Adv Physiol Educ 44: 26–31, 2020;
doi:10.1152/advan.00088.2019.

ILLUMINATIONS

An evidence-based framework for peer review of teaching

Sierra M. Dawson and Austin D. Hocker
Department of Human Physiology, University of Oregon, Eugene, Oregon

Submitted 5 July 2019; accepted in final form 8 November 2019

INTRODUCTION

Teaching evaluation is currently under scrutiny due to the overreliance on student evaluations of teaching (2, 6, 10). Teaching evaluation must employ a more holistic approach, including the student experience, faculty reflection, and peer review (5). Peer review is an expected practice at many institutions, but rarely includes a formal system, and often falls short of its potential. In order for peer review of teaching to be an effective component of teaching evaluation, a systematic approach is required, ideally grounded in a shared definition of teaching excellence. Peer reviews are often unstructured or completed just in time for tenure and promotion dossiers. Without criteria for evaluation and a definition of teaching excellence, unstructured reviews can often take the form of "love letters," espousing the virtues of a colleague's intellectual abilities and knowledge of the discipline, while void of insights regarding the use of evidence-based practices or student learning. Done well, peer review has the potential to open up conversations about teaching practices or philosophies of learning and assist the faculty with continuous and incremental improvement of their teaching, but only with the appropriate infrastructure as guide rails to the process.

Our departmental process for peer review was unclear to new faculty, and the department agreed to revise and implement a new system for peer reviews. A single faculty member (S.M.D.) drafted the first version of the new peer review process, which was first piloted in 2015, with agreement from the department. The goals were to implement a peer review system that promoted formative teaching reviews, engaged faculty in deeper conversations about evidence-based teaching practices, and also produced a useful report to be used during formal evaluation by the department chair and personnel committees for promotion, tenure, and merit.

An Evidence-Based Peer Review System

The following is a description of how our department redesigned peer review of teaching to make the best use of faculty time, while producing evidence-based observations and recommendations for continual improvement and a report that is used during formal teaching evaluation. Our department currently includes 17 tenure-related or career faculty, approximately evenly split among assistant, associate, full professors, and career instructors/lecturers, as well as 2–4 pro tempore (adjunct) instructors each year. Our university courses are taught in 11-wk terms (fall, winter, spring), with professors typically teaching 2.5 courses per year. Career instructors and lecturers are fully participating faculty in the department who teach the equivalent of six courses per year and contribute to department and university service alongside their tenure-track peers. Our department benefits from the teaching expertise of career instructors/lecturers, who often facilitate discussions on teaching throughout our department and provide peer support about evidence-based teaching practices to tenure-related faculty. Undergraduate class sizes in the department currently range from ~32 to 250.

The peer review process developed includes the following elements:

1. Consistent use of one observation tool
2. Self-reflection of evidence-based teaching practice utilization
3. Guided in-person discussion following observation
4. Template for peer review report
5. Training of peer reviewers
6. Systematic assignment of peer review pairs

The new peer review system described below has been in place since 2015, has been positively embraced by junior and senior faculty alike, and includes both career instructional faculty and tenure-related faculty as peer reviewers.

Consistent use of one observation tool. Without guidelines for peer review processes, reviewers will be guided by personal beliefs about good teaching, regardless of whether or not those beliefs are based on the scholarship of teaching and learning. Therefore, course observations should be structured and standardized to collect evidence about teaching practices, ideally in alignment with the definition of teaching excellence on which the unit has agreed. Thankfully, research on effective teaching practices has contributed a number of observation protocols for measuring teaching practices and student activities. For example RTOP (7, 8), COPUS (9), and PORTAAL (4) each have strengths and weaknesses in their ease of implementation and the data they provide. We chose the Classroom Observation Protocol for Undergraduate STEM (COPUS) from the Carl Wieman Science Education Initiative (http://www.cwsei.ubc.ca/), which measures evidence-based teaching practices and the amount of time that students are engaged in activities beneficial to their learning. The COPUS was well-aligned with the unit's definition of teaching excellence, which focused on evidence-based and student-centered teaching. The COPUS was also relatively easy to implement, a protocol for efficiently training observers existed, and the results provided considerable detail regarding what was happening in the classroom.

