teaching evaluations (surveys dedicated to evaluating the teaching academic's performance), are both strongly correlated with the teaching academic's demographics and other issues unrelated to the course or the academic's performance (Boring, Ottoboni, and Stark 2016; Uttl and Smibert 2017). For this reason, course and teaching evaluations are treated as one within this paper (unless otherwise stated).

CONTACT Troy Heffernan t.heffernan@latrobe.edu.au

This paper's collation and analysis of existing SET research makes several points clear: SETs are significantly biased due to the demographics of students completing them, and prejudice against the academic teaching the course, are dependent on subject areas, and are impacted on by myriad other aspects not connected to the teacher or course. Yet despite these clear prejudices and biases, SETs are used to gauge teaching quality and are a component in judging who is hired, who is let go and who is promoted. In addition to these discriminatory practices, the trend of abusive comments in SETs is increasing (Tucker 2014). These prejudices in SET results, and how the results are used, are subsequently leading to growing academic mental health and wellbeing issues that universities cannot ignore (Jordan 2011; Fan et al. 2019).

Thus, this paper provides much needed synthesis and analysis of the existing research for the benefit of academics in every field and discipline who are subjected to these practices. Cunningham-Nelson, Baktashmotlagh, and Boles (2019) have built upon the *International Association of Universities* 'World List of Universities' (2006) and estimated that globally the teaching staff of over 16,000 higher education institutions collect SETs at the end of each teaching period. The paper makes clear that while SETs are being used as an aid in gauging performance, women and marginalised groups are losing jobs, being promoted slower and/or less often, and are being negatively impacted at career progression junctures within the academy (Uttl and Smibert 2017). This paper aims to inform higher education institutions, policy-makers and administrators of the sexist, racist, homophobic and other biases that underpin SET data as the evidence demonstrates the way institutions are complicit in prejudicial practices associated with SET data.

Methods

This paper analyses the literature and evident themes around SETs and the prejudices and biases that influence their results. The paper follows the methodology of a systematic analysis to provide a transparent account of how data was gathered and informs the audience of the protocols used to provide an account of the literature and themes from a breadth of sources (Macpherson and Holt 2007; McCrae and Purssell 2020).

Data collection

This paper's literature review method used the following initial criteria. The search began by seeking out research published between 1990 and 2020. The period was selected to incorporate the growth of SET surveys to include how they are used today, how technology has changed how surveys are conducted, and how technology has contributed to new data analysis methods. The title, abstract and keyword search terms included all versions of 'student evaluations', 'student evaluations of teaching', 'SETs', 'SECTs', 'course evaluations' and 'teacher evaluations' in English language peer-reviewed articles and books from standard institutional databases including EBSCO, ProQuest and Web of Science.

The advantage of these search terms is that research surrounding these topics is not limited to publications explicitly dedicated to one field; in this case education researchers or education-based journals and books (Pittaway et al. 2004). As student evaluations impact on every university discipline, research has been conducted in essentially all disciplinary areas. These findings sometimes appear in journals relating specifically to the subject area (such as medicine or engineering) rather than journals focusing specifically on teaching that subject. These search terms thus resulted in literature beyond the usual scope of education journals and higher education researchers concerned with teaching practices.

The search resulted in 293 publications. The list was then refined to remove the papers addressing student surveys not connected to SETs (86), duplicate results (55), book reviews (9),

media articles (4) and non-English language publications (3). This resulted in 136 publications meeting the criteria. Analysis of these 136 articles then led to the identification of 47 articles that were not returned in the initial search. In most cases, the papers did not appear in the initial search because they were not publications dedicated to SETs; rather, the discussion around student surveys was carried out for the purpose of informing audiences concerned with course design or professional development. However, as they contributed to current scholarly thought around SETs and their impact, they were included in the following analysis.

Data analysis

After an initial reading of the final collection of material, themes were generated using Braun and Clarke's (2006) method of thematic analysis. This analytic method provides a system by which patterns of meaning can be generated from qualitative data; in this case it was themes surrounding student evaluations of courses and teaching (Clarke and Braun 2017). This process identified 58 individual themes, which after further analysis were categorised into 35 sub-themes, before finally being refined into the five main themes discussed in this paper.

Findings

Evaluations technically work

A starting point of many publications examining SETs is an acknowledgement that evaluations work in the sense that they provide the university with data relating to course design, delivery of the course and teaching staff performance. The inherent problem with SETs, however, is that this data disguises the prejudices and biases underpinning the data being gathered, and subsequently in the results being produced (Marsh 2007; Osoian et al. 2010; Stark and Freishtat 2014).

That SETs provide data that appears sound is arguably why institutions believe that evaluations are a measure of effectiveness (Osoian et al. 2010; Tucker 2014) and therefore play a role in academic hiring and promotions (Arthur 2009; Shah and Nair 2012; Boring, Ottoboni, and Stark 2016). At the same time, it has been noted that SET results have been used to aid in firing unproductive staff, or guiding staffing decisions during times of restructure (Jones, Gaffney-Rhys, and Jones 2014; Uttl and Smibert 2017).

Apparent data quality is also what leads institutions to use SETs as signifiers of teaching standards. Many institutions expect staff to achieve a certain SET result (e.g. 3.75 out of five or higher, or over 75% etc.) to be seen as fulfilling their duties. Stark and Freishtat (2014) also found that some universities use SETs to intentionally incite continuous cycles of competition amongst academics by making the acceptable result one that is above the cohort's average. That is to say, no matter what the cohort's results, half of the teaching staff will be below the average and susceptible to the negative repercussions of not meeting the target such as decreased promotion chances or leadership opportunities.

The issue many researchers examining SETs find is that looking at the data presented by compiling evaluation results appears to provide a somewhat objective picture of teacher and course success. However, what is rarely considered, or perhaps not seen, by universities and researchers praising SETs are the prejudices and conditions that shape the views that form the data (Marsh 2007; Osoian et al. 2010; Stark and Freishtat 2014).

Impact of students' cultural and demographic backgrounds

The review found that students' backgrounds and demographics in terms of gender, age, disciplinary area and study type can all impact on an academic's SET results. That student demographics alone impact on SET results demonstrates just how flawed the system of evaluations

is, and to what extent results are determined by factors not related to course content or teaching quality (Rosen 2018).

Tucker's (2014) study of 43,000 course evaluations found that across every discipline and course evaluated, and regardless of teaching staff, women students submitted SETs that were overall more favourable by two per cent. However, the standard level of positive SET score increased by as much as six per cent when completed by international students, students from older age groups, external students and students with higher grade point averages (GPAs). That student demographics alone can make a difference of between two and six per cent in SET results is a significant finding. However, it must be reiterated that this is just one of the many biases and prejudices that can accumulate to greatly disadvantage some groups more than others.

