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

1. Introduction

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

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

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

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

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

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

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

2. Assessment versus Evaluation

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

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

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

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

Evaluation:

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

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

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

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

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

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

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

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

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

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

3. Student Feedback

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

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

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

---

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

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

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

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

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

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

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

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

---

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

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

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

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

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

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

The bank is divided into four categories:

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

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

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

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

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

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

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

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

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

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

Recommendations:

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

4. Holistic Evaluation

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

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

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

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

4.1. The teaching Portfolio

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

A. Teaching and Supportive Learning

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

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

B. Professional Learning and Development

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

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

C. Mentorship

Mentorship comes in many forms, including:

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

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

D. Research, Scholarship and Inquiry

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

E. Educational Leadership

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

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

4.2. Self-Reflection

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

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

A. Teaching and Supportive Learning

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

---

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

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

B. Professional Learning and Development

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

C. Mentorship

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

D. Research, Scholarship and Inquiry

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

E. Educational Leadership

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

4.3. The role of the department chair

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

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

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

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

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

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

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

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

5. Guidance and oversight – next steps

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

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

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

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

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

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

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

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

Recommendations:

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

References

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

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

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

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

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

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

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

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

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

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

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

Membership

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

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

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

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

- Teaching evaluation on our campus presently appears to rely primarily on the Course and Professor Evaluations (CAPE) scores/comments. The CAPE instrument was developed by students for use by students, is not a research-validated instrument like the Student Assessment of their Learning Gains (SALG), and so is not a satisfactory tool.

- Although the APM requires that at least two methods be used to evaluate teaching in personnel reviews, this is not uniformly observed or undertaken in sufficient depth to be informative to CAP reviewers.

- Individuals appointed in the Security of Employment/Teaching Professor are primarily evaluated on the basis of their teaching performance, and the number of individuals employed in this series is increasing. We need to ensure that our campus employs an appropriate suite of methods for evaluating their work.

- Proposed changes to APM 285 related to the Security of Employment/Teaching Professor rank/step, would, if implemented, require more robust ways of evaluating teaching and learning for the 2018-19 evaluation cycle.

- The WASC reaccreditation visit in 2019 will expect us to demonstrate that we are assessing teaching, learning, and program impact in compelling ways.

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

- Identify and make available multiple reliable existing tools for teaching evaluation and also provide training on how to use and interpret them. It is conceivable that different units may deploy different combinations of tools, depending on what is appropriate to the discipline, curriculum, teaching method, etc.

- Establish a campus culture where both formative (critique to help one improve one's teaching on an ongoing basis) and summative (evaluative) assessment of teaching and learning are standard practice. This would, for instance, encourage faculty statements

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

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

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

- Examine the literature and consider best practices from peer institutions;

- Review current UC San Diego resources and practices; explore the expertise within the Teaching & Learning Commons, Rady School of Management and Jacobs School of Engineering;

- Determine what new resources, tools, and/or training programs are needed and how to coordinate them with each other and with what already exists;

- Review incentives (e.g., use of teaching in merit reviews, campus awards, external grant proposals) and see what others might be valuable, and whether existing incentives are aligned with our goals;

- Consider cultural factors: which units and/or faculty groups are already leaders in this area? Which can help others make progress? Which are fearful or resistant to discussing these issues?

- Consider a communications strategy: How can we best make case that change benefits students, faculty and the institution as a whole?

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

MEMBERSHIP

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

Senate

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

TIMELINE

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

Attachments: List of Holistic Teaching Evaluation Resources

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

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

Appendix B - Student Evaluation Question Bank

1. Student Participation

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

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

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

3. Student Learning

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

4. Structure/Inclusion

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

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

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

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

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

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

5.Open-ended prompts for student comments

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

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

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

6. Instructional assistant evaluation

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

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

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

The instructional assistant communicated effectively in section.

The instructional assistant communicated effectively in office hours.

Open-ended prompts to elicit comments:

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

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

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

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

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

- Remember that you are writing to your instructor. Your feedback can valuably influence the ways they teach this course and others in the future. (Unlike an online review site like "Rate My Professor," this is not a forum for saying whether or not you recommend a course to other students.)

- Specific constructive suggestions that focus on your learning are far more useful than general praise or critiques. See below for examples of ways you can provide feedback that helps instructors understand how their instructional choices facilitated or hindered your learning. Both positive and negative feedback is most helpful when very specific.

- Comments that are not related to your learning diminish the value of your feedback. For example, it is not helpful to comment upon an instructor's appearance or to include personal insults in your feedback.

Some examples of constructive feedback:

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

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

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