

Wearing the consultant's hat: Training faculty to perform quality peer reviews

Eric T. Metzler and Lisa Kurz

Abstract

The claim has been made for decades that college teaching, like research, should be peer reviewed if we are to see it as a serious scholarly activity; yet peer review as a method of evaluating teaching is not widespread. Interest in peer review is increasing, however, as institutions have sought ways of evaluating teaching informed by professional practice instead of solely data gathered from student evaluations of teaching, which are based on consumer satisfaction models and have proved biased against women and people of color. Yet if faculty are to engage in peer review of teaching, whether for formative or summative purposes, they should be trained in best practices to ensure that the reviews are of high quality and, most important, that they focus on improving the observed instructor's teaching. In this article we describe our program to train faculty to be peer reviewers, describe faculty reactions to our training, and outline the lessons we as consultants have learned in facilitating this program.

Keywords: peer review of teaching, faculty training, student evaluations of teaching, measuring teaching effectiveness

Peer Review of Research vs. Peer Review of Teaching

In the tenure and promotion economy in institutions of higher learning, peer review of research has long been the sine qua non of the process by which those on the tenure track, and increasingly, those in untenured positions, achieve career advancement and surety of employment with a college or university. Those conducting such peer review are well versed in their subject area, in the rules of research and publication—both explicit and unspoken—and in the weight various subject area journals carry. Peers performing research reviews are qualified to do so by dint of their subject area expertise and their insider's knowledge of the qualitative dimensions pertinent to review. Quantitatively speaking, review of research is a straightforward matter of counting the number of published works or those accepted for publication, which yields a figure that either supports or does not support the institution's notion of appropriate productivity. These qualitative and quantitative aspects of peer review of research make it a defensible, if not straightforward, process, which senior faculty are well prepared to undertake.

Yet when it comes to evaluating the other major responsibility of faculty—teaching—the waters become murky and the process ill defined. Unlike research review where faculty command expert knowledge of the research they are asked to review, faculty asked to perform peer review of teaching are much more likely to be unschooled in the general pedagogical principles that might serve as a benchmark of teaching effectiveness. At the same time, using their own teaching style—however effective—as the yardstick by which to measure good teaching can be both unfair and inappropriate. Hence, without formal training in peer review of teaching, many faculty members find themselves unsure of *what* to look for and *how* to judge teaching in its performative, preparatory, and conceptual dimensions. Further, unlike research review, the elegance of quantitative considerations plays a minor role, if any, in the teaching context. One may list the total courses, students, and new course preparations undertaken

in one's teaching duties, but these quantitative measures can only indicate an instructor's workload; they say nothing about the quality of the teaching itself.

Peer Review of Teaching Confers Status as a Serious Scholarly Activity

The idea of peer review of teaching arose from Ernest L. Boyer's (1990) argument that teaching, like research, should be considered a scholarly activity, which has since found its manifestation in the Scholarship of Teaching and Learning, or SoTL. Yet it was Pat Hutchings (1994, 1996) and Lee Shulman (1993) who proposed that the intellectual work of teaching must be peer reviewed if it is to be taken seriously as scholarly work. In *Making Teaching Community Property*, Hutchings (1996) based her thoughts and writing on a collaboration with colleagues at a dozen universities around the country exploring different models of peer review, including teaching circles, mentoring relationships, and reciprocal class observations. In this early monograph, she commented on how difficult it is for busy faculty to see and reflect on what they do in the classroom to improve their teaching and their students' learning. Faculty need the help of peers to see more fully what they do, and she argued that improving teaching and learning should be the primary purpose of any peer review, whether its ostensible purpose is formative or summative. Although Hutchings promoted the practice of enriched peer reviews that include looking over syllabi and course documents, she found particular value in observing a class: "while classroom observation focuses on only one facet of teaching—what actually goes on in the classroom—it is also a strategy with special power: (1) to prompt concrete, substantive discussion of teaching and learning, (2) to create an occasion for reflection and self-assessment, and (3) to foster collegiality and community among faculty" (p. 17). This powerful argument for peer review and classroom observation remains as relevant today as when it was published nearly 30 years ago.

In the years following Hutchings's seminal work, however, the peer review of teaching did not spread as widely as might have been hoped. Part of the reason for this failure might lie in the difficulty mentioned above, that faculty find themselves unsure and unconfident when asked to review their peers' teaching. Their hesitation might be ameliorated by formal training in peer review procedures, but the idea of training in how to conduct a peer review is touched on only briefly by Hutchings (1996) and her colleagues as well as in other early work advocating for peer review (e.g., Arreola, 2000; Chism, 2007). Faculty diffidence is further compounded by the difficulty, if not impossibility, of laying out a set of pedagogical standards that can be applied evenly irrespective of discipline and instructor style: teaching is infinitely complex and contingent on so many factors that finding a common yardstick by which to measure the effectiveness of an instructor's teaching in every circumstance is simply not possible. Indeed, an informed, sensitive, and nuanced approach to evaluating both the intellectual work instructors have invested in their teaching and the craft with which they perform in the classroom is essential for high-quality, useful, and fair peer reviews. Yet without training, such an approach is often beyond the capacity of faculty.