Variances in SET results being provided by students from different disciplinary areas have also been noted. Beran and Violato's (2005) study of 370,000 evaluations and Centra's (2009) study of 238,000 evaluations found that, across multiple universities and countries, academics teaching science-based subjects receive lower results than those teaching social science/humanities-based subjects. Uttl and Smibert (2017) study of 325,000 SETs divided subject areas into quantitative and qualitative subjects; those being subjects with assessments based on right/wrong assessment (such as correct calculations and formulas) and those with assessor judgements such as essays. They found that those teaching quantitative subjects were 3.3 times more likely to score in a lower evaluation bracket than those being evaluated by students in qualitative subjects. Those teaching quantitative subjects were also 1.88 times more likely to fail their evaluations. With quantitative subjects continually being found to be evaluated lower by students, Uttl and Smibert (2017, p. 8) concluded that:

> Professors who teach quantitative vs. non-quantitative classes are not only likely to receive lower [SETs], but they are also at a substantially higher risk of being labelled unsatisfactory in teaching, and thus, more likely to be fired, not re-appointed, not promoted, not tenured, and denied merit pay.

Students are also influenced by factors not related directly to the course or teacher. Several studies have found high correlations between students' grade expectations and the SET scores they deliver. These studies have examined expectations based on GPAs and mid-term results from institutions across several countries. Repeatedly, the findings are that students who are graded higher, or expect to gain a high grade if the SET is completed before results are released, provide higher scoring evaluations (Worthington 2002; Short et al. 2008; Stark and Freishtat 2014; Boring, Ottoboni, and Stark 2016). These findings also raise pedagogical concerns as these results provide motivation for academics to set easier assessment, or perhaps grade easier, to facilitate better SET results. Considering SET results are used to aid in hiring, firing and promotional decisions, it cannot be ignored that academics may be motivated to alter their assessment processes given that their livelihoods are at risk (Carter and Lara 2016; Bachan 2017).

Studies have also found that SET results are being driven by student biases irrelevant to course content and effectiveness. Benton, Cashin, and Manhattan (2012) found class size to be a major factor. Issues including classroom design, cleanliness of the university, quality of course websites, library services, food options available on campus, and difficulty in the admissions process (for first year students) have all been found to play a larger role in influencing SET results than teaching quality or course design (Osoian et al. 2010).

Academic gender, ethnicity, sexual identity, and other demographics

Arguments have also been made that an academic's gender, ethnicity, language, perceived sexual identity, age or visible disabilities impacts on student evaluations (Valencia 2020). Such is the bias against gender and perceptions of ethnicity, sexuality, age and disability that, in 1993, Ambady and Rosenthal (1993) found that student reactions to a 30 second silent video of their

teacher played at the start of the semester correlated to the SET results the academic received. More recently, Boring, Ottoboni, and Stark (2016) study of 23,000 SETs, and Fan et al.'s (2019) study of 22,000 SETs, found that male students express a significant bias in favour of male academics.

Numerous studies have also found statistically significant differences between how gender influences academic evaluations. MacNell, Driscoll, and Hunt's (2015) work determined that women academics consistently receive lower scores relating to course design, clarity of assessment, class engagement, turnaround time of essays, and question response times regardless of their performance. Boring, Ottoboni, and Stark (2016) also found statistically significant examples of student expectations being amplified due to the academic's gender. Thus, they found that not only did SET results highly correlate with grade expectations, when grade expectations were met, male academics were rewarded with higher scores. When grade expectations were not met, the impact on evaluation scores was lower for male academics. To explore the extremes of gender prejudice in SETs, MacNell, Driscoll, and Hunt (2015) conducted a study that found that online classes led by male avatars (regardless of the academic's actual gender) received higher SET scores than those led by women avatars. Thus, students even perceiving the teaching academic to be a woman delivered lower scores.

A consistent theme within these studies is that gender, and even *perceived* gender, makes a difference to SET scores and is highly prejudiced against women. In Boring, Ottoboni, and Stark's (2016) summary of their study's findings and those of the existing studies they analysed, they declared that SETs are 'biased against female instructors by an amount that is large and statistically significant' (p. 1).

Many of these studies find that these biases and prejudices result in large and significant variations in SET results, but this raises the question of what these differences can look like in practice. Fan et al.'s (2019) study of 22,000 SET results concluded that at the extreme, women academics are receiving SET scores 37 percentage points lower than male academics. This figure represents science subjects with high numbers of male students and being led by younger female academics (approximately under 35 years old). Boring, Ottoboni, and Stark's (2016) study found similar results and concluded that the biases at play were so great that more effective female academics are being placed in lower SET grading brackets than their less effective male counterparts.

In all studies relating to gender, the analyses indicate that the highest scores are awarded in subjects filled with young, white, male students being taught by white English first language speaking, able-bodied, male academics who are neither too young nor too old (approx. 35–50 years of age), and who the students believe are heterosexual. Most deviations from this scenario in terms of student and academic demographics equates to lower SET scores. These studies thus highlight that white, able-bodied, heterosexual, men of a certain age are not only the least affected, they benefit from the practice. When every demographic group who does not fit this image is significantly disadvantaged by SETs, these processes serve to further enhance the position of the already privileged.

Prejudice in SETs stemming from the academic's ethnicity is also a common finding in studies concerning evaluations and biases, as are issues around age, disability, sexual identity and appearance (Andersen and Miller 1997; Cramer and Alexitch 2000; Worthington 2002). A primary issue with these studies is that for large-scale quantitative research, the current state of higher education inclusion and diversity means results concerning marginalised groups are deemed too small to be considered a valid sample size (Hendrix 1998; Rubin 1998; Fan et al. 2019). However, smaller scale surveys and studies relying on qualitative methods consistently find prejudices against academics of colour (DiPietro and Faye 2005; Hamermesh and Parker 2005).

Significant bias in SETs have been found to impact negatively on academics of colour (Hendrix 1998; Rubin 1998), and academics whose native language is not that of the university. In most studies this means the English language, though some studies have explored issues of SET prejudice in European non-English speaking countries and found similar results (DiPietro and

Faye 2005; Hamermesh and Parker 2005). Fan et al.'s (2019) study also found that academics from diverse backgrounds or non-English first language backgrounds faced different levels of prejudice according to subject area; more liberal subjects were less prejudiced but were prejudiced nonetheless. Domestic students are also more likely to provide lower SET scores to academics from these groups than international students (Tucker 2014).

Providing further evidence of the significant prejudice women academics face in the SET process, even though all academics from ethnically diverse and marginalised backgrounds received lower SET scores than their white, English first language speaking colleagues, women from ethnically diverse backgrounds are graded more harshly than men from ethnically diverse backgrounds. At the extreme, Fan et al. (2019) found that in science faculties, a woman from a non-English speaking background was half as likely to receive the same SET result as a white English-speaking male.