To conduct the COPUS, the observer codes the instructor and student activities at 2-min intervals (9). Examples for

Address for reprint requests and other correspondence: S. M. Dawson, 1258 University of Oregon, Eugene OR, 97403 (e-mail: sdawson@uoregon.edu).

1043-4046/20 Copyright © 2020 The American Physiological Society

Downloaded from journals.physiology.org/journal/advances (193.255.177.234) on April 2, 2020.

student codes include listening, asking a question, whole class discussion, individual thinking/problem solving, working in groups, making a prediction, presenting to the class, waiting, taking a quiz, and others. Examples for instructor codes include lecturing, real-time writing, showing a simulation/demonstration, giving feedback, posing a nonrhetorical question, answering student questions, moving around the class guiding student work, administration, and others. We chose to employ the Microsoft Excel (2010 Microsoft Corp.) version of the COPUS, which we further augmented to automatically produce visual representations of what the students and instructor are doing during the observation (https://tep.uoregon.edu/files/copus_with_visualization.xlsx). The results can be viewed with all the codes represented for a more detailed analysis (Fig. 1), or collapsed into instructors "guiding" or "presenting" and students "working" or "receiving" for a bigger picture understanding (Fig. 2). Both visuals are useful during in-person discussions by allowing for visualization of trends in the incorporation of active learning, as well as a deeper analysis of specific teaching practices, which may be worth discussing. In addition to coding student and instructor behaviors every 2 min, the observer can also make notes in the right-hand column of the COPUS about what else they are observing (topics being discussed, student engagement, level of challenge, classroom sense of community and inclusiveness, or other nuances), which can be discussed or clarified at the follow-up meeting.

Self-reflection of evidence-based teaching practice utilization. Holistic peer review benefits from the perspective of the faculty instructor to provide course context and to better understand the instructor's goals for the course. Furthermore, it promotes instructor reflection about teaching practices and student outcomes and increases the agency of the faculty being reviewed. The Department of Human Physiology chose to include another of the Carl Wieman Science Education Initiative tools: the Teaching Practices Inventory (TPI) (11). This inventory of teaching practices is completed as an online survey, and the output is a score of the extent to which evidence-based teaching practices are utilized (11). Each teaching practice selected is given 0–3 points based on the strength of the evidence that the specific teaching practice improves student learning. Not every evidence-based teaching practice is practical in every context, so the goal is not to get the highest possible TPI score. However, the range within which the scores falls sheds light on the degree to which evidence-based practices are utilized.

The benefit of including the TPI into the peer review process is that it provides lists of evidence-based practices that are and are not currently incorporated into the course. It is common for