The purpose of this article is to lay out the various strands contributing to the need for formal, structured peer review training in today's academy and to share our popular training program, now in its sixth year, with colleagues across the academy, enabling others to adapt our program to their own contexts and needs. The article also considers some of the feedback we requested of our faculty participants, which both refined our training course and revealed some of the ways instructors were transformed by it. Finally, as seasoned educational developers, we share some of our own reflections and learnings from facilitating the course. To begin, let us consider some of the intractable problems associated with student course evaluations, which many institutions use as the primary, if not sole, measure of teaching effectiveness.

The Trouble With SETs

Lack of faculty confidence and training in peer review as well as other structural limitations have resulted in a heavy reliance on Student Evaluations of Teaching (SETs) to stand in for a more comprehensive review of teaching effectiveness such as the review typically performed for research productivity. Proponents of using SET data will claim that the information they provide is generated by those who have the most experience observing an instructor's performance throughout the semester and who rely on effective teaching for their own success—students. They also argue that SETs provide decision-makers with tidy numerical data that can be very alluring as a method of evaluation and comparison precisely *because* they are numerical. These arguments are not without merit, and we must concede that SET data are indeed useful as *one* of the data points by which to measure teaching effectiveness. Yet they alone are insufficient indices of good teaching, for there are many more facets that determine excellence in teaching than how a particular class lands with students. Nevertheless, many, if not most, institutions of higher learning rely on SET data as the sole measure of teaching effectiveness, and this unitary reliance necessarily limits the usefulness of the data. SET data allow for coarse distinctions between unsatisfactory and satisfactory but cannot reveal the many specifics that constitute pedagogical excellence.

The blunt nature of SET data notwithstanding, they have been used to evaluate faculty teaching effectiveness for nearly 100 years (Stroebe, 2020). In recent decades, however, there have been growing concerns about the appropriateness of using SET data exclusively or weighting them heavily when judging faculty teaching effectiveness—especially when the judgment is used for tenure and promotion decisions (Godbout-Kinney & Watson, 2022; Stark & Freishtat, 2014). First, although students can give us an accurate assessment of an instructor's (or course's) clarity, organization, and appeal, they are

not trained in pedagogy and cannot therefore judge the demanding intellectual work of developing and delivering a course; nor are they qualified to evaluate the appropriateness or currency of the course's content. Further, it is quite possible that an instructor may assign challenging or tedious work for the learner's benefit or give difficult but honest feedback to help the learner improve. In either case, learners may react negatively, yielding SET data that suggest ineffective teaching when, in fact, the opposite may be the case (National Academies of Sciences, Engineering, and Medicine, 2020).

Another problem with using SET data as a primary means of evaluating an instructor's teaching is the low rate of response we see as institutions have transitioned from paper forms to electronic surveys. At our large, public Midwestern university we have seen rates of return drop from 70%–80% with paper forms to 30%–45% when moved online. This precipitous drop leads one to wonder whether the data accurately reflect students' assessment of the course and its instruction or whether they represent only students who either loved or hated the course (He & Freeman, 2021).

Adding to the inherent shortcomings of using SETs to evaluate a faculty member's teaching expertise are the seemingly inescapable biases, particularly of gender and race, that result when students offer their subjective evaluations of the instructor. Recent research has established and confirmed that SET data skews positive toward white, male instructors, leaving women and people of color at a disadvantage, even when the teaching quality and effectiveness are at parity (Andersen & Miller, 1997; Buser et al., 2022; Kreitzer & Sweet-Cushman, 2022; Peterson et al., 2019). Similar biases in SETs are found against instructors whose first language is not English, with female international instructors being particularly disadvantaged (Fan et al., 2019). These findings have grown in salience as the professoriate has diversified, making it increasingly important to identify equitable methods by which to evaluate teaching effectiveness and ensure that those who assess teaching are appropriately trained to do so.

Peer Review to Measure Teaching Effectiveness

Fair, equitable, and professionalized evaluation of teaching has only grown in importance as economic realities have led institutions of higher learning to hire teaching faculty who are not eligible for tenure. For these faculty, job retention or promotion rests squarely on how decision-making bodies perceive teaching effectiveness, as typically evidenced by a portfolio of documents. Hence, all the more reason to identify fair and comprehensive ways to evaluate teaching quality that go beyond SETs. Two examples that can make evident an instructor's teaching process are teaching portfolios and course portfolios; both can show thoughtful, reflective teaching and provide explanations of teaching innovations that improve student learning. Yet perhaps the most powerful evidence is well-executed teaching reviews completed by trained faculty peers.

