

Developing Peer Review of Instruction in an Online Master Course Model

John Haubrick

Deena Levy

Laura Cruz

The Pennsylvania State University

Abstract

In this study we looked at how participation in a peer-review process for online Statistics courses utilizing a master course model at a major research university affects instructor innovation and instructor presence. We used online, anonymous surveys to collect data from instructors who participated in the peer-review process, and we used descriptive statistics and qualitative analysis to analyze the data. Our findings indicate that space for personal pedagogical agency and innovation is perceived as limited because of the master course model. However, responses indicate that participating in the process was overall appreciated for the sense of community it helped to build. Results of the study highlight the blurred line between formative and summative assessment when using peer review of instruction, and they also suggest that innovation and presence are difficult to assess through short term observation and through a modified version of a tool (i.e., the Quality Matters rubric) intended for the evaluation of an online *course* rather than the *instruction* of that course. The findings also suggest that we may be on the cusp of a second stage for peer review in an online master course model, whether in-person or online. Our findings also affirm the need for creating a sense of community online for the online teaching faculty. The experiences of our faculty suggest that peer review can serve as an integral part of fostering a departmental culture that leads to a host of intangible benefits including trust, reciprocity, belonging, and, indeed, respect.

Keywords: Peer review, online teaching, teaching evaluation, master course model, statistics education, instructor presence

Haubrick, J., Levy, D., & Cruz, L., (2021). Developing peer review of instruction in an online master course model. *Online Learning*, 25(3), 313-328. doi:10.24059/olj.v25i3.2428

Peer review has a long history in academia, originating in the professional societies of the early Enlightenment. The practice first arose to address the need for an evaluation/evaluative metric of the quality of research in an era replete with amateur scientists. In this same context, peer review also functioned as a foundation for establishing collective expertise that was not dependent on the approval of an external body, whether political fiat or divine consecration. The present study examines one way in which this long-standing practice of peer review has evolved to embrace new professional modes (i.e., teaching), new modalities of instruction (i.e., online), and new roles for instructors within the current context of higher education.

Literature Review

Peer review had long been the gold standard for academic research, but it was not until the learning-centered revolution, begun in the 1970s, that the practice found application in education. At first, peer review was confined largely to volunteers who were experimenting with pedagogical changes stemming from recent developments in learning science research. As one leading scholar writes, there was “a general sense...that teaching would benefit from the kinds of collegial exchange and collaboration that faculty seek out as researchers” (Hutchings, 1996). Further, contrary to the conservative bias often attributed to the peer review of research (Roy & Ashburn, 2001), peer review of teaching (PRT) has increasingly proven to foster both personal empowerment and teaching transformation (Chism, 2005; Hutchings, 1996; Lomas & Nicholls, 2005; Smith, 2014; Trautman, 2009). As one set of scholars state, “the value of formative peer assessment is promoted in the exhortative literature...justified in the theoretical literature...and supported by reports of experimental and qualitative research” (Kell & Annetts, 2009; Hyland et al., 2018; Thomas et al., 2014).

Those early experiments led to dramatic breakthroughs in evidence-based practice in teaching and learning and, by extension, changes in how these activities are evaluated. Since the early 2000s, universities have responded to a growing imperative to assess teaching effectiveness, both as a means of evaluating work performance and as a way of demonstrating collective accountability for the student learning experience. An increasing number of studies have linked effective instruction to desired institutional outcomes, including recruitment, persistence, and graduation rates, upon the latter of which many funding models rest. Because the drive towards accountability is fueled by student interests, it is perhaps not surprising that the most common strategy for evaluating teaching are student evaluations of instruction (SETs). At a typical U.S. university today, students are asked to complete an electronic survey at the end of each semester comprised of a series of scaled survey items along with a handful of open-ended questions.

Over the years, the use of SETs as a measure of teaching effectiveness has been both affirmed and disputed (Seldin, 1993). The reliability of the practice has been strengthened through increasing sophistication of both the design of the questions and the analysis of the results. At the same time, however, it has also been questioned as the basis of personnel decisions (Nilson, 2012; Nilson, 2013).