SET comments

A key component of SETs in the literature that promotes their use is their anonymity. Universities argue that anonymity allows students to provide honest feedback without fear of retribution for speaking negatively against the teaching academic; the implication being that students may feel unable to make negative comments if they can be identified (Tucker 2014; Uttl and Smibert 2017). Studies have nonetheless determined that the issue with anonymous comments in SETs is that a portion of the comments are abusive, that the abuse is growing, and that the abuse is mostly directed towards race, gender, sexual identity, ethnicity, age and other marginalising characteristics.

The literature regarding comments in SETs is somewhat limited. Tucker's (2014) study of 43,000 SETs suggested that only around one per cent of comments were abusive (though she noted sharp increases across studies, and hypothesised that the trend was growing rapidly). However, even if the figure of one per cent is accurate, it is imperative to point out that the one per cent is not distributed evenly across academic demographics. Additionally, to the person receiving the abusive comments, the emotional damage and stress is real, and the overall rate of abusive comments is irrelevant to their stress, anxiety or mental wellbeing (Tucker 2014).

As might be expected considering the clear trends within the findings of this review, a white male who is perceived to be heterosexual and is in the 35–50-year-old age group will, statistically speaking, receive few, if any, abusive comments. Abusive comments are mostly directed towards women and marginalised groups, focus on marginalising characteristics, and they are cumulative. For example, women receive abusive comments, and academics of colour receive abusive comments, thus, a woman of colour is more likely to receive abuse because of her gender and her skin colour (Oliver et al. 2008; Jordan 2011). One way to consider this finding is that some groups are so underrepresented in the sector that they do not constitute a valid sample size in large-scale studies. It is women and academics from these underrepresented groups that receive a majority of the abusive comments which makes the notion of 'only one per cent of comments being abusive' a highly distorted figure.

The result of SET comments is that they are a source of anxiety for academics. The process of SETs alone is cause for concern being that they are used for firing, hiring and promotion purposes (Jones, Gaffney-Rhys, and Jones 2014; Uttl and Smibert 2017). However, that comments for the sector's women and marginalised groups are then likely to be at best unconstructive and unjustified, and at worst racist, sexist, homophobic or ageist (among other prejudices) is only a further cause for concern and mental distress for the academics receiving these comments (Jordan 2011; Tucker 2014).

Jones, Gaffney-Rhys, and Jones (2014) considered the legal implications of universities continuing to allow SETs to be collected when they are known to be a cause of distress. They discuss defamation and the university's potential breaches of duty of care, and their suggestions

add further reason for SETs to be removed. Where comments are concerned, however, it is likely that if a student who could be identified provided racist, sexist, homophobic or other abusive feedback they would be removed from classes or face other consequences.

What is known from previous SET research?

For all the research and studies conducted around SETs, their conclusions largely all point to similar findings. A frequent conclusion across large-scale quantitative and smaller qualitative studies is that SETs rarely measure course or teacher quality or effectiveness, or if they do, these elements are significantly outweighed in the final results by other factors unrelated to course or teacher quality (Stark and Freishtat 2014; Tucker 2014; Boring, Ottoboni, and Stark 2016; Fan et al. 2019).

As has been discussed, SETs are heavily influenced by student demographics and subject area, but most studies argue that the two greatest influences are the academic's gender and culture (Stark and Freishtat 2014; Boring, Ottoboni, and Stark 2016; Boring 2017; Fan et al. 2019). It is also critical to note that these researchers, and those of many more studies within the literature review, have suggested that it is impossible to account for these biases because from one class, subject, let alone one university, to another, the variances cannot be predicted or accounted for in the data.

Many researchers have not only questioned the use of SETs, but also the value of the content they provide. Cunningham-Nelson, Baktashmotlagh, and Boles (2019) make the point that even if SETs were not guided by biases and prejudice, the students completing the evaluation will not benefit from the process. The surveyed group will have moved on before changes can be made and the next cohort may value different attributes in the course or teacher. Researchers additionally argue that SETs are limited in what they provide because the qualities students place value on are already known (Sunindijo 2016). The existing literature declares that students want engaging lectures, assessment to be explained clearly and graded fairly, assessment returned in a timely manner and their questions answered promptly (Sunindijo 2016; Park and Dooris 2020). Vivanti, Haron, and Barnes (2014) came to similar conclusions, and add that students want guidance, and teaching staff who were interested in them and the subject they are teaching.

Similar arguments also appear in the literature alongside the conclusion that SETs are measuring customer satisfaction (Enache 2011; Osoian et al. 2010). These researchers acknowledge that the marketisation of the higher education system means customer satisfaction plays a valuable role in the modern university because enrolment numbers matter (Marginson 2013). However, these researchers are all highly critical that SETs are customer satisfaction surveys alleging to measure course outcomes and teacher effectiveness when many studies argue that this is untrue. These studies point out that customer satisfaction can be measured via methods that are not SETs, and suggest opportunities within the growing programs of universities embracing student voice and students as partners initiatives.

Discussion

This paper's analysis of the existing literature makes it clear that SET results are strongly influenced by external factors unrelated to course content or teacher performance. In addition, these factors are frequently based on student demographics, and students' biases and prejudices based on the teaching academic's gender, sexuality, ethnicity, age or disability as well as other marginalising factors. Ultimately, this analysis raises the question of how can any university justify the continued use of SETs?

Official university responses to issues regarding SETs are rare. Tucker (2014) sought clarification of why a university continued using SETs when they were known to attract abusive

comments. The university's response was that the rate of abusive comments was too low to alter the existing procedures considering the value (they perceived) in the attained data. This response provides some insights into institutional administrative thinking. The first is that universities consider there to be an acceptable level of abuse that staff must endure. A factor here, however, is that Tucker collected and evaluated this information almost a decade ago. Considering the current focus on staff wellbeing and mental health (Henning 2018), it is possible that institutions today might have different views on academics receiving abuse. Universities might also be influenced by the knowledge that it is now known that abuse is directed heavily towards women and academics from marginalised groups. Thus, institutions continuing to conduct SETs are allowing the sector's most underrepresented and marginalised groups to be subjected to possible hate speech (Jones, Gaffney-Rhys, and Jones 2014; Uttl and Smibert 2017).