At our large, Midwestern, R1 university, administrative bodies, including the faculty council, have recognized the need to broaden the bases on which a faculty member's teaching effectiveness is evaluated. Accordingly, they have moved to both limit the weight carried by SET data and increase the importance of regular peer review of teaching when evaluating a faculty member's teaching effectiveness (University Faculty Council, 2022). These administrative decisions have resulted in a three-fold uptick in requests for peer reviews of teaching among faculty applying for tenure and promotion. Concomitant with this trend has been an uptick in requests for assistance and training in the peer review process as faculty come to recognize the ultimate import and influence their reviews could have on employment decisions affecting the lives of their peers. This demand in turn has sparked considerable interest in peer review training among educational developers who are often asked to teach faculty how to perform peer reviews and write evaluative letters for tenure or promotion dossiers. At our institution, senior administrators have reported that summative letters have been largely unhelpful. They note the letters are typically full of glowing platitudes about the instructor's teaching but lack specific detail to

support any judgments about their pedagogical skills—good or bad. Hence, in order to ensure the quality, professionalism, and utility of peer review of teaching, it is important to identify best practices and train faculty accordingly.

Standardizing Peer Review With Checklists and Rubrics

One approach to professionalizing peer review of teaching has been to codify best teaching practice into a tabular rubric as has the Bay View Alliance (n.d.) with the TEval (n.d.) form, used to evaluate teaching in STEM disciplines. Here, the reviewer sees a list of pedagogical dimensions (e.g., teaching practices or class climate), each described at varying levels of expertise (e.g., developing, proficient, or expert). Another approach to supporting untrained faculty in performing peer review of teaching is checklists, such as those proposed in Nancy Chism's (2007) seminal work *Peer Review of Teaching: A Sourcebook*. Here, various dimensions of effective teaching appear in a list with checkboxes, offering the peer reviewer a comprehensive set of teaching moves to look for when observing a class in action. An online version of such a checklist, for example, has been created by Carl Wieman and his associates (Smith et al., 2013) for STEM courses.

Whether a peer reviewer uses checklists or rubrics to guide their observation and evaluation of teaching, we observe some clear strengths and advantages to highly systematic and formal approaches such as these. First, standardized forms such as rubrics and checklists are developed by skilled and experienced instructors, drawing on the wisdom of many to capture best general pedagogical practices. Second, the forms generate quantified data that can be compared across instructors. Third, the specific articulations of various teaching dimensions can help the novice or untrained reviewer determine what to attend to when reviewing teaching documents or observing a class, both of which can be overwhelming to an untrained reviewer without guidance. Hence, using standard rubrics and checklists to perform

peer review of teaching does much to yield fair, appropriate, and useful data about an observed class session at a baseline level.

At the same time, the formulaic, one-size-fits-all structure of rubrics and checklists necessarily misses the rich singularity of an instructor's craft on many fronts, including the intellectual work of designing a course, the instructor's performance when interacting with students, grading and feedback strategies, and other dimensions of the infinitely complex undertaking of teaching. Further, standardized instruments are often lengthy, complex, and packed with information, making it difficult to train one's attention fully on the live teaching dynamic while also attending to the checklist or rubric. Standardized instruments also articulate a set number of teaching dimensions, purportedly common to university teaching; yet such standard articulations may not necessarily apply to the discipline, particular class, or signature pedagogy unique to a discipline or instructor precisely because they are standardized and calibrated to generalized notions of teaching. Hence, although we see both the benefits and the necessity of rubrics and checkboxes for faculty untrained in peer review, we believe intentionally trained faculty are best positioned to see both generally accepted best teaching practice and unique teaching choices that set instructors apart. In fact, peer review (especially when performed by a peer outside an instructor's discipline) might be seen as a way of helping an instructor define and hone their discipline-specific signature pedagogy. Shulman's (2005) writing on signature pedagogies provides a basis for this line of thinking.

Training Faculty to Perform Peer Reviews as Instructional Consultants

To meet institutional demand for trained faculty peer reviewers, we developed a year-long course designed to teach faculty how to perform peer reviews as would instructional consultants. The training course includes instruction on best practices of peer review followed by a

practicum that gives participants the opportunity to practice performing reviews in a safe, controlled environment. Faculty participate in the program voluntarily, completing an application to participate, which enables us to communicate the demands of the course in advance and manage the size of each year's cohort, as the course is quite popular. Participants must attend all five hour-long sessions in the fall and then observe two participants and be observed by two participants in the spring to achieve certification as trained faculty peer reviewers. We first offered the course in the fall of 2019 and have offered it every year since (including offering it online during the pandemic). One hundred faculty have participated in the training thus far.