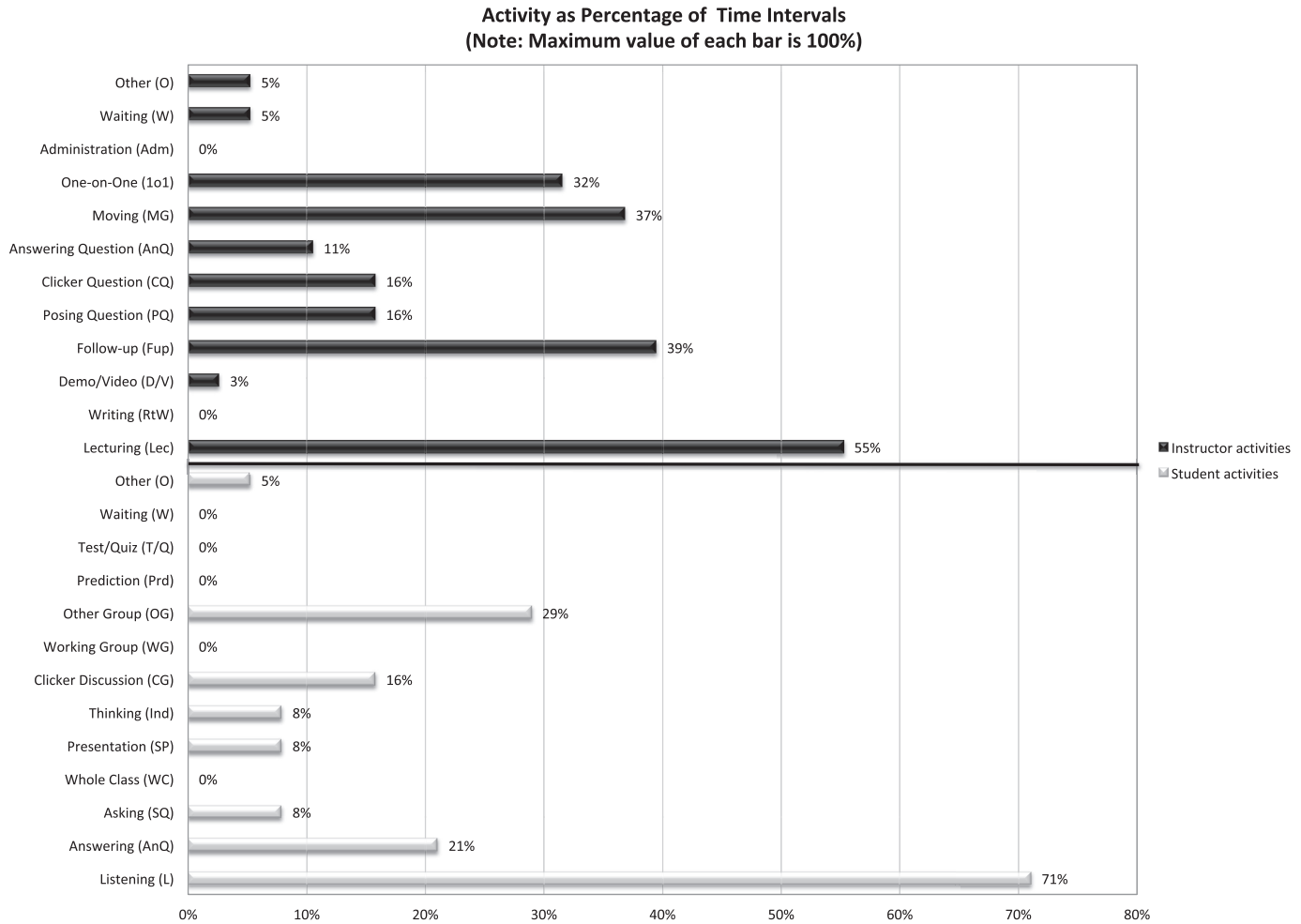


Fig. 1. Output from the Classroom Observation Protocol for Undergraduate STEM (COPUS) from one class period observation. The codes on the *top* indicate what the instructor is doing, and codes on the *bottom* indicate what the students are doing, during each 2-min time interval.

Downloaded from journals.physiology.org/journal/advances (193.255.177.234) on April 2, 2020.