The second issue Tucker's (2014) research raised (as have other researchers in this literature review) is that universities believe the data they are collecting via SETs provides accurate assessments of course content and outcomes, and teacher quality and effectiveness. This paper's findings make it clear why universities might believe this to be the case. Methodologically, the process appears sound. Students provide anonymous feedback so as to not be concerned with potential repercussions from academics, and this data informs the faculty and administrators of students' perspectives concerning the course and teacher (Arthur 2009; Shah and Nair 2012; Jones, Gaffney-Rhys, and Jones 2014; Boring, Ottoboni, and Stark 2016). However, this review's findings also demonstrate that the methodology surrounding SETs is inherently flawed because the data being input into the system is influenced by biases and prejudices that are invisible in the data's results (Marsh 2007; Osoian et al. 2010; Stark and Freishtat 2014). These elements are also invisible because unless researchers specifically focus on biases in SETs, there is little reason to believe that the questions being asked will not be answered somewhat objectively across a class of respondents and provide an overall, somewhat accurate measure. However, this is not the case; this paper has shown that statistically significant biases in these data collection methods exist.

The groups most impacted by these prejudices are clear. This study demonstrates that at best SETs disadvantage women, and at worst, see women academics placed in untenable positions (MacNell, Driscoll, and Hunt 2015; Boring, Ottoboni, and Stark 2016). These results have also been theorised as a reason why women are underrepresented in both the professoriate and upper levels of university leadership (Fan et al. 2019).

The results are worse still when ethnicity, language, perceptions of sexuality or disability are considered. People from ethnic backgrounds, and/or who do not speak English as their first language, are receiving much lower SET results (Fan et al. 2019). Other researchers have examined disability, sexual identity, and cultural and linguistic diversity and found similar results in prejudice against these groups (Hendrix 1998; Rubin 1998; DiPietro and Faye 2005; Hamermesh and Parker 2005), though it cannot be ignored that researching prejudice against some of these groups is difficult because they are so underrepresented within the sector.

Researchers have also argued for decades that subject areas impact on SET results with the sciences being evaluated more harshly (Beran and Violato 2005; Centra 2009; Uttl and Smibert 2017). Similarly, that studies have routinely found high correlations between student grades and SET results makes it clear that student evaluations are being influenced by factors far exceeding the intended purpose of SET questions (Worthington 2002; Short et al. 2008; Stark and Freishtat 2014; Boring, Ottoboni, and Stark 2016). Studies have also determined that issues including classroom design, cleanliness of the university, quality of course websites, library services and food options available on campus all have a larger influence on SET results than practices concerning courses and teaching (Felton, Mitchell, and Stinson 2004; Felton et al. 2008; Osoian et al. 2010; Benton, Cashin, and Manhattan 2012).

In addition to the above findings, many researchers also argue that SETs are not needed because universities already know what students want from a course and teacher (Vivanti, Haron, and Barnes 2014; Sunindijo 2016; Park and Dooris 2020). These researchers claim that focus

groups or student interviews would be a more equitable way of gaining information and judging potential biases and prejudice. What these views make clear is that the growth in student voice, and students as partners initiatives, in recent years are well placed to provide discussions to improve learning opportunities rather than relying on the current prejudice practices.

Conclusion

This paper has shown that no university, and indeed the higher education sector as a whole, can declare to be a gender equal employer or have an interest in growing a safe, inclusive and diverse workforce if they continue using SETs to evaluate course and teacher quality.

This paper provides an evidence base which can be used as part of the growing material and argument against the practice of collecting SET data. When SET data is known to be highly prejudiced against many groups, methods must be changed, and using SET data as a component of hiring, promotion and/or firing decisions must be seen as the blatantly discriminatory practice that it is.

The need for immediate policy changes is clear. Women and marginalised groups are losing jobs, failing to achieve promotion, and are being negatively impacted at every step where SETs are concerned, and will continue to be discriminated against every time SET data is collected until the practice is stopped. These practices not only harm the sector's women and most underrepresented and vulnerable, it cannot be denied that SETs also actively contribute to further marginalising the groups universities declare to protect and value in their workforces.

ORCID

Troy Heffernan http://orcid.org/0000-0002-8156-622X

References

Ambady, N., and R. Rosenthal. 1993. "Half a Minute: Predicting Teacher Evaluations from Thin Slices of Nonverbal Behavior and Physical Attractiveness." *Journal of Personality and Social Psychology* 64 (3): 431–441. doi:10.1037/0022-3514.64.3.431.

Andersen, Kristi, and Elizabeth D. Miller. 1997. "Gender and Student Evaluations of Teaching." *PS: Political Science and Politics* 30 (02): 216–219. doi:10.1017/S1049096500043407.

Arthur, L. 2009. "From Performativity to Professionalism: Lecturers' Responses to Student Feedback." *Teaching in Higher Education* 14 (4): 441–454. doi:10.1080/13562510903050228.

Bachan, R. 2017. "Grade Inflation in UK Higher Education." *Studies in Higher Education* 42 (8): 1580–1600. doi:10.1080/03075079.2015.1019450.

Benton, S., W. Cashin, and K. Manhattan. 2012. Student ratings of teaching: A summary of research and literature. *IDEA Center*. http://www.ideaedu.org

Beran, T., and C. Violato. 2005. "Ratings of University Teacher Instruction: How Much Do Student and Course Characteristics Really Matter?" *Assessment and Evaluation in Higher Education* 30 (6): 593–601. doi:10.1080/02602930500260688.

Boring, A. 2017. "Gender Biases in Student Evaluations of Teaching." *Journal of Public Economics* 145: 27–41. doi:10.1016/j.jpubeco.2016.11.006.

Boring, A., K. Ottoboni, and P. Stark. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *Science Open Research*. doi:10.14293/s2199-1006.1.sor-edu.aetbzc.v1.

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. doi:10.1191/1478088706qp063oa.

Carter, M., and P. Lara. 2016. "Grade Inflation in Higher Education: Is the End in Sight?" *Academic Questions* 29 (3): 346–353. doi:10.1007/s12129-016-9569-5.

Centra, J. 2009. *Differences in Responses to the Student Instructional Report: Is It Bias?* Princeton: Educational Testing Service.

Clarke, V., and V. Braun. 2017. "Thematic Analysis." *The Journal of Positive Psychology* 12 (3): 297–298. doi:10.1080/17439760.2016.1262613.

Cramer, K., and L. Alexitch. 2000. "Student Evaluations of College Professors: Identifying Sources of Bias." *Canadian Journal of Higher Education* 30 (2): 143–164. https://journals.sfu.ca/cjhe/index.php/cjhe/article/view/183360.

Cunningham-Nelson, S., M. Baktashmotlagh, and W. Boles. 2019. "Visualizing Student Opinion through Text Analysis." *IEEE Transactions on Education* 62 (4): 305–311. doi:10.1109/TE.2019.2924385.

DiPietro, M., and A. Faye. 2005. "Online Student-Ratings-of-Instruction (SRI) Mechanisms for Maximal Feedback to Instructors." *30th Annual Meeting of the Professional and Organizational Development Network*. Milwaukee, WI.

