Transformative Dialogues: Teaching and Learning Journal
Fall 2024, Volume 17, Issue 3
https://doi.org/10.26209/td2024vol17iss31826

Integrating Educational Development and Assessment of Student Learning Through Peer Review of Teaching

Melissa Eblen-Zayas
Carleton College

Morgan Iommi
Nevada State University

Abstract
Educational developers seek to improve faculty practice. Assessment offices look to assess student learning. Connecting educational development and student learning assessment efforts can provide a better understanding of how faculty practices are directly impacting student outcomes and identify areas for improvement. Peer review of teaching can help create the collaborative culture needed to support shared understanding of student learning outcomes and how faculty practice can help achieve those outcomes. Peer review of teaching adds an important dimension of triangulation in addition to individual faculty reflection and student evaluations of teaching by providing a more holistic context for the artifacts of student learning used for assessment. The frameworks of peer review of teaching can foster conversations among colleagues about student learning outcomes, practices for implementing instructional approaches that help achieve those outcomes, and areas for faculty growth. Faculty can learn from each other about what is working effectively and where they might need to modify their approaches. However, the goals of peer review of teaching can vary based on the approach and structural design. In this conceptual piece, we will explicate several peer observation models, review the key elements of peer review of teaching for varying goals, and consider how these various models can connect the objectives related to educational development and assessment of student learning.

Keywords
educational development, assessment of student learning, peer observation of teaching, peer review of teaching

Introduction

Continuous improvement and professional development are important aspects of being an educator. Best practices in teaching, learning, and educational technology are rapidly advancing

while the landscape of higher education and the type of students enrolling are shifting. Due to these factors, being a reflective educator is increasingly important to ensure that instructors can identify where students' learning falls short and what instructional changes they can make to address shortcomings and ensure student success. However, many areas of higher education are still siloed or outdated. Instructors often enter the profession with no formal training in teaching and thus mimic the teaching practices they received unless the institution provides support for instructor professional development. Similarly, instructors often have no background in assessment of student learning and being asked to contribute to assessment activities can feel like disconnected extra work on top of teaching, research, and other responsibilities. Integrating practices in student assessment and educational development could provide a more robust understanding of student learning and instructional effectiveness. This practice could help identify where instructors could make changes and how educational developers can better support them in doing so.

Reflective educators who are focused on improving student learning need to collect and analyze data about the effectiveness of their teaching in an effort to ensure student success. Student evaluations, peer assessment, and self-evaluation all provide data about one's teaching (Smith, 2014). However, because teaching is so complex, each data source alone may not give enough information to effectively analyze one's teaching effectiveness (Berk, 2005). Therefore, it is generally agreed upon that using three or more data sources, commonly referred to as triangulation, is most effective in understanding the connection between instructional choices, student experiences in the classroom, and students' learning outcomes. Using multiple data sources (both indirect and direct) can help increase the validity and reliability of the feedback provided to instructors by highlighting themes and linkages across the multiple data sources (Appling et al., 2001).

However, many institutions still prioritize data from student evaluations, relying on student feedback to make decisions ranging from what classes an instructor teaches to the tenure and promotion process. Yet, student evaluations are not indicative of the learning that occurred in the course. Additionally, student evaluations are biased and often unreliable (Esarey & Valdes, 2020). Therefore, more comprehensive approaches are needed to provide both formative and summative feedback as a foundation for evidence-based teaching (Appling et al., 2001).

To improve instructional effectiveness and address the need for more robust data on teaching effectiveness, many institutions have centers for teaching and learning or other initiatives to offer educational development programs and support instructors who teach on campus or virtually (Reder, 2010). These centers offer many resources, from one-off workshops to certificate programs to online asynchronous resources (Herman, 2012). Each of these resources can fill the needs of different types of instructors while also meeting various goals to support continuous improvement. However, centers for teaching and learning are often separate from offices focused on assessment of student learning. Therefore, efforts do not always connect the effectiveness of instructor approaches to student learning data. One type of educational development offered by many centers that can help instructors triangulate data about teaching effectiveness, with the goal of connecting it to student learning, is peer review of teaching. In particular, peer review opportunities that move beyond simply watching colleagues teach and instead encourage instructors to engage in robust conversations about the instructional choices they make have the potential to connect educational development and assessment of student learning efforts.

Peer Review of Teaching

Several recent review papers have synthesized reports on the impact of peer review of teaching (Zeng, 2020), developed a framework for formative peer review based on the existing literature (Esterhazy et al., 2021), and summarized key factors that are important for development of peer review programs (Cutroni & Paladino, 2023). One of the themes that emerges from these reviews is that peer review of teaching takes many different forms with a variety of goals. Even the language around peer review of teaching varies. Different programs use "peer observation of teaching" and "peer review of teaching," for example, often without defining the meanings of these terms.