The curriculum for the fall term includes an overview of the nuts and bolts of reviewing a colleague's teaching, a variety of readings, and homework between sessions. Participants learn about the pre-observation and post-observation meetings by watching video demonstrations, reading short articles, and role-playing in session. They also practice performing an actual observation by watching a curated video and taking structured notes. To supplement the instruction on observational review, we introduce participants to strategies for reviewing the course syllabus, as it distills the intellectual work of planning and teaching the course into one (hopefully) concise document and can serve as a good source of information for summative evaluation of teaching (Arreola, 2000). To assist participants in syllabus review, we provide a highly regarded and thoroughly researched rubric (Palmer et al., 2014), which our future peer reviewers then apply to sample syllabi for practice. Finally, participants review sample letters and discuss how best to document their review for summative purposes.

The instruction on how to conduct a pre-observation meeting is designed to help participants see it as an opportunity to learn about various contexts for the course and the session to be observed as well as any teaching challenges the instructor is facing. This meeting also serves to establish an all-important human connection between instructor and reviewer, the latter presenting as a supportive and

collaborative peer, not a critical superior (Bell et al., 2019). Participant feedback on the impact of the training on their attitudes toward peer review, which we describe in greater detail below, confirms the establishment of this supportive, collaborative relationship before the observation takes place.

To perform the observations, we furnish participants with a detailed note-taking template that includes a column for impartial descriptions of *what* was observed, a column for *when* it was observed, and a third column for any *comments* the observer wishes to make in the moment. Accompanying the template is a brief list of instructor characteristics to look for, such as organization and clarity, classroom management, and presentation skills. We include a similar list of student characteristics to watch for, such as preparation, engagement, and civility. The characteristics we suggest take the form of neither a rubric nor a checklist; rather, we offer them as guidelines to help participants train their attention on aspects of the classroom experience we have found most germane in helping instructors to see their own teaching through a second set of eyes and, if warranted, to solve teaching problems.

Our instruction on the keystone post-observation meeting focuses on helping reviewers learn how to share their impressions of the observed session and collaborate with the instructor to solve any potential teaching problems. Participants learn to begin by asking the instructor to self-assess the observed session and indicate whether it was a typical session (Bell et al., 2019). Reviewers then pose information-seeking or clarifying questions to better understand the session they observed. Next, the observer moves to the most important part of the post-observation meeting: posing probing questions (National School Reform Faculty, n.d.-a). These are questions that often have no quick or simple answer and help instructors reframe or ponder challenging teaching issues so they can discover their own solutions rather than be told what to do (Newman et al., 2019). Because probing questions can be quite challenging to construct for those new to them, we provide participants stems such as “What would it look

like if . . . ?” or “What’s another way you might accomplish . . . ?” Throughout the post-observation meeting (and explicitly at this stage of the conversation), we firmly instruct faculty participants to maintain a kind, positive attitude and to keep the conversation collegial and collaborative. This approach has resulted in participant feedback describing post-observation meetings as comfortable and supportive rather than anxiety producing or judgmental. Overall, the training helps participants realize that while the observation itself is important, the conversations before and after the observation are where the valuable work of reflecting on and strengthening an instructor’s teaching is done.

In the spring practicum each participant is observed twice and observes two colleagues. We require that all observations be strictly formative and that the peer pairs be from different disciplines. We impose these constraints to enable our participants to develop their reviewing skills in a safe environment without the added stress of having to evaluate a colleague and write a summative document. Peer pairs must be from different disciplines to enable them to follow best practice for peer review, focusing on pedagogy rather than content. The benefits of cross-disciplinary peer reviews have also been noted by Barrios-Rodríguez et al. (2023) and O’Keeffe et al. (2021). Our course wraps up with a final meeting in which the newly trained peer reviewers debrief their experiences reviewing their colleagues and offer us feedback about the training program.

In the fall of 2022, we rolled out an online version of the course to enable faculty with scheduling conflicts and instructors at other universities to access and complete the program at no charge. The online course teaches the basics of peer review (i.e., our fall curriculum) by leveraging new technologies such as Powtoon animations to demonstrate key moves in the pre-observation and post-observation meetings and live videos in PlayPosit with embedded questions and discussion prompts to engage viewers. The online course includes three modules, each with activities and quizzes to help learners review key concepts:

1. Addressing the fundamentals of observing and debriefing a peer's teaching
2. Procedures for reviewing course documents
3. Guidance—including templates—for documenting reviews

After completing the online modules, participants arrange a practicum—usually through their local teaching centers—to observe a colleague's teaching and to be observed. (To access the online course, visit <https://expand.iu.edu/courses/peer-review-of-teaching-certification-course>.)

Our training is designed to teach participants how to gather sufficient information from observing a single class session (and reviewing course documents) to be prepared to write a summative memo about the instructor they reviewed. Some scholars have argued that one classroom observation is inadequate for drawing conclusions—especially summative conclusions—about an instructor's teaching (Arreola, 2000; Chism, 2007). However, in practice busy schedules can make it very difficult for a faculty peer to observe more than one class session, especially when the observation rightly includes a pre- and a post-observation meeting (Greenhoot et al., 2022). We argue that a single observation preceded by a well-executed pre-observation meeting, followed by a rich post-observation conversation, and supplemented with thoughtful syllabus review yields ample material on which to base formative and summative conclusions.