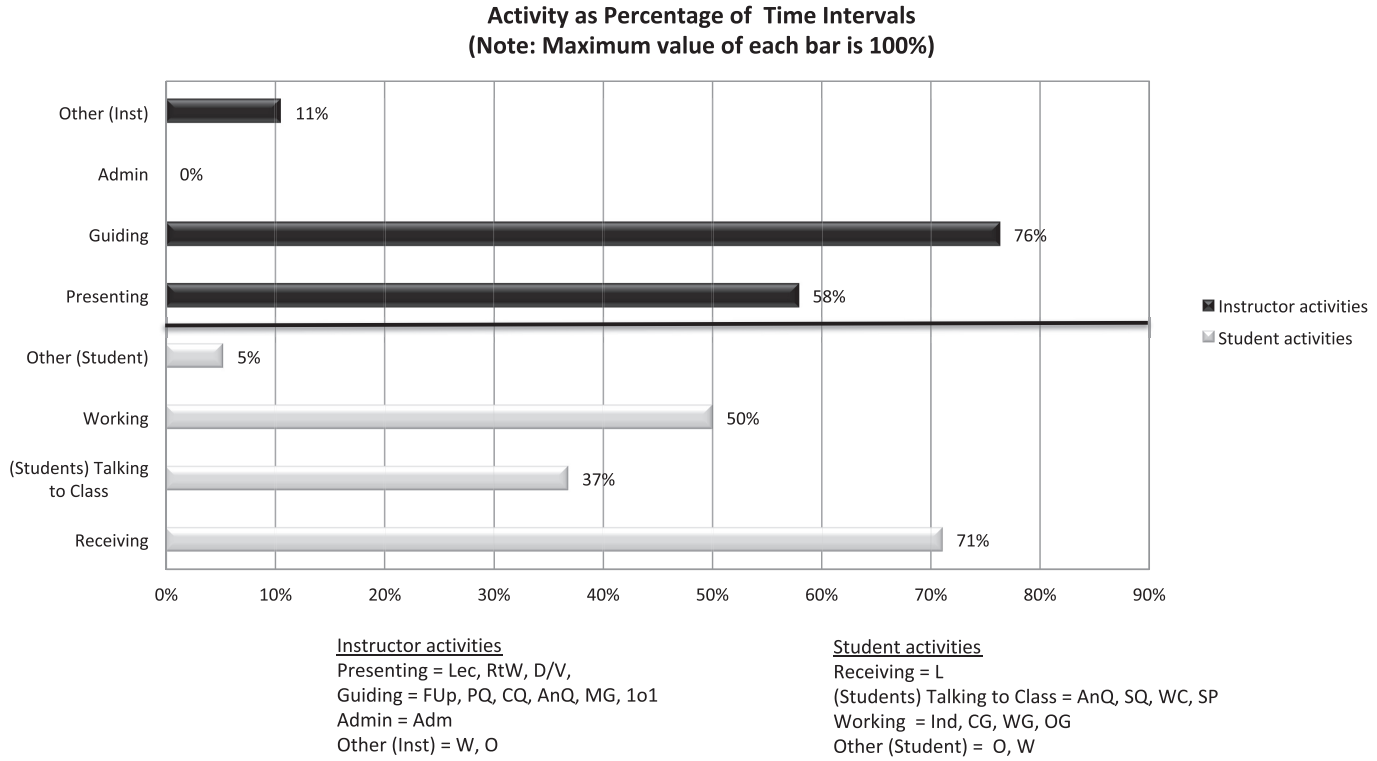


Fig. 2. Output from the Classroom Observation Protocol for Undergraduate STEM (COPUS) from one class period observation, with the individual codes collapsed to indicate the degree to which the instructor was guiding vs. presenting (*top*), and the students were working vs. talking vs. receiving (*bottom*), during each 2-min time interval. See Fig. 1 for definition of code acronyms.

the reviewer to identify one or two evidence-based practices that are not yet utilized and recommend those be included in future iterations of the course. A preliminary analysis of TPI scores from instructors who have been peer reviewed twice for the same course suggests instructors do implement more evidence-based teaching practices between their first and next peer review cycle (Fig. 3). As an example, by using the list of evidence-based practices as a guide, a new assistant professor with little previous teaching experience moved from a TPI score of 35 to 53 from her first peer review to her next peer review of the same course 1 yr later (department mean TPI score is currently 41.7 ± 6.7; Fig. 3). These types of improvements suggest the TPI is a practical tool to help guide faculty toward evidence-based course changes. Importantly, there are limitations to self-reflection tools (3), and we emphasize that the intention of TPI in our peer-review system is the formative reflection on current teaching practices and discussion of strategies to include more evidence-based teaching practices from the inventory, when applicable.