Although not definitively proven, there is a persistent perception that SETs are biased, particularly in the case of faculty members from under-represented populations, including those for whom English is a second language and, in some disciplines, women (Calsamiglia & Loviglio, 2019; Zipser & Mincieli, 2019). Other scholars have called the validity of the results into question, suggesting that students are not always capable of assessing their own learning accurately or appropriately, leading to claims that SETs are more likely to measure popularity

rather than effectiveness (Schneider, 2013; Uttl et al., 2017). Perhaps the only safe and definitive conclusion to draw is that the implications of the practice are complex and contested.

Higher education institutions have navigated these stormy waters in multiple ways, most by encouraging the use of multiple forms of measurement for teaching effectiveness, often in the form of a portfolio, or similar collection tool (Chism, 1999; Seldin et al., 2010). This practice is supported by the research literature, which aligns the practice with the multi-faceted nature of teaching as well as the importance of direct (e.g., not self-reported) measures of student learning. To potentially counterbalance the limitations of SETs, practitioners have suggested the use of PRT, which places disciplinary experts, rather than amateur students, in the driver's seat. In this evaluative mode, PRT typically takes the form of either peer review of instructional materials and/or peer observation of teaching.

While PRT may appear to be a neat solution to a pervasive issue, the practice had previously been used largely for formative purposes on a voluntary basis. The transition to compulsory (or strongly encouraged) evaluative practice has proven to be fraught with dangers, both philosophical and practical (Blackmore, 2005; Edström 2019; Keig, 2006; McManus, 2002). Practically speaking, the PRT process requires a considerable investment of time, energy, and attention, not only to conducting the reviews but also to developing shared standards and practices. Philosophically, several scholars have predicted that several of the primary benefits of PRT as a developmental tool might suffer when transposed into a summative context (Cavanagh, 1996; Gosling, 2002; Kell & Annetts, 2002; Morley, 2003; Peel, 2005). It has proven to be difficult to substantiate these fears, however, as one of the downsides of utilizing summative assessment is the challenges it presents to research.

The PRT problem is confounded by the rise of new modes of instruction, especially online and hybrid modalities (Bennett & Barp, 2008; Jones & Gallen, 2016). Since its inception, online education has carried with it a burden of accountability that traditional in-person instruction has not, and the onus rests with online instructors to prove that the virtual learning experience is of comparable quality to other modalities (Esfijani, 2018; Shelton, 2011). This has, in turn, led to the development and refinement of shared quality standards for online courses (notably, the Quality Matters (QM) rubric), the application and evaluation of which often rely on the collective expertise of other online instructors, i.e., pedagogical (rather than disciplinary) peers (Shattuck et al., 2014). The QM peer-review process, for example, designates two reviewer roles, a subject matter expert and online pedagogy practitioner, the latter of whom undergoes a QM-administered certification process.

The proliferation of online courses, however, has been accompanied by design and implementation changes. Because it takes time and sustained engagement to master the techniques and approaches needed to meet the quality standards for online courses, the role of the instructional designer (ID) as expert in these areas has become increasingly commonplace. A typical role for an ID might be to collaborate closely with faculty members to design and develop online courses that effectively deliver content in a manner that meets (or exceeds) quality standards. Once created, it is certainly possible for the same course to be taught by multiple faculty members.

In a typical ID-faculty scenario, the faculty member often has considerable input on the design as it evolves and provides primary instruction, but peer review of instruction is complicated both by the medium and the role of the third party (the ID) (Drysdale, 2019). For example, the observation protocols developed for the classroom may not apply to a virtual space, at least not to the same degree, and a review of instructional strategies, as reflected in artifacts

such as the syllabus, may be the product of both the ID and/or the faculty member. It is perhaps for these reasons that peer review of online instruction has tended to focus on the course rather than the instructor. The Quality Matters rubric, for example, emphasizes attributes of course design rather than teaching effectiveness. Yet, the need for evaluative measures of instruction and instructor persists, perhaps even more so as trends point to a growing number of adjunct faculty teaching online courses for whom such measures can provide both accountability and professional development. (Barnett, 2019; Taylor, 2017).