Feedback From Faculty Participants

In the last session of our training program, we ask participants to reflect on the main takeaways from their training and practicum experiences and to describe how the training has changed their attitudes toward peer review and teaching more generally. In addition to gathering this reflective feedback from every cohort, we solicited additional formative feedback from our first two cohorts, via a feedback questionnaire and informal focus groups, to help us refine and improve the training

program. (We obtained approval from our university's Institutional Review Board to collect questionnaire and focus group feedback, specifically for the purpose of ensuring that we could quote participants' feedback anonymously in publications and conference presentations about our training program.) Taken together, all the formative feedback we have received from our participants has enabled us to improve the training and practicum experiences for future cohorts, while also giving us some insight into the impact of peer review training on participants' teaching practices and attitudes toward teaching and peer review.

One of our main goals in the training program is to help participants understand the purpose of peer review: not as a process for critiquing a colleague or offering advice, but rather as a collaborative, collegial interaction in which peers work together to appreciate teaching wins and address teaching challenges. Our participants' feedback confirmed their understanding of this basic precept. As one instructor noted in referring to the post-observation meeting, "I used to think peer review involved providing suggestions and advice. Now I think providing suggestions and advice should only be done when specifically requested, and the conversation should be collaborative and focused on questions and answers."

Another of our major goals for the training is to help participants learn to focus on pedagogical strategies rather than disciplinary content in their peer reviews. Early in the training, some participants expressed skepticism about their ability to achieve this goal, but their feedback after the practicum confirmed the empowering effects of observing and offering feedback to a colleague outside their discipline. Many noted (with some surprise) that they *could* discuss teaching challenges with a colleague even if they were not in the colleague's department or discipline (and in extreme cases, even if the class they observed was taught in a language they were unfamiliar with). The benefits of cross-disciplinary peer review were emphasized in comments like this one: "I used to think peer review was best completed by faculty in your home department. Now I see the value of having peers from different

departments add to the growth and development of teaching and how this can create more collaboration across campus.”

On a related note, we hope the training program changes participants’ attitudes toward peer review, and the feedback from both the individual reflections and the questionnaire responses confirm the desired change. Before the training, many reported seeing peer review as an uncomfortable, unfriendly, subjective procedure that would result in judgment. After the training, participants noted that while the procedure can be systematic and even rigorous, it can also be a collegial, collaborative process. As an instructor noted in their end-of-program reflection, “this is not about giving advice, but about listening to instructors and helping to guide them to their own insights.” From the survey data: “Before [the training] I was skeptical. Now I see a spirit of peers who want to help improve my skills and give honest feedback.” Almost all reported being more comfortable with the peer review process, more likely to seek a peer review in the future, and more likely to recommend peer review to their colleagues. Even participants who had a positive attitude toward peer review before the training pointed out the difference the training made. As one participant noted:

My attitude [about peer review] has been the same, but the experience has been different because with this training I have been observed by people who have been trained and who had a positive approach to the process. It is hard to be observed by someone who has not been trained.

The feedback from our participants seemed to confirm that the training helped them to see the peer review process as a collaborative, positive interaction among peers and empowered them to focus on pedagogy rather than disciplinary content. As a corollary of these goals, our participants noted the impact of the training program on their own teaching methods, showing an eager willingness to use or adapt teaching moves they had seen while observing their peers. They described a new appreciation for the diversity of teaching styles while

also recognizing that many teaching challenges are universal. As one instructor noted, "I certainly felt less alone as I saw great teachers struggling with distracted students and trying to find creative ways to engage them."

While the feedback about both the training sessions and the practicum was in general strongly positive, we did receive a few less-than-positive comments from participants regarding their practicum experiences. For example, one peer reported that their observer failed to maintain the confidentiality of the observation; instead, the observer commented about the observation in a crowded hallway surrounded by students and faculty colleagues! Another participant reported to us that they felt disrespected by their observer in the post-observation meeting. They said the observer reported problems with their teaching but failed to note any positive or effective pedagogical moves; the conversation left the peer questioning their teaching style and doubting their ability. Learning of these incidents has helped us to improve the training for subsequent cohorts. For example, when providing instructions for the practicum we now put more weight on confidentiality, to ensure that all our participants appreciate its importance and know how to maintain it. And since the post-observation meeting is the most challenging part of the training for faculty observers, we have also worked to improve participants' approach to this interaction. We have doubled down on our emphasis on maintaining a collegial and supportive relationship with the observed instructor, and we stress the importance of acknowledging their vulnerability at this stage. We also provide more practice in asking probing questions to ensure the observer avoids judgment and instead encourages their peer to reflect deeply on their teaching.