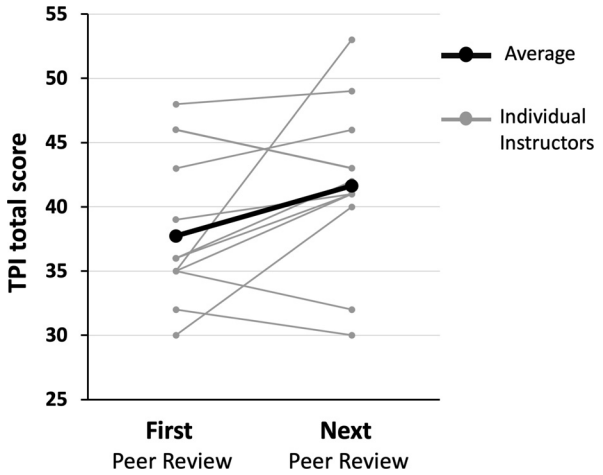


Fig. 3. Most faculty Teaching Practices Inventory (TPI) scores improve in a single peer-review cycle. Self-reported TPI scores from different iterations of the same course taught by the same instructor at least 1 yr apart were compared. Eight of eleven faculty increased their TPI score, suggesting increased implementation of evidence-based teaching practices by instructors.

Guided in-person discussion following observation. One of the most enjoyable parts of the new peer review system is the requirement for a follow-up conversation. Teaching can be such a private endeavor on a research-focused campus, and the follow-up discussion breaks this pattern and allows peers to discuss their teaching practices.

The two main goals for the meeting are to *1*) discuss the data collected by the COPUS and TPI, and *2*) allow the faculty under review to answer the six standard questions, which were adapted from Bain's book, *What the Best College Teachers Do* (1).

The questions are as follows:

1. How do people learn?
2. What will your students be able to do as a result of learning in your class?
3. What do you do to help students achieve these goals (described in no. 2)?
4. How do you and the students get feedback about student learning along the way?
5. How do you evaluate your efforts to foster student learning?

Downloaded from journals.physiology.org/journal/advances (193.255.177.234) on April 2, 2020.

6. What can I and/or the department do to further support your efforts?

While analyzing the COPUS data together, the conversation usually focuses on whether or not the documented use of class time matches the faculty member's expectations and intentions. The discussion might also include suggestions, such as breaking up long periods of direct lecture into smaller chunks with student-centered activities between, or remembering to move around the classroom during student-centered activities, or adding the opportunity for students to predict the outcome before showing a simulation or demonstration.

The output from the TPI indicates specifically which evidence-based practices are included and which are not, along with a numerical representation of the extent to which evidence-based practices are utilized overall. A peer reviewer can use the list of unimplemented practices to make suggestions, such as providing time for students to discuss their answers with one another after they individually answer a clicker question, or starting a class period with why the topic is useful or interesting from the student's perspective, or the addition of a short reflective exercise at the end of class.

The focus of the COPUS and TPI discussion is often about what small changes can be made to increase the utilization of teaching practices that have been documented to improve student learning. The six guiding questions are asked to dig deeper into the faculty member's teaching philosophy and the reasons he/she is making particular choices. In our experience, these conversations have uncovered mismatches in faculty expectations and student outcomes and most importantly have helped instructors realize they could design a new project or learning activity to better guide student learning in their course. Through open dialog with peers about our teaching goals and practices, our blind spots as educators can sometimes be revealed.