Enache, I. 2011. "Customer Behaviour and Student Satisfaction." *Economic Sciences* 4 (53): 41–46.

Fan, Y., L. Shepherd, D. Slavich, D. Waters, M. Stone, R. Abel, and E. Johnston.2019. "Gender and Cultural Bias in Student Evaluations: Why Representation Matters." *Plos ONE* 14 (2): e0209749. doi:10.1371/journal.pone.0209749.

Felton, J., P. Koper, J. Mitchell, and M. Stinson. 2008. "Attractiveness, Easiness, and Other Issues: Student Evaluations of Professors on Rate My Professors." *Assessment and Evaluation in Higher Education* 33 (1): 45–61. doi:10.2139/ssrn.918283.

Felton, J., J. Mitchell, and M. Stinson. 2004. "Web-Based Student Evaluations of Professors: The Relations between Perceived Quality, Easiness and Sexiness." *Assessment and Evaluation in Higher Education* 29 (1): 91–108. doi:10.1080/0260293032000158180.

Hamermesh, D., and A. Parker. 2005. "Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity." *Economics of Education Review* 24 (4): 369–376. doi:10.1016/j.econedurev.2004.07.013.

Hendrix, K. 1998. "Student Perceptions of the Influence of Race on Professor Credibility." *Journal of Black Studies* 28 (6): 738–764. doi:10.1177/002193479802800604.

Henning, M. 2018. *Wellbeing in Higher Education: Cultivating a Healthy Lifestyle Among Faculty and Students*. Abingdon: Routledge.

International Association of Universities. 2006. *World List of Universities, 25th Edition: And Other Institutions of Higher Education*. 25th ed. London: Palgrave Macmillan.

Jones, J., R. Gaffney-Rhys, and E. Jones. 2014. "Handle with Care! An Exploration of the Potential Risks Associated with the Publication and Summative Usage of Student Evaluation of Teaching (SET) Results." *Journal of Further and Higher Education* 38 (1): 37–56. doi:10.1080/0309877X.2012.699514.

Jordan, D. W. 2011. "Re-Thinking Student Written Comments in Course Evaluations: Text Mining Unstructured Data for Program and Institutional Assessment." PhD diss., California State University, Stanislaus.

MacNell, L., A. Driscoll, and A. Hunt. 2015. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40 (4): 291–303. doi:10.1007/s10755-014-9313-4.

Macpherson, A., and R. Holt. 2007. "Knowledge, Learning and Small Firm Growth: A Systematic Review of the Evidence." *Research Policy* 36 (2): 172–192. doi:10.1016/j.respol.2006.10.001.

Marginson, S. 2013. "The Impossibility of Capitalist Markets in Higher Education." *Journal of Education Policy* 28 (3): 353–370. doi:10.1080/02680939.2012.747109.

Marsh, H. 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*, edited by R. Perry and J. Smart, 319–383. Dordrecht: Springer.

McCrae, N., and E. Purssell. 2020. *How to Perform a Systematic Literature Review: A Guide for Healthcare Researchers, Practitioners and Students*. Berlin: Springer.

Oliver, B., B. Tucker, R. Gupta, and S. Yeo. 2008. "eVALUate: An Evaluation Instrument for Measuring Students' Perceptions of Their Engagement and Learning Outcomes." *Assessment and Evaluation in Higher Education* 33 (6): 619–630. doi:10.1080/02602930701773034.

Osoian, C., R. Nistor, M. Zaharie, and H. Flueras. 2010. "Improving Higher Education through Student Satisfaction Surveys." *Proceedings of the 2nd International Conference on Education Technology and Computer*. doi:10.1109/icetc.2010.5529347.

Park, E., and J. Dooris. 2020. "Predicting Student Evaluations of Teaching Using Decision Tree Analysis." *Assessment and Evaluation in Higher Education* 45 (5): 776–793. doi:10.1080/02602938.2019.1697798.

Pittaway, L., M. Robertson, K. Munir, D. Denyer, and A. Neely. 2004. "Networking and Innovation: A Systematic Review of the Evidence." *International Journal of Management Reviews* 5 (3–4): 137–168. doi:10.1111/j.1460-8545.2004.00101.x.

Rosen, A. 2018. "Correlations, Trends and Potential Biases among Publicly Accessible Web-Based Student Evaluations of Teaching: A Large-Scale Study of RateMyProfessors.com Data." *Assessment and Evaluation in Higher Education* 43 (1): 31–44. doi:10.1080/02602938.2016.1276155.

Rubin, D. 1998. Help! My Professor (or doctor or boss) doesn't talk English. In *Readings in cultural contexts*, edited by J. Martin, T. Nakayama, & L. Flores, 149-159. Mountain View, CA: Mayfield.

Shah, M., and C. Nair. 2012. "The Changing Nature of Teaching and Unit Evaluations in Australian Universities." *Quality Assurance in Education* 20 (3): 274–288. doi:10.1108/09684881211240321.

Short, H., R. Boyle, R. Braithwaite, M. Brookes, J. Mustard, and D. Saundage. 2008. "A Comparison of Student Evaluation of Teaching with Student Performance." In *Proceedings of the 6th Australian Conference on Teaching Statistics*, edited by H. MacGillivray, 1–10. Melbourne, Victoria, Australia.

Stark, P., and R. Freishtat. 2014. "An Evaluation of Course Evaluations." *Science Open Research* doi:10.14293/s2199-1006.1.sor-edu.aofrqa.v1.

Sunindijo, R. 2016. "Teaching First-Year Construction Management Students: Lessons Learned from Student Satisfaction Surveys." *International Journal of Construction Education and Research* 12 (4): 243–254. doi:10.1080/15578771.2015.1121937.

Tucker, B. 2014. "Student Evaluation Surveys: Anonymous Comments That Offend or are Unprofessional." *Higher Education* 68 (3): 347–358. doi:10.1007/s10734-014-9716-2.

Uttl, B., and D. Smibert. 2017. "Student Evaluations of Teaching: Teaching Quantitative Courses Can Be Hazardous to One's Career." *Peer Journal* 5: e3299. doi:10.7717/peerj.3299.

Valencia, E. 2020. "Acquiescence, Instructor's Gender Bias and Validity of Student Evaluation of Teaching." *Assessment and Evaluation in Higher Education* 45 (4): 483–495. doi:10.1080/02602938.2019.1666085.

Vivanti, A., N. Haron, and R. Barnes. 2014. "Validation of a Student Satisfaction Survey for Clinical Education Placements in Dietetics." *Journal of Allied Health* 43 (2): 65–71.

Worthington, A. 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. doi:10.1080/02602930120105054.