The feedback from our faculty participants confirms the observations of others about the responses of faculty to peer review and classroom observation. For example, Kohut et al. (2007) noted the beneficial impact of peer review on the teaching of both observer and the observed instructor. Barrios-Rodríguez et al. (2023) confirmed that negative impressions of peer review vanish after going through the

peer review process. Several authors (Bell et al., 2019; Hendry et al., 2021; Newman et al., 2019) have confirmed that when a strong, trusting relationship is established between the observer and the observed instructor, peer review can increase the observed instructor's confidence as a teacher. Several of these themes are also mentioned in the comprehensive review of peer review of teaching programs provided by Cutroni and Paladino (2023). The beneficial effects of cross-disciplinary peer review, which allows both parties to focus on pedagogy rather than curriculum, have been reported previously (Barrios-Rodríguez et al., 2023; O'Keeffe et al., 2021). Several authors have also noted the importance of establishing an atmosphere of respect and trust between the peer pair, which our training accomplishes through pre-observation and post-observation meetings (Kohut et al., 2007; O'Keeffe et al., 2021). This trusting, yet honest and open relationship between peers has been described by others (Lomas & Nicholls, 2005; O'Keeffe et al., 2021) as one of "critical friends" (based on the concept of a "critical friends group" of teaching professionals, developed by the National School Reform Faculty; National School Reform Faculty, n.d.-b). These revelations demonstrate that peer review not only fosters a sense of community among faculty across departmental and disciplinary silos; it can also lead to more collaborative problem-solving among instructors, which often leads to better teaching and enhanced student learning (Cutroni & Paladino, 2023).

Lessons Learned

While participant feedback from surveys and interviews confirmed a broad range of learning and growth as well as a high level of satisfaction with the peer review training course, we also learned some important lessons from facilitating the course. First, we came to realize that while formative and summative reviews have different purposes, we believe that *all* reviews should emanate from a place that is supportive, collaborative, and essentially formative (Bell et al., 2019; Engin, 2016;

Gosling, 2013; Greenhoot et al., 2022; Newman et al., 2019; Roberson, 2006). Whether arising from an informal, spontaneous observation or a formal, high-stakes evaluative review or something in between, we instruct our peer review trainees to let the observed instructor determine the course of the review and to approach the interaction with kindness and understanding. The content of a summative memo should fairly represent the post-observation discussion between the instructor and the peer observer, including both recognition of teaching strengths and articulations of areas for growth; thus, the written document should come as no surprise to the instructor. In both the post-observation debrief and in the summative document, the idea is not to judge, but to give the instructor fair opportunities to grow and improve.

As we have led the course, we have also noticed additional spots where faculty typically struggle, such as how to write an effective summative memo. Faculty worry about the memo's tone, its structure, and whether their words will have a negative effect on a colleague's chances for tenure or promotion. To assist faculty in writing summative memos, we have developed a template with extensive annotations for participants to explore during the course and later use when needed.

Another challenge we have come to recognize is our request that participants avoid immediate judgment during observations and subsequently when discussing the observation. Trained to judge and critique academic work as quickly and efficiently as possible, faculty commonly expect to apply the same approach to peer review. It takes concerted effort on our part to insist that they objectively describe what they saw when debriefing the video that all participants collectively observe as part of our curriculum. Many participants jump to judgment immediately and have to be gently corrected when they begin to do so. While their judgments may not be unwarranted, they nevertheless skip the important step of collecting observational data, upon which later productive conversations will rely. Similarly, during the post-observation discussion, trainees must learn to resist the "criticize and suggest" approach and instead employ a consultative strategy, encouraging the observed instructor to reflect critically on

their own teaching, asking questions rather than making statements, referring to observed behaviors rather than making critical judgments. Highlighting what was observed followed by an invitation to consider its meaning typically allows instructors to reflect and guide the conversation in a productive manner. Along the same lines—and equally difficult for our trainees—we insist they offer suggestions only when the instructor specifically asks for them.

Finally, we have come to realize that much of our work throughout the course entails the deceptively simple project of helping faculty learn how to talk to each other about teaching in a manner that is sensitive, nuanced, humble, and solicitous, qualities that do not come naturally to many and must be practiced. As others have pointed out, if we want peer review to promote university teaching as a collaborative and public act, faculty must learn how to talk about teaching—both others' and their own—in a supportive, congenial fashion (Gosling, 2013; Greenhoot et al., 2022; Hutchings, 1994, 1996). They need to wear the proverbial consultant's hat, because it is the *conversation*—before and after an observation or review of course documents—that matters as much as what is seen and heard in the class session. When reviewers manage conversations as would a consultant, instructors can really *hear* feedback about their teaching because they feel comfortable, safe, and secure hearing it from a well-trained peer.

Conclusion

Even though there have long been strong arguments in support of peer review of teaching generally, we began our work specifically in response to an unmet need at our university for trained peer reviewers. University administration began to signal expectations for peer reviews in tenure and promotion dossiers but had no plans in place to train faculty to perform them. Thus did we develop a course to help faculty achieve the competence and confidence they needed to review their peers' teaching.