Template for peer review report. Before implementation of this new system of peer review, writing the report was one of the hardest parts of the informal process. What specifically should be included? For most peer reviewers, it felt like a creative writing assignment. Therefore, our new system includes a template for the report with three sections, which has significantly reduced the burden of the report. Additionally, the template is useful for transitioning the formative discussions of teaching improvement to a useful report for formal evaluation of teaching. Our report template met the guidelines and expectations from the university's tenure and promotion process and has been commended for its clarity and usefulness during formal evaluation of teaching.

Section 1 is an overview and includes context for the review, such as the course name and number, the topics under discussion the day of the observation, whether the students were majors or nonmajors, their class standing, and whether it was an elective or required course.

Section 2 provides the results of the data collected via the COPUS, TPI, and the in-person discussion [i.e.: the answers to the six questions adapted from Bain (1)]. The template includes a brief explanation of each tool, as context for readers outside the unit (e.g., university personnel committees or teaching award committees). The unit's mean and standard deviation for the TPI are listed to provide context for the TPI score reported. The template includes an appendix, which lists all of the evidence-based practices from the TPI, so the reviewer can simply electronically cut and paste the practices not utilized into a separate document, which can also be used as fodder for evidence-based recommendations to improve the course. A second appendix includes a placeholder for the COPUS graphs, illustrating the student and instructor activities as a percentage of time intervals, or as a percentage of all activities or activities across time. The peer reviewer is reminded in the template to include a caption for each graph that explains its meaning to a reader outside the unit.

Section 3 provides recommendations to the individual being reviewed that will continue to increase the use of evidence-based teaching methods and continue to create the best learning environment for the students. A note on the template reminds reviewers to go back through the list of evidence-based practices not yet utilized and consider suggesting two or three that make sense in the context of the course and learning goals.

At the end of the template, there is a reference list that includes the three primary sources on which the peer review process was built (1, 9, 11). This reinforces the fact that, not only do we care about implementing evidence-based practices in our classrooms, but that we also have an evidence-based process for the peer review of teaching.

Training of peer reviewers. In September 2015, we launched our new peer review of teaching system with a 2-h training for all faculty and graduate students in the unit. This training was conducted as a part of our unit's annual day of teaching development we call the Teaching Academy, in which all faculty and graduate students are required to participate. We trained all faculty so that both peer reviewers and those being reviewed would understand the evidence-based tools used (COPUS and TPI) and their relationship to evidence-based teaching practices. Additionally, as faculty rotate on and off the peer review committee, we wanted all faculty to understand the process and be prepared to act as peer reviewers.

The learning objectives for the training included the following:

1. Describe the three new components to the Peer Review of Teaching.
2. Conduct the COPUS accurately and confidently.
3. Interpret the data collected via the TPI.
4. Develop confidence in your ability to complete the new Peer Review of Teaching written report, including evidence-based recommendations.

The COPUS portion of the training included practice coding class sessions that had been videotaped, including Eric Mazur from Harvard and Paula Lemons from the University of Georgia (videos and training material generously shared by Michelle Smith currently at Cornell University). Participants coded for 2-min intervals and then discussed the codes applied and the rationale for their selection. Although coding a video can be more challenging than being physically present in the class, due to the limited lens of the camera, it is still an effective method for training and onboarding.

The TPI portion of the training included participants completing the TPI in hard copy and then swapping with a neighbor and scoring the responses using the provided rubric (11). They then took the list of evidence-based teaching practices in the peer review report template and placed an asterisk next to each one included. Lastly, the pair took turns sharing the results of the TPI (score and practices utilized) and discussed one or two of the evidence-based practices not yet included in the course

Downloaded from journals.physiology.org/journal/advances (193.255.177.234) on April 2, 2020.

and whether or not they might include it in the future, assuming the practice made sense in the context of the course.