UC RIVERSIDE Academic Senate

COMMITTEE ON ACADEMIC FREEDOM

February 9, 2021

To: Jason Stajich, Chair
Riverside Division Academic Senate

From: Frederick Wilhelm, Chair
Committee on Academic Freedom

Re: Final Report from the Ad Hoc Committee on Evaluation of Teaching

The UCR Senate Committee on Academic Freedom reviewed the Final Report from the Ad Hoc Committee on Evaluation of Teaching, and did not find any eminent concerns regarding Academic Freedom.

UC RIVERSIDE Academic Senate

COMMITTEE ON ACADEMIC PERSONNEL

February 18, 2021

To: Jason Stajich, Chair
Riverside Division Academic Senate

From: Yinsheng Wang, Chair
Committee on Academic Personnel

Re: [Campus Review] Report Review: Final Report from the Ad Hoc Committee on Evaluation of Teaching

CAP discussed the Final Report from the Ad Hoc Committee on Evaluation of Teaching. CAP agrees that there are limitations of current student evaluations for classes and that it is important to consider other forms of teaching and teaching evaluations in merit/promotion processes. In this vein, CAP suggests that faculty peer assessment could play a useful role in evaluating faculty teaching efficacy. In particular, this could provide a more objective measure for large service courses where the current student assessments are often not objective.

CAP endorses the recommendations of the two-phase plan outlined in the report. In addition, CAP would like to note the following:

CAP recognized a decreasing trend in response rate from student evaluations, which makes it challenging to evaluate teaching effectiveness. CAP suggests the resumption of the use of incentives (e.g. earlier release of grades) and reduce the number of questions in the evaluation form to encourage students to complete course evaluations.

Some CAP members voiced the concern that inclusion about reminder of the presence of bias in teaching evaluation may discourage students from completing the evaluations, thereby further reducing response rate. Moreover, some suggested that training can be provided to students to improve their awareness of potential bias while evaluating teaching effectiveness of faculty.

CAP also suggests that effective teaching evaluation focuses more on evaluating learning outcome of students than evaluating faculty skills.

UC RIVERSIDE Academic Senate

COMMITTEE ON EDUCATIONAL POLICY

March 10, 2021

To: Jason Stajich, Chair
Riverside Division

From: Stefano Vidussi, Chair
Committee on Educational Policy

RE: Ad Hoc Committee on Evaluation of Faculty Teaching's Final Report

The Committee on Educational Policy (CEP) reviewed the Ad Hoc Committee on Evaluation of Faculty Teaching's Final Report at their March 5, 2021 meeting. The Committee noted concern that the report does not provide a plan to implement the recommendations. The Committee noted a need for questions to be added to evaluations that measure students' preparedness for the course, understanding of course materials, and documentation of academic integrity.

UC RIVERSIDE

College of Humanities, Arts, and Social Sciences
EXECUTIVE COMMITTEE

March 31, 2021

TO: Jason Stajich, Chair
Riverside Division of the Academic Senate

FROM: Lucille Chia, Chair Lucille Chia
CHASS Executive Committee

RE: Final Report of the Ad Hoc Committee on Evaluation of Faculty Teaching

---

The following are comments based on the CHASS Executive Committee's review of this report. The main points which the EC noted with approval are:

1. the detailed recommendations to revise the current Student Evaluation of Teaching (SET) to reduce the bias toward women faculty and non-White faculty, and to emphasize student learning rather than unclear evaluation of the instructor
2. the insistence on evaluating a faculty member's pedagogy through not just SET but other kinds of evaluations such as a teaching portfolio, a teaching self-statement, and observations by a colleague.

The EC was not surprised that the survey (p. 1) of faculty experience and opinions of the current iEval system revealed "widespread dissatisfaction with the current system."

The EC agrees with Subcommittee A's recognition that "the problem of bias as among the most serious flaws in our current SET and its uses on our campus." On Subcommittee A's overall recommendations: the EC

--agrees with #1, that students should be better educated to understand the purposes of the evaluations and learn how to give suitable ("helpful") feedback, although the challenge is how to find the resources (time, instructors, training program, etc.) to do so

--#2: agrees that there should be fewer and more precise questions

--#3: questions referred to don't seem to be consistent with those listed in Appendix C

--#4: thinks that the stress "Student evaluation of learning" rather than instructor performance seems like a good idea, but this would require effective implementation of #1, in particular, that the students are educated in understanding how to evaluate their own learning, a difficult task, and as the report suggests, students may not accurately appreciate what they have learned until some time after they have taken the course, possibly after they have left UCR—any way to make this recommendation practical?

--#5: "Consider including the instructor's stated learning outcomes or learning objectives for the course as a component of the student evaluation." Again, this requires educating the

students to be able to evaluate appropriately the instructor's goals (e.g., a student looking for infotainment may not appreciate an instructor's less than entertaining approach to lecturing)

#6: "Narrative comment fields should be retained, but the iEval survey should provide carefully crafted prompts that guide the scope of student comments rather than simply having an open-ended Comments field." This is a good suggestion, perhaps even more than the subcommittee's original intent: it provides those reading the evaluations a fuller sense of a student's approach toward the class and provides clues to how much effort the student put into the class, and what the student's notion of a college course requires.

#7: "A statement or "preamble" about bias should be presented to students at the beginning of the survey (see example below)." --a good suggestion, though how effective remains to be seen

#8: "A statement or "preamble" about bias should be presented to students at the beginning of the survey (see example below)." definitely

#9: the problematic correlation of the numerical evaluation and the related narrative answer is important to address: "The placement of prompts for written comments relative to scale-based questions may influence numerical ratings. It is possible that providing written comments first, particularly if the student is invited to reflect on their role in the learning process, may provoke a more thoughtful and less biased consideration of scores on scale-based questions."

This may alert those reading the evaluations to be aware of comments from students who seek to punish instructors by giving them the lowest possible numerical evaluations and then extremely negative and even false comments.

The EC also notes that the persistent problem of low response rates for the current SET system, due to a number of factors, which the report did not address. For example, iEval is not a user-friendly system for students using a mobile phone—something that ITS should be invited to work on. In any case, until response rates can be increase for the SET, skewed numerical data may lead to problematic assessment, especially for departments that put a great deal of emphasis in these data.

As for the part of the report by Subcommittee B, which had the broader task of giving the campus "recommendations to the campus on how to interpret and use teaching evaluations responsibly for merit/promotions", the EC found it generally more problematic than the discussion of Subcommittee A.

For the four categories of questions discussed in recommendations 1 and 2,: category 3 should **not be replaced** by questions by those that "reflect those aspects of the course that directly impact student learning and the use of technology and degree of engagement in the absence of personal contact."

Considering that online instruction is an inevitable part of future UC teaching. And the subcommittee admits, "Many institutions of higher education are still struggling with how to effectively evaluate online formats for courses." We therefore need as much information from these evaluations of teaching and learning as possible.