Peer review of teaching usually includes at least three elements: pre-observation preparation that involves the observer and the instructor sharing information and clarifying goals for the observation, the observation itself, and reflection/discussion after the observation. The length and number of observations vary significantly between peer review programs. In addition, programs differ in terms of who is observing, whether the observations are reciprocal, and what type of training or guidance participants receive.

Gosling (2002) provided an early categorization of the three main types of peer observation of teaching: evaluative (instructors are observed by more senior instructors or administrators for an institutional review process), developmental (instructors are observed by an "expert" who has more knowledge than the individual being observed), and collaborative peer review (instructors observe each other). The three models are summarized in Table 1. Gosling noted that even the term "peer" in "peer observation of teaching" takes on different meanings in each model. Our focus will be on the latter two models, which are non-evaluative and have the most potential for integrating educational development and assessment of student learning. Unlike Gosling, we use the term peer review of teaching as the overarching term, rather than peer observation of teaching, but in the subsequent discussion of particular studies, we vary terms to reflect the word choice used by authors.

Table 1

Gosling's Models of Peer Observation with Frameworks Identified by Zeng

	Evaluative	Developmental	Collaborative Peer Review
Purpose	Making personnel decisions about reappointment, tenure, or merit raises; identify underperformance	Identify strengths areas for improvement; increase use of evidence-based effective teaching practices	Engage in self and mutual reflection through discussion about teaching

Participants	Senior faculty or administrators observe instructors as required by policies	Educational developers or trained peers observe instructors who have chosen to participate	Instructors at any level or in any discipline observe each other – often in a reciprocal manner
Observation framework	Observer uses a required framework with criteria defined by the institution	Observer may choose a framework that addresses specific pedagogical goals; no required framework	No framework; freedom and flexibility to make one's own choices about what is observed and how

Zeng (2020) suggested that rather than sorting approaches into three clear types, one might consider peer observation activities/programs to occur on a continuum from developmental to evaluative. The type of framework provided to guide the observation is important in differentiating between approaches (Zeng, 2020). Table 1 identifies these frameworks. Gosling's collaborative peer review category of observations is a self-initiated developmental approach without an institutional or expert framework provided, while Gosling's developmental category of observations is an approach supported by educational developers or other experienced individuals who may use a guiding framework for the conversations. Yiend et al. (2014) called into question the quality of feedback in Gosling's collaborative peer review model and suggested there are benefits to peer observation opportunities that integrate developmental observations and collaborative peer review.

Cutroni and Paladino (2023) noted that the peer review program outcomes reported for the 82 articles in their review generally fell into several categories. One group of outcomes focused on the peer review process itself, gathering information from participants about attitudes towards collegial professional development and perceptions about the structure of the peer review program. The other group of outcomes was related to how peer review influenced teaching and learning through providing evidence of effective teaching practices (either through feedback from the observer or through reflection on their own practices), producing teaching practice improvements, or impacting student learning and/or engagement. By far, the outcome reported least frequently (in less than 10% of the articles) was the impact of peer review programs on student learning or engagement (Cutroni & Paladino, 2023).

Peer review often focuses on the mechanics of what the teacher does in the classroom (e.g., the tone and volume of voice, visuals used in lecture, etc.). In such scenarios, the connection between what the instructor does and student learning is completely missing. However, when structured appropriately, peer review programs can provide an opportunity for reflection on learning objectives and curricular design, how to make student learning visible during class time, and opportunities for adjusting instruction in response to assessment of student learning and self-reflection (through triangulation of direct and indirect evidence). These programs can integrate

educational development and assessment efforts by working directly with instructors to improve student learning and instructional effectiveness.

Goals and Models of Peer Review of Teaching

While peer review programs can have a variety of goals, our focus is on how peer review of teaching can be used to support both effective teaching and enhanced student learning. This outcome typically occurs in peer review programs in one of three ways: by explicitly promoting constructive alignment (Drew & Klopper, 2014; Tobiason, 2022; 2023), by serving as an additional source of evidence that can be used in efforts to triangulate data about both instructor teaching and student learning (Bernstein, D. J. et al., 2000; Drew & Klopper, 2014; Galbraith & Merrill, 2012), and by fostering reflection and conversations beyond the level of individual instructors about learning objectives and how they are met (Hammersley-Fletcher & Orsmond; Stieha et al., 2016).