Each year since the commencement of the new system in 2015, there has been a 2-h training for the faculty who are serving on the peer review committee. The meeting in 2016 was an opportunity to discuss how things had gone the past year and any suggested changes to the template or clarifications to the process. As new members have rotated on the committee in 2017 and 2018, the 2-h training was more similar to the original version in which new and experienced committee members learn the coding procedures for COPUS and practice using the classroom video recordings as a norming exercise.

Systematic assignment of peer review pairs. An extremely important part of designing a new peer review of teaching process for your unit is implementing a system to assign reviewers to faculty under review. It does not matter if you have an evidence-based observation tool or a helpful report template if the peer review never occurs. In speaking to units across campus about peer review, it is surprisingly common that the peer review process is completely forgotten until a dossier is being assembled for a third- or sixth-year review process for promotion or tenure. For example, in one case at our institution, the required three peer reviews for a 6-yr time period were all conducted within 3 wk on the same course. Do not underestimate the importance of maintaining a system for coordinating peer reviews.

The system we implemented followed the service model of our unit and ensured that all needed reviews could be completed without overloading anyone. In our department of 17 tenure-related or career faculty plus 2–4 pro tempore (adjunct) faculty, 5 faculty agreed to serve on the Peer Review Committee. Each individual review takes <6 h (1- to 2-h class observation, 1-h follow-up discussion, and 1–2 h to complete the report template). In the first 3 yr of the new system, there were 16 reviews conducted in 2015–16 when some catching up was needed (S.M.D. conducted more than one per term), followed by 12 reviews in 2016–17 and 13 reviews in 2017–18. The reviews were distributed equally across three terms so that the five peer reviewers would conduct no more than one review each term. The chair of the committee organizes and directs the peer review schedule and assigns peer review pairs (a reviewer and reviewee). The chair is responsible for tracking all faculty and their rank, which determines the frequency of required peer reviews at our institution. At the beginning of each academic year, the chair identifies faculty in need of peer review and distributes the work to committee members. It is helpful to map out required peer reviews over a 3-yr period to distribute the needed peer reviews over time, since some faculty ranks require a review every year, and others only once in a 3-yr period. The peer review mapping is shared at the first faculty meeting of the year to prevent scheduling conflicts and address any concerns. Then the chair of the peer review committee notifies the peer review pairs at the beginning of each term, including a reminder that the person being reviewed needs to initiate the process by sending their reviewer days and times that would be best for the class observation and the follow-up meeting. The peer review report has a due date, is submitted to the committee chair, the office manager, and the department chair, and is included in the reviewed faculty member's personnel file.

Future Directions and Next Steps

Our unit's definition of teaching excellence was recently updated to a more formal statement of values. It now reads:

> We value engaging, inclusive and evidence-based teaching. Teaching is engaging when students participate in active learning, inclusive when students belong, are represented and have a voice, and evidence-based when course design and classroom activities are grounded in the scholarship of teaching and learning.

The peer review system currently documents the degree to which the learning environment is engaging (or uses active learning) and is evidence based, but does not collect information about the faculty's intentional actions to make the class inclusive. This demonstrates the need to continually update any process, like peer review, to keep up with the changing focus of the unit. In this particular case, the unit is considering an adjustment to the follow-up discussion questions to ask how the faculty is intentionally ensuring that students belong, are represented, and have a voice in their course. Ideally, the peer review process should be aligned to the unit's teaching excellence definition in order for it to provide the information needed for continual improvement and formal evaluation of teaching.

Our unit has also just recently added its first online course, and another is being prepared. Since the current peer review system expects a classroom observation using COPUS, there needs to be an alternative provided to "observe" the online learning environment in a way that documents student engagement and the degree to which evidence-based practices are utilized. The Quality Matters Peer Course Review Rubric (5) is a tool designed specifically for online courses and may be the best tool to replace the COPUS observation tool for online courses.

As the make-up of the unit changes over time, the number of faculty serving on the peer review committee will have to change too. On our campus, junior faculty (prepromotion or pretenure) require peer review every year, whereas tenured faculty require peer review every other year, and promoted career instructors and lecturers require peer review every 3 yr. As our faculty become more senior, fewer members will be required for the peer review committee to meet the frequency of the required reviews.