Under Items 4&5: "Teaching effectiveness does not refer only to in-class teaching but includes all other aspects of educating students such as, but not limited to, general guidance, mentoring, supervising, and advising at all levels." **This should be put in bold and caps.**

The last several paragraphs of Subcommittee B's report make worthy recommendations about facilitating (and "enforcing"—the subcommittee's term) the serious consideration of teaching effectiveness through means other than SET, but again, the implementation of these recommendations is challenging. For instance, for what the subcommittee terms "peer review" (visits to classes, examining the syllabus and teaching materials, etc.) require much work by the candidate's peers will be time consuming and may prove difficult to assess knowledgeably and equitably if the observers are insufficiently familiar with subject/field of the course in general.

UNIVERSITY OF CALIFORNIA
LET THERE BE LIGHT
RIVERSIDE

UC RIVERSIDE | UNIVERSITY OF CALIFORNIA
College of Natural and Agricultural Sciences
Executive Committee

1 April 2021

To: Jason Stajich, Chair
Riverside Division

From: Theodore Garland, Jr., Chair, Executive Committee
College of Natural and Agricultural Science

Re: Campus Review: Report Review: Final Report from the Ad Hoc Committee on Evaluation of Teaching

Overall, the Committee expressed support of the report and noted that it had concrete, short-term fixes to address issues of bias and also to make sure that we follow our own evaluations as evidence of teaching. There was some brief discussion of student response rates, which were not mentioned at length in the report. Members also shared some examples of their own positive experiences with faculty peer evaluation.

Cheers,

Ted Garland

UC RIVERSIDE Academic Senate

COMMITTEE ON DIVERSITY, EQUITY, & INCLUSION

February 23, 2021

To: Jason Stajich, Chair
Riverside Division Academic Senate

From: Xuan Liu, Chair
Committee on Diversity, Equity, & Inclusion

Re: Final Report from the Ad Hoc Committee on Evaluation of Teaching

CoDEI reviewed the Final Report of the Ad Hoc Committee on Evaluation of Faculty Teaching. We are generally positive on the steps that are being taken to address bias in student evaluations of teaching (SETs) and the process of teaching evaluation for merit and promotion at UC Riverside. Following the format of the report itself, we have divided our response into short-term (Phase I) and long-term (Phase II) comments. Our short-term comments include the following:

- It could be helpful to compile evidence of the existing bias in the evaluation of teaching at the university, based on historical data. Although these statistics cannot serve as proof of bias, it will allow administrators to (1) discover where bias may be present within our own evaluation process, which can guide the decisions about the benefit of potential changes, and (2) later observe if the implemented changes appear to reduce evidence of bias in SETs and the merit and promotion process.
- Add information to the proposed preamble about the purpose of SETs, which will help students understand how they are contributing to teaching evaluation and how biased responses could be detrimental to the individuals involved.
- We look favorably on the option for faculty to include other forms of evidence of teaching quality, particularly the self-reflective teaching statement. It may be helpful to give faculty information on how they can use a teaching statement to address and alleviate bias in their SETs.

In response to the long-term proposals, we have the following comments:

- Any future changes would apply to all faculty, regardless of their likelihood of being discriminated against in the evaluation process. It is possible these changes could benefit the unintended group more than the intended group, which would widen any unfair gaps. For example, there is evidence that giving parental leave to all parents actually puts female faculty at a disadvantage when applying for promotion because male faculty use the leave to improve their research more than contributing to household production (https://www.iza.org/publications/dp/9904). Any new policies should be created with this possibility in mind.

- Assuming that numerical teaching evaluation scores remain, removing numerical department means could result in increased bias at CAP because it leaves CAP members to construct their own expectations and comparison groups.
- Several members of the committee have broader concerns about the value of SETs and prefer that the university consider more meaningful reforms, including the possibility of eliminating SETs entirely. As a minimum, they suggest that SETs be used as a formative, rather than summative assessment. This requires building mechanisms to assess how instruction was modified in response to SETs, rather than looking at the SETs themselves to assess teaching effectiveness.
- Reducing bias is the primary goal of this endeavor, but a legitimate secondary goal should be to incentivize faculty to increase their commitment to student learning and success. We believe that using this secondary goal as a guide in the reform will benefit faculty, students and the university community in the long run.

UC RIVERSIDE

Academic Senate

COMMITTEE ON COURSES

March 15, 2021

To: Jason Stajich, Chair
Riverside Division

From: Ming Lee Tang, Chair
Committee on Courses

Re: Ad Hoc Committee on the Evaluation of Faculty Teaching Final Report

The Committee on Courses reviewed the Ad Hoc Committee on the Evaluation of Faculty Teaching's Final Report at their March 10, 2021 meeting and is supportive of the report's conclusions.

UC RIVERSIDE Academic Senate

COMMITTEE ON FACULTY WELFARE

April 5, 2021

To: Jason Stajich
Riverside Division Academic Senate

From: Patricia Morton, Chair
Committee on Faculty Welfare

Re: [Campus Review] Report Review: Final Report from the Ad Hoc Committee on Evaluation of Teaching

The Committee on Faculty Welfare met on March 16, 2020 to consider the Final Report from the Ad Hoc Committee on Evaluation of Teaching.

The report has been long in coming and it is a good evaluation of the diversity of opinions on the value and use of teaching evaluations, both for improving teaching and for evaluating teaching in the merit and promotion review process. Survey responses suggest a range of faculty responses that varied from dropping evaluations altogether to leaving the system as is, but with the majority suggesting that change is necessary. The ad hoc committee proposed a two phased approach to initiating change.

The first phase is immediate, and an attempt to provide a focus on the importance of recognizing and avoiding bias in teaching (=learning) evaluations. This will be addressed by an ieval preamble and the replacement of question 5 with a question on the recognition of bias. These two approaches form a positive educational tool for all of our students. The other aspects focus on evaluation, which include stopping of comparative departmental and campus metrics, communication at all levels of the importance of recognizing bias, and also modifying efile to allow for a diversity of teaching evaluation assessments to be added and utilized. The FW committee was strongly in favor of all of these approaches.

The second phase entails a redesign of the current student evaluations, with a greater focus on questions that greater address learning outcomes and try to avoid bias, and to develop a "Student Evaluation of Learning". The ad hoc committee recommended that this be outsourced to faculty that have research programs that have relevance to the field of bias and learning outcomes. FW supports this approach, but while this is laudable, the same difficulties may be faced with the faculty accepting those recommendations, and also, making sure that questions are appropriate to all fields of study across the campus. However, whatever recommendations that are made must

come to the senate for acceptance, and hopefully a new and improved survey can be developed that is agreed to by the majority.