The popularity and demand for our course surprised us, and, as if to emphasize the point, when we presented our work at a recent POD conference, we were greeted by a standing-room-only crowd of over 100 attendees. This signaled to us that we needed to share our work with the broader educational development community to enable others to benefit from our work.

We hope we have succeeded in showing both the theoretical basis and the practical need for our course. We believe the curriculum, created with busy faculty in mind, balances the knowledge one needs and the skills one must practice and master to perform an effective and successful review. We invite our colleagues to adapt it to their own contexts so that faculty across the academy will be prepared to review their peers' teaching in a way that is useful to their institution and to the peers engaged in the review process.

Biographies

Eric T. Metzler (emetzler@iu.edu) is the Instructional Support and Assessment Specialist at the Kelley School of Business, Indiana University Bloomington and Indianapolis. In his role as instructional consultant, Eric works confidentially with business faculty who wish to improve, refine, or otherwise perfect their teaching performance. As assessment specialist, Eric helps instructors, departments, programs, centers, and institutes at the Kelley School measure student learning and use the data collected to improve student learning and the student experience.

Lisa Kurz (kurz@indiana.edu) is Principal Instructional Consultant for Non-Tenure Track Development in the Center for Innovative Teaching and Learning at Indiana University Bloomington. Her main work as an educational developer focuses on supporting the professional and career development of teaching faculty (those not on the tenure track) through individual consultations, teaching institutes, and faculty learning communities. She also specializes in course and curricular

design and in helping faculty implement innovative assessment and grading strategies.

Acknowledgments

We are grateful to the many people who made this article possible. To our reviewers, who offered the friendly, helpful criticism that resulted in a stronger article, we give our thanks. We also thank our participants, who shared written feedback and generously gave of their time to be interviewed individually and in focus groups. Finally, we would like to thank our colleague, Michael Morrone, director of Indiana University's Faculty Academy on Excellence in Teaching (FACET) Community, who gave us opportunities to make our work public, both in conference settings and online. These opportunities enabled us to think deeply about our work, revising and refining it into its current form.

Conflict of Interest Statement

The authors have no conflict of interest.

References

- Andersen, K., & Miller, E. D. (1997). Gender and student evaluations of teaching. *PS: Political Science & Politics*, 30(2), 216–219. <https://doi.org/10.1017/S1049096500043407>
- Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system* (2nd ed.). Anker.
- Barrios-Rodríguez, R., Salcedo-Bellido, I., Jiménez-Moleón, J. J., Lozano-Lorca, M., Galiano-Castillo, N., Cobos, E. J., Vilchez Rienda, J. D., Olmedo-Requena, R., Amezcua-Prieto, C., Martín-Peláez, S., González Domenech, C. M., Arrebola Moreno, J. P., Rica, R. A., García-Rubiño, M. E., & Requena, P. (2023). Peer review of teaching: Using the nominal group technique to

- improve a program in a university setting with no previous experience. *International Journal for Academic Development*, 28(4), 385–397. <https://doi.org/10.1080/1360144X.2022.2032718>
- Bay View Alliance. (n.d.). *Evaluating teaching: Transforming the evaluation of teaching (TEval)*. <https://bayviewalliance.org/racs/evaluating-teaching/>
- Bell, A., Meyer, H., & Maggio, L. (2019). On the same page: Building best practices of peer coaching for medical educators using nominal group technique. *MedEdPublish* 8:95. <https://doi.org/10.15694/mep.2019.000095.1>
- Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Carnegie Foundation for the Advancement of Teaching. Wiley.
- Buser, W., Batz-Barbarich, C. L., & Kearns Hayter, J. (2022). Evaluation of women in economics: Evidence of gender bias following behavioral role violations. *Sex Roles*, 86, 695–710. <https://doi.org/10.1007/s11199-022-01299-w>
- Chism, N. V. N. (2007). *Peer review of teaching: A sourcebook* (2nd ed.). Anker.
- Cutroni, L., & Paladino, A. (2023). Peer-ing in: A systematic review and framework of peer review of teaching in higher education. *Teaching and Teacher Education*, 133, 104302. <https://doi.org/10.1016/j.tate.2023.104302>
- Engin, M. (2016). Enhancing the rigour of peer observation through the scholarship of teaching and learning. *International Journal for Academic Development*, 21(4), 377–382. <https://doi.org/10.1080/1360144X.2016.1225576>
- Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., & Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLoS One*, 14(2), e0209749. <https://doi.org/10.1371/journal.pone.0209749>
- Godbout-Kinney, K., & Watson, G. P. L. (2022). Institutional approaches to evaluate teaching effectiveness: The role of summative peer review of teaching for promotion and tenure. *Canadian Journal of Educational Administration and Policy / Revue canadienne en administration et politique de l'éducation*, 201, 2–14. <https://doi.org/10.7202/1095479ar>
- Gosling, D. (2013). Collaborative peer-supported review of teaching. In J. Sachs & M. Parsell (Eds.), *Peer review of learning and teaching in higher education* (Professional Learning and Development in Schools and Higher Education, vol. 9, pp. 13–31). Springer. https://doi.org/10.1007/978-94-007-7639-5_2
- Greenhoot, A. F., Austin, A., Cornejo Weaver, G., & Finkelstein, N. D. (2022, May 24). How peer review could improve our teaching. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/how-peer-review-could-improve-our-teaching>.
- He, J., & Freeman, L. A. (2021). Can we trust teaching evaluations when response rates are not high? Implications from a Monte Carlo simulation. *Studies in Higher Education*, 46(9), 1934–1948. <https://doi.org/10.1080/03075079.2019.1711046>