Summary

Peer review of teaching has the potential to be an important contributor to the continuous improvement and formal evaluation of teaching. It has the added benefit of opening up dialog between faculty colleagues about what happens in the classroom, a place where faculty often feel isolated from one another. No longer does peer review need to be a rushed exercise in which love letters are produced only to check a box in a tenure or promotion file. Through development of a system that includes a classroom observation tool, self-reflection exercise, structured follow-up discussion, template for the report, and systematic assignment of trained peers, units can ensure that the time spent on peer review leads to a process that supports the teaching culture of the unit and provides meaningful data for formal evaluation.

Downloaded from journals.physiology.org/journal/advances (193.255.177.234) on April 2, 2020.

ACKNOWLEDGMENTS

We would like to thank our colleagues in the Department of Human Physiology at the University of Oregon for participation and engagement in peer review. Thank you to Julie Mueller, who created the Excel sheets to automatically summarize COPUS results.

DISCLOSURES

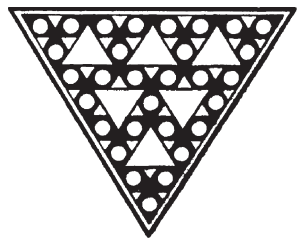
No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

S.M.D. conceived and designed research; A.D.H. analyzed data; S.M.D. and A.D.H. prepared figures; S.M.D. drafted manuscript; S.M.D. and A.D.H. edited and revised manuscript; S.M.D. and A.D.H. approved final version of manuscript.

REFERENCES

1. **Bain K.** *What the Best College Teachers Do* (Online). Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674013254 [22 Sept 2019].
2. **Berk R.** Start spreading the news: use multiple sources of evidence to evaluate teaching. *J Fac Dev* 32: 73–81, 2018.
3. **Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE.** What we say is not what we do: effective evaluation of faculty professional development programs. *Bioscience* 61: 550–558, 2011. doi:10.1525/bio.2011.61.7.9.
4. **Eddy SL, Converse M, Wenderoth MP.** PORTAAL: a classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE Life Sci Educ* 14: ar23, 2015. doi:10.1187/cbe.14-06-0095.
5. **Little BB.** The use of standards for peer review of online nursing courses: a pilot study. *J Nurs Educ* 48: 411–415, 2009. doi:10.3928/01484834-20090615-10.
6. **Mengel F, Sauermann J, Zölitz U.** Gender bias in teaching evaluations. *J Eur Econ Assoc* 17: 535–566, 2019. doi:10.1093/jeea/jvx057.
7. **Piburn MD.** *Reformed Teaching Observation Protocol (RTOP) Reference Manual. Technical Report* (Online). https://eric.ed.gov/?id=ED447205 [22 Sept 2019].
8. **Sawada D, Piburn MD, Judson E, Turley J, Falconer K, Benford R, Bloom I.** Measuring reform practices in science and mathematics classrooms: the reformed teaching observation protocol. *Sch Sci Math* 102: 245–253, 2002. doi:10.1111/j.1949-8594.2002.tb17883.x.
9. **Smith MK, Jones FHM, Gilbert SL, Wieman CE.** The classroom observation protocol for undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE Life Sci Educ* 12: 618–627, 2013. doi:10.1187/cbe.13-08-0154.
10. **Stark P, Ottoboni K, Boring A.** Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Res* 2016: 1–11, 2016. doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.
11. **Wieman C, Gilbert S.** The teaching practices inventory: a new tool for characterizing college and university teaching in mathematics and science. *CBE Life Sci Educ* 13: 552–569, 2014. doi:10.1187/cbe.14-02-0023.



Downloaded from journals.physiology.org/journal/advances (193.255.177.234) on April 2, 2020.