Increased student participation is necessary but difficult. This was recognized by the committee. Early grade release is seen across many different universities as a means of increasing participation, but FW agrees that without the option in banner we will likely continue to see low participation. Several CFW members felt that incentives to foster greater participation such as early grade release have led in the past to evaluations that were done simply for grade access and were not necessarily meaningful evaluations.

One aspect not covered in the report is whether students should also have access to evaluations as is done at UCSD.

FW noted that in teaching evaluations, (1) there should be greater recognition by departments of the ability to provide anonymous student letters to address teaching at any level of assessment, and (2) that there was not enough focus in the report on the need for greater guidance on evaluation by the Academic Personnel Office.

Overall, FW felt that this is a good first step to revising the evaluation of learning on campus.

UC RIVERSIDE Academic Senate

GRADUATE COUNCIL

March 26, 2021

To: Jason Stajich, Chair
Riverside Division

From: Amanda Lucia, Chair
Graduate Council

Re: [Campus Review] Report Review: Final Report from the Ad Hoc Committee on Evaluation of Teaching

Graduate Council reviewed the final report from the Ad Hoc Committee on Evaluation of Teaching at their March 18, 2021 meeting. The Council was supportive of the report's recommendations and suggests that the report be updated with a timeline for the implementation of recommendations.

GSOE Response to Final Report of the Ad Hoc Committee on Evaluation of Faculty Teaching

- We thought it was a strength that the report makes specific recommendations on which questions to exclude and which to include
- A concern is that the bias in student evaluations does not completely go away; At the same time, we like that students have a voice in this process. We don't have a full solution to this.
- Overall, we worry that evaluation metrics are not a good way to improve teaching. Better teaching would result from building school/dept wide commitment to and interest in teaching, discussing it, etc. We need to promote that a culture concerned about teaching and accompany it with support.

School of Medicine
Division of Biomedical Sciences
Riverside, CA, 92521

April 5, 2021

To: Jason Stajich, Ph.D., Chair, Academic Senate, UCR Division

From: Declan McCole, Ph.D., Chair, Faculty Executive Committee, UCR School of Medicine

Subject: SOM FEC Response to: Final Report from the Ad Hoc Committee on Evaluation of Teaching

Dear Jason,

The SOM FEC welcome the thorough efforts made by the committee to re-imagine the model for teaching evaluations. We acknowledge that student feedback can be of value and is important to empower students to become engaged in improving the teaching mission of the university. However, concerns still remain regarding student biases exerting a disproportionate negative impact on evaluations of faculty teaching, and the lack of clarity on how other modes of evaluation will be structured, collated and eventually weighted during the merit & promotion decision process.

- Concerns include Page 7, Q5 with the 1 (minimal), 3 (moderate) 5 (strong) scale. At the end of blocks when students are evaluated, they tend to be quite exhausted and frustrated/angry. This may prompt a lot of '1' scores in an evaluation based on such negative feelings and thus will have a disproportionately negative impact on faculty who include rigorous questions to evaluate topic understanding on a final exam. FEC suggests that if a student selects such a negative score, then they should have to include a clear justification in a comment box for such a score so as to provide context. This can then be evaluated for merit i.e. a negative score with a comment of "The instructor put hard questions on the exam", could be considered as lacking merit.
- The suggested guidance in providing comments to extract suggestions for improvement are worthwhile and could provide constructive feedback to the instructor.
- FEC advocates that student participation in the new SELs be made compulsory as the response rate is very poor and just favors negative comments. One suggestion that has been enacted elsewhere i.e. University of Arizona, is withholding of a student's grade until they submit the evaluation so as to ensure a meaningful number of respondents. There was some additional discussion as to whether this could be enacted within a School/College or if it requires Senate Executive approval.
- The preamble regarding bias should also include a reminder of professionalism in the tone of their responses, particularly to students enrolled in professional courses i.e. medicine, business, engineering; and to focus on constructive comments.
- The report acknowledges that the "other" methods of evaluations are likely not going to be weighed as heavily as SETs, so even though faculty will be allowed to upload more evidence of their teaching, will committees weigh that evidence appropriately, or will they continue to rely on the data that they are familiar with – i.e. SETs?

- The reframing of the questions doesn't solve the underlying problem. No matter how you frame the question, students will upvote the things they like, not the things that are good teaching practice. For example, the question *What approaches or materials did you find most effective or successful in aiding your learning in this course?* The problem is students ***are not qualified*** to answer that question. They are not knowledgeable on the scholarship of teaching and learning. They will generally respond to this prompt by evaluating negatively the methods they disliked – regardless of how effective those methods were in helping them learn. Ask any group of students how many of them want more quizzes. Not a single hand will go up. Yet, we know that quizzing is one of the most impactful strategies for learning. With SETs you are dealing with student *preferences* not with what is most beneficial to their learning.
- Who will code all these qualitative answers generated by the revised SET/SEL to come up with some kind of useful metric?
- There are also concerns regarding what the "other" means of providing evidence of teaching effectiveness will involve as this is not specified. What kind of evidence is going to be supplied? Who is going to provide that evidence? If this involves teaching observation, then who will perform the observation and what metrics will be used?

Yours sincerely,

Declan F. McCole, Ph.D.
Chair, Faculty Executive Committee School of Medicine

UC RIVERSIDE
School of Public Policy

School of Public Policy
University of California, Riverside
INTS 4133 | 900 University Ave
Riverside, CA 92521

TO: Jason Stajich, Chair
Riverside Division

FR: Richard M. Carpiano, Chair
Executive Committee, School of Public Policy

RE: [Campus Review] Report Review: Final Report from the Ad Hoc Committee on Evaluation of Teaching

Date: March 29, 2021

The Executive Committee of the School of Public Policy reviewed and discussed the "Final Report from the Ad Hoc Committee on Evaluation of Teaching." Our committee was in general agreement that the report is thorough and we concur with most of its recommendations. Notably, we like the critique offered regarding the iEval system. It would be good to see it reformed.

That said, the report does not seem to discuss the low response rate by students. A few years ago—perhaps through 2017/2018—students could receive their grades earlier if they completed their course evaluations. This created a very strong incentive for students to complete evaluations and response rates were very high. However, they changed the grading system and removed this incentive. Response rates plummeted dramatically and have been very low ever since. Currently, we routinely see faculty with approximately 10% response rates on evaluations. It could be argued that low response rates bias evaluation scores downwards and magnify unfair/overly negative evaluations. Hence, the committee needs to find a way to boost response rates to evaluations. As long as we are going to use evaluations, this is a key issue that needs to be addressed. Response rates are likely even worse during this pandemic with the online teaching experience. It will be great if the committee can think of ways to incentivize students to submit their evaluations.

Sincerely,

Richard M. Carpiano, Ph.D., M.P.H.
Professor of Public Policy and Sociology