The Peer-Assisted Reflections On Student Learning (PAROSL) program, developed at the University of California – Los Angeles, is a reciprocal, non-evaluative peer observation focused on encouraging instructors to design student-centered instruction aligned to learning outcomes, also referred to as constructive alignment (Tobiason, 2022; 2023). The pre-observation meeting is designed to get the instructor thinking about student learning and assessment through pre-observation questions such as: "What do you want students to learn during this lesson, and how will you know if they have learned it?" "What will students be doing to move towards the intended learning?" and "How will current learning (relative to intended learning) be made visible?" (Tobiason, 2023). Debrief meetings include similar types of questions, such as, "Which tasks/activities were most useful in making student learning visible?" (Tobiason, 2023). While the program focuses on lesson-level alignment of in-class activities to learning objectives, rather than the program-level alignment often used for assessment, faculty participants in PAROSL noted that the program helped them realize the importance of using learning objectives to focus what they did in the classroom. Additionally, PAROSL encourages instructors to make student progress toward meeting those objectives visible through class activities. Such an approach supports instructors in planning for student assessment embedded in instruction, which is one of the emerging trends in the assessment of student learning (Kinzie et al., 2019).

Other peer review programs have made efforts to support constructive alignment by facilitating the triangulation of multiple types of data. An Australian university developed a peer observation program where pairs of observers – one a disciplinary expert and one a learning and teaching expert – observed instructors in the STEM disciplines (Drew & Klopper, 2014). The observations from the two observers (who observed both what the instructor did and what the students were doing) were combined with student feedback. The student feedback included minute papers from the end of class responding to the question, "What is the most important thing you learned?" and student evaluations of the instructor along ten dimensions of teaching effectiveness. The observers were also encouraged to pay attention to the same ten elements. The instructor was observed twice; they received feedback from the first class visit and had conversations with the observers about future actions before the second class visit. The data from students and observers were triangulated with instructor reflections to identify how peer observation impacted ten dimensions of teaching effectiveness across the STEM disciplines. Aggregated across STEM departments, the data suggested that the most improved dimension of teaching effectiveness (from the peer review of

teaching process) was curricular design and structure within lessons. On the other hand, the dimension of teaching effectiveness that showed the least change was instructor likelihood to use formative assessment (Drew & Klopper, 2014).

A large American land-grant university had a peer review of teaching program that involved pairs of instructors in the same discipline participating in a three-part peer review process that included 1) reviewing intellectual content of the course, including the statement of course goals and rationale for those goals, which was observed through annotated syllabi and narrative accounts, 2) articulating specific goals for class time with students and rationale for the planned activities, which was observed through class visits, and 3) making sense of what kinds of understanding students reached, which was observed through copies of exams and assignments, examples of student work, and the feedback the instructors provided to students (Bernstein et al., 2000). Unlike PAROSL, which was focused on constructive alignment for single-class sessions, this approach was much more expansive. By collecting information about student performance on assignments and exams in addition to instructor rationale for their instructional choices, this program connects direct evidence of student learning to what is observed through class visits. Unfortunately, less than half of the instructors participating in this program saw student learning increase. Those instructors whose classes did show an improvement primarily did so primarily through 1) an increase in the usefulness of feedback that instructors provided to students on their work and 2) changing the class assessments to ones that required more application, analysis, and synthesis, rather than a straightforward assessment of comprehension. Many instructor participants in the program continued to hold the view that students are solely responsible for their learning and failed to connect how the quality of teaching might impact student learning (Bernstein et al., 2000).

Peer review opportunities that move beyond individual instructors and encourage larger groups of instructors to articulate learning objectives and reflect on how choices about teaching promote student learning have the potential to bridge the gap that often exists between educational development and assessment of student learning. Such groups could include instructors who teach within a sequence in a department or instructors who teach a particular type of general education course. Kinzie et al. (2019) note that one area for CTLs to partner with assessment staff is to "work with [groups of] faculty and departments to articulate goals for student learning that are shared and broadly understood" (p. 52). These are the types of reflective conversations that peer review of teaching can support if structured in a particular manner.

For example, at a mid-sized American doctoral university, a peer observation program that employed the Reformed Teaching Observation Protocol (RTOP) was designed as a way to encourage instructors of general education STEM classes to adopt evidence-based instructional practices and connect it with assessment data (Sawada et al., 2002; Stieha et al., 2016). In this program, the peers who conducted the observations were graduate students trained in the RTOP instrument, and each instructor was observed three times. The researchers found that, although observations added a certain intentionality to instructional teaching choices, “disconnections between teaching practice and learning outcomes assessment was a persistent theme” (Stieha et al., 2016, p. 99). What the program did succeed in doing, primarily through focus group conversations among instructors, was to leverage instructor-instructor interactions to develop a common vocabulary about teaching and assessment and shift perspectives, particularly with regard to evidence-based instructional practices.