- Hendry, G. D., Georgiou, H., Lloyd, H., Tzioumis, V., Herkes, S., & Sharma, M. D. (2021). "It's hard to grow when you're stuck on your own": Enhancing teaching through a peer observation and review of teaching program. *International Journal for Academic Development*, 26(1), 54–68. <https://doi.org/10.1080/1360144X.2020.1819816>
- Hutchings, P. (1994). Peer review of teaching: "From idea to prototype." *American Association of Higher Education Bulletin*, 47(3), 3–7.
- Hutchings, P. (1996). *Making teaching community property: A menu for peer collaboration and peer review*. American Association for Higher Education.
- Kohut, G. F., Burnap, C., & Yon, M. G. (2007). Peer observation of teaching: Perceptions of the observer and the observed. *College Teaching*, 55(1), 19–25. <https://doi-org.proxyiub.uits.iu.edu/10.3200/CTCH.55.1.19-25>
- Kreitzer, R. J., & Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, 20(1), 73–84. <https://doi.org/10.1007/s10805-021-09400-w>
- Lomas, L., & Nicholls, G. (2005). Enhancing teaching quality through peer review of teaching. *Quality in Higher Education*, 11(2), 137–149. <https://doi.org/10.1080/13538320500175118>
- National Academies of Sciences, Engineering, and Medicine. (2020). *Recognizing and evaluating science teaching in higher education: Proceedings of a workshop—In brief*. National Academies Press. <https://doi.org/10.17226/25685>
- National School Reform Faculty. (n.d.-a). Probing questions. In *Evolving glossary of NSRF terms*. <https://nsrfharmony.org/glossary/>
- National School Reform Faculty. (n.d.-b). *What IS Critical Friends Group work?* <https://nsrfharmony.org/whatiscfgwork/>
- Newman, L. R., Roberts, D. H., & Frankl, S. E. (2019). Twelve tips for providing feedback to peers about their teaching. *Medical Teacher*, 41(10), 1118–1123. <https://doi.org/10.1080/0142159X.2018.1521953>
- O’Keeffe, M., Crehan, M., Munro, M., Logan, A., Farrell, A. M., Clarke, E., Flood, M., Ward, M., Andreeva, T., Van Egeraat, C., Heaney, F., Curran, D., & Clinton, E. (2021). Exploring the role of peer observation of teaching in facilitating cross-institutional professional conversations about teaching and learning. *International Journal for Academic Development*, 26(3), 266–278. <https://doi.org/10.1080/1360144X.2021.1954524>
- Palmer, M. S., Bach, D. J., & Streifer, A. C. (2014). Measuring the promise: A learning-focused syllabus rubric. *To Improve the Academy: A Journal of Educational Development*, 33(1), 14–36. <https://doi.org/10.1002/tia2.20004>
- Peterson, D. A. M., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS One*, 14(5), e0216241. <https://doi.org/10.1371/journal.pone.0216241>

- Roberson, W. (Ed.). (2006). *Peer observation and assessment of teaching: A resource book for faculty, administrators, and students who teach*. Center for Effective Teaching and Learning and Instructional Support Services, University of Texas at El Paso. https://www.utep.edu/faculty-development/_Files/docs/utep_peer_observation_booklet.pdf
- Shulman, L. S. (1993). Teaching as community property: Putting an end to pedagogical solitude. *Change: The Magazine of Higher Learning*, 25(6), 6–7. <https://doi.org/10.1080/00091383.1993.9938465>
- Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus*, 134(3), 52–59. <https://doi.org/10.1162/0011526054622015>
- Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, 12(4), 618–627. <https://doi.org/10.1187/cbe.13-08-0154>
- Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1>
- Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Psychology*, 42(4), 276–294. <https://doi.org/10.1080/01973533.2020.1756817>
- TEval. (n.d.). *Frameworks and rubrics*. <https://teval.net/index.php?frameworks>
- University Faculty Council. (2022, April 26). Faculty and librarian promotions. UFC policy ACA-38. <https://policies.iu.edu/policies/aca-38-faculty-librarian-promotion/index.html>