The challenge is further compounded by the emergence of instructional standards and/or competencies for online (or hybrid) courses that are distinctive to the virtual environment, both in form and context (Baran et al., 2011). The popular community of inquiry model, for example, differentiates between cognitive presence (content and layout), social presence (engagement), and teaching presence in online courses; all are facets of instruction that are less emphasized in in-person instruction. These insights have led to the development of several exemplary protocols specifically intended for reviewing online instruction (McGahan et al., 2015; Tobin et al., 2015). Each of these tools are firmly grounded in an extensive body of evidence-based practice for online teaching, but still, the handful of studies that have been conducted on the PRT process itself have tended to be limited to case studies and/or action research (Barnard et al., 2015; Swinglehurst et al., 2014; Sharma & Ling, 2018; Wood & Friedel, 2009). As one researcher put it, it is simply “difficult to find quantitative evidence due to its nature and context” (Bell, 2002; Peel, 2002).

The challenge of peer review of teaching is even further complicated by the increasing use of the master course model (Hanbing & Mingzhuo, 2012; Knowles & Kalata, 2007). For courses in which stakes are higher and student populations larger, such as gateway or barrier courses, an institution may choose to adopt a master course model in which an already designed course is provided to all instructors, thereby ensuring a consistent experience for all students (Parscal & Riemer, 2010). In this scenario, instructors have little to no control over the content, design, and, in many cases, delivery of the course, all of which serve as major components of most peer review of instruction models, whether for online or in-person courses. However, even within a master course model, instruction varies and opportunities remain to provide both formative (for individual improvement) and summative (for performance evaluation) feedback. Yet, the question of how to evaluate teaching within these boundaries is a subject that has received less attention in both research and practice. Our study explores the implementation of a peer review of teaching process for an online statistics program that uses master courses at a large, public, research-intensive university.

Methods

Context

The Pennsylvania State University is a public research university located in the northeastern part of the United States. The statistics program offers 24 online courses, with approximately 1500 enrollments per semester, including those for its online graduate program and two undergraduate service courses. Statistics courses have been identified as barrier courses at many institutions, including this one. Therefore, the program at The Pennsylvania State University bears the responsibility for high standards of instructions that contribute to student success, especially persistence.

Each of the program’s 24 courses is based on a master template of objectives, content, and assessments. The courses are delivered through two primary systems, the learning management system (LMS) and the content management system (CMS). Each section has its own unique LMS space for each iteration of the course. Students and instructors use the LMS for announcements, communication/email, assessments, grading, discussion and any other assignments or interactions. The lesson content for each course is delivered through a CMS, which in this case has a public website whose content is classified as open educational resources under a creative commons license. The CMS is unique to the course and is not personalized or changed from semester to semester. Similarly, the lesson content, developed and written by program faculty members, does not change from semester to semester, aside from minor fixes and/or planned revisions.

Instructor agency in the LMS context varies depending on the course taught, how long the instructor has taught it, and how many sections are offered in that semester. Instructors who are teaching a course that has only one section have more agency to change appearance and interactions within the LMS than instructors who are teaching a course with multiple sections. In this statistics department, only one section of most of the online graduate courses is offered per semester, while more than one section of undergraduate courses is typically offered. The largest of these undergraduate courses is a high enrollment, general education requirement course that runs 10-12 sections per semester. Courses with multiple sections use the same CMS as well as the same master template in the LMS to maintain consistency in the student experience. Therefore, in a single section course the instructor could modify the design of their course space within the LMS by choosing their home page, setting the navigation, and organizing the modules while still delivering the content and objectives as defined by the department for that course. Such modifications are less likely to occur in multi-section courses. The following table highlights the level of agency possessed by the instructor in both the CMS and LMS according to the varied teaching contexts in this department.