Bernstein et al. (2000) planned to expand their peer review of teaching program to include targeted efforts to bring instructors together from within the same department or instructors who teach courses that meet general education requirements. That way, instructors could engage in conversations about the connections between their teaching and the broader curricular expectations or trends. Even if instructors do not have clear connections by discipline or institutional affiliation, peer observation can help instructors think about how they teach skills (such as communication skills) by observing how those skills are taught in different contexts (O'Keeffe et al., 2021). Designing peer review programs that allow sustained conversation among multiple instructors teaching within a coherent aspect of the curriculum (a particular general education course, a core sequence for a major, etc.) are programs that have the potential to promote consideration of the connection between course design, teaching effectiveness, and student learning in a more authentic way than programs that involve one-time observations or simply bring together pairs of instructors.

Recommendations

Peer review of teaching can be beneficial to understanding and improving teaching effectiveness both by providing feedback from peers and encouraging instructor self-reflection. If the program is carefully designed, instructors can also gain a better understanding of how to assess student learning. However, in order to best utilize peer review of teaching to triangulate data and integrate assessment of student learning, the literature suggests several program design considerations.

High-Quality Feedback Comes When Observers Have Expertise or Training

One of the common concerns about peer review is the quality of feedback given. Collaborative peer review (in Gosling's model) tends not to include the type of feedback that encourages critical reflection and revision of one's instructional approaches (Yiend et al., 2014). Among the ways to address this are to have reviewers include both a disciplinary and a pedagogical expert (O'Keeffe et al., 2021) or to have the review by a disciplinary educational researcher in the same field as the person being observed (Georgiou et al., 2018). Training instructor peer observers is another option. Instructors who have seen models of feedback that foster critical reflection and encourage its application to improve instructional practice are more likely to provide higher quality and more in-depth feedback to their peers (Yiend et al., 2014). Regardless of who is observing, if not well-structured, peer observation reflections can remain superficially focused on the course-level teaching and delivery process rather than moving into a more wholistic realm of inquiry that connects quality teaching with student learning (Hammersley-Fletcher & Orsmond, 2005; O'Leary & Savage, 2020).

Participants Need Time to Build Trust with Each Other, and How Observers are Paired Impacts the Quality of the Observation

For peer review of teaching to be effective in informing and strengthening the intersection of educational development and assessment of student learning, the programs need to be structured in a way that there is sufficient relationship building and participants trust each other. In one cross-institutional peer observation program, participants noted that they could only move from an initial focus on mechanical issues of practice to holistic consideration of how class activities feed into learning outcomes once trust had been established between individuals in the program (O'Keeffe et al., 2021).

Although it might be tempting to allow instructors to choose their observation partners because a trusting relationship already exists, Galbraith and Merrill (2012) found that peer observation evaluations of teaching effectiveness better correlated with student learning outcomes when a peer observer was assigned rather than when the instructor chose their peer observer. This finding suggests that if the goal of a peer observation program is to inform efforts at assessment of student learning, peer observers should not be selected by the instructor being observed.

Bring Together Groups of Instructors Who Teach a Particular Skill or General Education Requirement

Design reflection opportunities that move beyond discussions of what happens in individual classrooms to considerations of broader issues of learning within a department or even across an institution (Hammersley-Fletcher & Orsmond, 2005); such discussion can develop a shared vocabulary around learning objectives and instructional approaches. In a peer observation program that includes multiple instructors who have some commonality in what they teach, collective reflection can ensure that the discussion is not focused on superficial elements of teaching. Instead, the consideration becomes what the group can do better to support student learning within a department or as part of the general education curriculum. These collective conversations can then be used to identify strategic professional development opportunities to improve practice and strengthen student learning (Drew & Klopper, 2014). As Kinzie et al. (2019) note, "The purpose of assessment is not to generate reports but to inform ongoing efforts to improve what happens for students in their courses" (pg. 53). Educational developers and assessment professionals can use peer observation conversations to inform the design of programming to support instructors in changing what they do in the classroom to better support student learning.

Future Directions

Peer review benefits education by giving important feedback to instructors and administrators about teaching effectiveness. However, the literature is still ambiguous and inconsistent in the terms and definitions used for such programs. Future efforts to clarify the language used and structures of such programs will benefit the field, particularly when creating programs that can effectively integrate assessment of teaching effectiveness and student learning. Furthermore, because the terms and structures are so inconsistent, the frequency and effectiveness of different types of programs have not been well established. Future research should seek to understand the makeup of these programs, develop more consistent frameworks, and begin to study the effectiveness of program elements for meeting various goals in different contexts.