Table 1
Levels of Instructor Agency in Various Course Types Offered

If the instructor teaches...	Content Management System (CMS)	Learning Management System (LMS)
Undergraduate, single section	Low	High
Graduate, single section	Low	High
Undergraduate, multiple sections	Low	Low
Graduate, multiple sections	Low	Low

During the fall 2019 semester, the faculty members in the department who teach online courses were comprised of full-time teaching professors (n=13), tenure-track professors (n=6), and adjuncts (n=10). Peer review of instruction has been practiced since the onset of the program. In its current iteration, the process takes place annually over an approximately three-week period in the fall semester. The primary purpose of the peer-review process is to offer formative feedback to the instructors, but the results are shared with the assistant program director and faculty members are permitted (though not required) to submit the results as part of their reappointment, promotion, and tenure dossiers. For the fall 2019 semester, 27 of the 29 (93%) faculty members participated in the peer-review process.

Peer Review of Instruction Model

In the fall of 2018, the instructional designer for these statistics courses piloted a new peer-review rubric, which is a modification of the well-known Quality Matters Higher Ed rubric. In this modification, 21 out of 42 review standards were determined to be applicable to the instructors in the master course context. The rubric serves as the centerpiece of a two-part process, in keeping with identified best practices (Edkey, & Roehrich, 2013). First, the faculty member completes a pre-observation survey and the reviewer, who is added to the course as an instructor, evaluates the course according to each of the twenty-one standards in the rubric. The observation is followed by a virtual, synchronous meeting with the peer-review partner. Faculty members are paired across various teaching ranks and course levels, and the pairings are rotated from year to year. Both the observation and the peer meeting are guided by materials created by the instructional designer, who provides both the instructor intake form and two guiding questions for discussion.

In keeping with evidence-based practice for online instruction, the first discussion prompt addresses how the faculty establish social, cognitive, and teaching presence within their course. Along with the prompt, definitions and examples of each type of presence are provided to the instructor.

Discussion prompt 1 in the online statistics program peer-review guide:

Prompt #1: *Share with your peer how you establish these three types of presence in your course.*

Notes: How does your peer establish these three types of presence in their course?

The second prompt provides an opportunity for the instructors to share changes or innovations they have implemented within the past year.

Discussion prompt 2 in the online statistics program peer-review guide:

Prompt #2: *Share with your peer if you are trying anything new this semester (or year)? If yes, share your innovation or change you've made this semester (or year).*

- *Has the innovation or change been successful?*
- *What challenges have you had to work through?*
- *How could others benefit from what you've learned?*
- *What advice would you share with a colleague who is interested in trying this or something similar?*

Notes: What has your peer done this semester (or year) that is innovative or new for them?

The process seeks to evaluate and promote not only quality standards through the rubric, but also collegial discussion around innovation, risk-taking, and instructor presence.

Study Design

The IRB-approved study was originally intended to be a mixed methods study, in which input from participating instructors, collected in the form of a survey, would be supplemented with an analysis of the peer-review artifacts, especially the instructor intake form and the peer-review rubric (which includes the 2 discussion prompts). The instructors provided mixed responses to the requests for use of their identifiable artifacts, which limits their inclusion in the study, but the majority did choose to participate in the anonymous survey (14 out of 27, 54%) which was administered in the Fall semester of 2019. The online survey, sent to instructors by a member of the research team not associated with the statistics department, consisted of 11 questions, comprised of 1 check all that apply, 8 five-point Likert scale, 1 yes/no, and 3 open-ended questions.

Results

Quantitative Results

With the small sample size (n=13) we are limited to basic descriptive statistics to analyze the results of the Likert questions. The most infrequently chosen category on the Likert scale of this survey was “neither agree nor disagree” (n=10), while “somewhat agree” (n=37) was the most frequently chosen. In looking at the responses to specific prompts, we note that the statement with the highest score was *The steps of the peer-review process were clear*. For this statement, 13/13 responded with somewhat agree or strongly agree (mode = “strongly agree”). Consistent with our qualitative findings, the next highest scoring statement was *The peer-review process was collegial*, where 12/13 responded with somewhat agree or strongly agree and one responded as neither agree nor disagree (mode = “strongly agree”). The statement *The peer-review process was beneficial to my teaching* received the third highest rating with 10/13 respondents saying that they somewhat agree (n=7) or strongly agree (n=3) (mode = “somewhat agree”).

We do want to note that consistent with best survey design practice, one of the statements was purposely designed as a negative statement: *The peer-review process was not worth the time spent on doing it*. For this prompt, 8/13 responded with strongly disagree or somewhat disagree, while 3/13 somewhat agreed with that statement and 2 chose neither agree nor disagree (mode = “strongly disagree”).

Qualitative Results

The findings suggest that the participants operated under several constraints. When asked how they assess student learning in the intake form, for example, the majority indicated that the assessments are part of the master class and largely outside of their control, e.g. *All... sections have weekly graded discussion forums (might not be the same question), same HWs and same exams. All instructors contribute for exams and HWs. Assessment of learning outcomes mainly occur through these*. This was evident both in the content and tone of their responses, with passive voice predominating, e.g., *quiz and exam questions are linked to lesson learning objectives*. The presence of constraint also came to the fore in the survey questions about changes; for those who did make changes (6/11), these largely took the form of micro-innovations (e.g., *so far just little things, small modifications*), tweaks primarily focused either on course policies (e.g., *new late policy*); enhancing instructor presence (e.g., *try new introductions; I am using announcements more proactively*) or fostering community (e.g., *increasing discussion board posts, add netiquette statement*).

Space for personal pedagogical agency and innovation is perceived as limited because of the master course model employed in this context. This sentiment is evidenced by the tone of the survey responses related to assessments, and as just discussed. On the other hand, the instructor intake form shows that instructors can innovate and experiment with those course components that can be characterized broadly as relating to instructor presence, particularly regarding communication in the course. There is a marked shift in the tone of response when asked, for example, *Please describe the nature and purpose of the communications between students and instructors in this course.* Responses to this question show agency and active involvement on the part of the instructor in this aspect of the course:

I post announcements regularly and am in constant communication with the class. The discussion forums have a fair bit of chatter and I have replied with video and images as well there with positive feedback.

I respond very quickly to student correspondence. I use the course announcements feature very often and check Canvas multiple times a day.

I would like to promote the use of the Discussion Boards more, but students still do not use those as much as I would like them to.

In this last example, we see that the instructor is forward looking and discusses changes that he or she would like to make even in the future. The data suggest that instructors are trying to make space for their own unique contribution to the course and for more personalized choices in their interactions with students. They are also eager to get feedback from their peers on practices that fall into this space of agency:

I would appreciate any feedback on my use of course announcements. Do you feel that they are appropriate in both content, frequency, and timing?

Our findings indicate that many of these instructors are operating within the constraints of a master course model, as discussed earlier, and they are most enthusiastic in their responses and innovative in their teaching when they can identify areas over which they can exert some degree of control in the course design and delivery process.

As evidenced in the quantitative findings previously discussed, these qualitative findings also tell us that instructors who participated in the survey appreciate the collegiality of the process. Their open-ended responses indicate an appreciation of the collegiality and connection, the informal learning, that the peer-review process afforded them. For example, one instructor comments, “I have enjoyed the opportunity to discuss teaching ideas and strategies with other online faculty. As a remote faculty member, I particularly value that interaction.” Responses primarily indicate that participating in the process was overall appreciated for the sense of community it helped to build. What we see emerge is another space—a space where instructors can negotiate together the limitations for innovation that exist in this sequence of Statistics courses, and where they can also share experiences. As one participant comments, *The direct communication with the peer is great for sharing positive and negative experiences with different courses.* As we see in our findings, faculty members clearly find value in the process, regardless of the product. This insight suggests the presence of a lesser known third model, distinct from

either formative or evaluative formats, called collaborative PRT (Gosling, 2002; Keig & Waggoner, 1995). In collaborative PRT, the end goal is to capture the benefits of turning teaching from a public to a more collaborative activity (Hutchings, 1996).