References

Appling, S. E., Naumann, P. L., & Berk, R. A. (2001). Using a faculty evaluation triad to achieve evidence-based teaching. *Nursing Education Perspectives*, *22*(5), 247.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, *17*(1), 48–62.

Bernstein, D. J., Jonson, J., & Smith, K. (2000). An examination of the implementation of peer review of teaching. *New Directions for Teaching and Learning*, *83*, 73–86.

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.

Drew, S., & Klopper, C. (2014). Evaluating faculty pedagogic practices to inform strategic academic professional development: a case of cases. *Higher Education*, *67*(3), 349–367. https://doi.org/10.1007/s10734-013-9657-1

Esarey, J., & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assessment & Evaluation in Higher Education*, *45*(8), 1106-1120. https://doi.org/10.1080/02602938.2020.1724875

Esterhazy, R., de Lange, T., Bastiansen, S., & Wittek, A. L. (2021). Moving beyond peer review of teaching: A conceptual framework for collegial faculty development. *Review of Educational Research*, *91*(2), 237-271. https://doi.org/10.3102/0034654321990721.

Galbraith, C. S., & Merrill, G. B. (2012). Predicting student achievement in university-level business and economics classes: peer observation of classroom instruction and student ratings of teaching effectiveness. *College Teaching*, *60*(2), 48–55. https://doi.org/10.1080/87567555.2011.627896

Georgiou, H., Sharma, M., & Ling, A. (2018). Peer review of teaching: What features matter? A case study within STEM faculties. *Innovations in Education and Teaching International*, *55*(2), 190–200. https://doi.org/10.1080/14703297.2017.1342557

Gosling, D. (2002). Models of peer observation of teaching. *The Generic Centre: Learning and Teaching Support Network*.

Hammersley-Fletcher, L., & Orsmond, P. (2005). Reflecting on reflective practices within peer observation. *Studies in Higher Education*, *30*(2), 213–224. https://doi.org/10.1080/03075070500043358

Herman, J. H. (2012). Faculty development programs: The frequency and variety of professional development programs available to online instructors. *Journal of Asynchronous Learning Networks*, *16*(5), 87-106.

Kinzie, J., Landy, K., Sorcinelli, M. D., & Hutchings, P. (2019). Better together: How faculty development and assessment can join forces to improve student learning. *Change: The Magazine of Higher Learning*, *51*(5), 46–54. https://doi.org/10.1080/00091383.2019.1652076

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

O'Leary, M., & Savage, S. (2020). Breathing new life into the observation of teaching and learning in higher education: moving from the performative to the informative. *Professional Development in Education*, *46*(1), 145–159. https://doi.org/10.1080/19415257.2019.1633386

Reder, M. (2010). Effective practices in the context of small colleges. In: Gillespie, K.J., and Robertson, D. L. (Eds.), *A guide to faculty development.* 2nd ed., 293–308. Jossey-Bass,

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*(6), 245–253. https://doi.org/10.1111/j.1949-8594.2002.tb17883.x

Smith, S. L. (2014). Peer collaboration: Improving teaching through comprehensive peer review. *To Improve the Academy*, *33*(1), 94–112.

Stieha, V., Shadle, S. E., & Paterson, S. (2016). Stirring the pot: Supporting and challenging general education science, technology, engineering, and mathematics faculty to change teaching and assessment practice. *The Journal of General Education*, *65*(2), 85–109. https://doi.org/10.5325/jgeneeduc.65.2.0085

Tobiason, G. (2022). Going small, going carefully, with a friend: Helping faculty adopt lesson-level constructive alignment through non-evaluative peer observation. *Active Learning in Higher Education 25*(1), 53–66. https://doi.org/10.1177/14697874221092977

Tobiason, G. (2023). From content-centered logic to student-centered logic: can peer observation shift how faculty think about their teaching? *International Journal for Academic Development, 28*(3), 287–300. https://doi.org/10.1080/1360144X.2021.2015691

Yiend, J., Weller, S., & Kinchin, I. (2014). Peer observation of teaching: The interaction between peer review and developmental models of practice. *Journal of Further and Higher Education*, *38*(4), 465–484. https://doi.org/10.1080/0309877X.2012.726967

Zeng, L. M. (2020). Peer review of teaching in higher education: A systematic review of its impact on the professional development of university teachers from the teaching expertise perspective. *Educational Research Review 31*, 100333. https://doi.org/10.1016/j.edurev.2020.100333.