Discussion

Our findings should not be overstated. This study was conducted for a single program at a single university over the course of one semester; as such, the results may or may not be replicable elsewhere. Replication may also be hindered by the challenges inherent in studying peer review as a process. Because the results of peer review in this case may be used for summative or evaluative purposes, any evidence generated is considered part of a personnel file and, as such, subject to higher degrees of oversight in the ethical review process. The ethical review board at The Pennsylvania State University, for example, did not classify this study as exempt research, but rather put the proposal through full (rather than expedited or exempt) board review, and has required additional accountability measures. And the evaluative nature of those documents also contributed to low faculty participation (n=3) in the first stage of our study, where we asked to include copies of their peer-review documents (an intake form, review rubric, and meeting notes). There is a reason why there are comparatively few studies on peer review as a process.

In the case of the statistics program, the primary rationale for establishing a peer review of teaching process was intended to be formative assessment, i.e., providing feedback to instructors so that they might improve the teaching and learning in online statistics courses. In practice, however, the boundaries between formative and summative assessment blurred. While instructors were not required or compelled to disclose the results of their peer review, many did choose to include comments and/or ratings in their formal appointment portfolios, especially when the only other evidence of teaching effectiveness (a primary criterion) available are student evaluations of instruction (SETs). At The Pennsylvania State University, SETs are structured so that students provide feedback on both the instructor and the course, at times separately and, at other times, together. In a master course model, however, instructors have limited control over many components of the course, making the results of student evaluations challenging to parse out and potentially misleading if treated nominally or comparatively.

The distinction between formative and evaluative assessment is not the only blurred line that arose from this study. In this case, peer review of *instruction* was accomplished with a modified version of a tool (the QM rubric) intended to be used for the evaluation of an online *course*. The modification of the QM rubric took the form of removing questions or sections pertaining to course components deemed to be outside the control of the master course instructors. In addition to the modified QM rubric, two supplemental items—open-ended questions—were added to the review process. These items focused on presence and innovation, which are difficult to assess through short-term observation. Our results suggest that this strategy has led to partial success, i.e., the majority (10/13) of faculty members who responded to our survey strongly or somewhat agreed that the process was beneficial, but its impact on teaching practice has been limited. This may be partially a result of the limited scope of the study (one academic year) which may or may not be an appropriate time frame for capturing changes to teaching practice, but it may also stem from limitations in the current iteration of the peer-review process itself.

If we look back over the history of peer review of instruction for online courses, a pattern emerges in which first, an existing tool, developed for a different purpose or context, is importuned and adapted into a new environment. This occurred, for example, when peer evaluation tools designed for in-person courses were adapted to suit online courses. In the next stage, the adaption process reveals limitations of the existing tool which, in turn, spur the development of new instruments or processes that are specifically designed for the context in which they are being used. The creation of the QM Rubric is a clear example of this latter step.

The findings of our study suggest that we may be on the cusp of this second stage for peer review of teaching in online master courses, which constitutes a quite different teaching environment than other types of courses, whether in-person or online. In the case of master courses, there is a distinctive division of labor where, primarily, instructional designers work with authors to develop courses, course leads manage content, and instructors serve as the primary point of contact with students. It may be time to develop a new rubric (or similar tool) that takes this increasingly popular configuration more into consideration.

Adoption of the master course model is fueled by the need for both efficiency and consistency in the student learning experience, and both experience and research suggest that it has been effective in serving these goals. That being said, like all models, it also has its limitations. Our study suggests that one of those tradeoffs may be that the model constricts both the space for and the drivers of change. Without being able to make changes to the master course itself, the faculty in our study tried to find ways to make small changes, i.e., micro-improvements in those areas over which they held agency. Larger or more long-term changes, on the other hand, would need to come from instructional designers and program managers, who may be one or even two steps removed from the direct student experience. Although instructors frequently make suggestions for course improvements, large changes to courses are not frequently implemented. In other words, the division of labor needed to support the master course model also divides agency, and the challenge remains to find systematic ways to re-integrate that agency in the service of continuous improvement.

The limitations on faculty agency inherent in the master course model have led some institutions to further devalue the role, substituting faculty-led courses for lower-paid, lesser recognized, and more easily inter-changeable instructor roles (Barnett, 2019). Such a path would be at odds with the culture of The Pennsylvania State University, but it does suggest the need for faculty development, i.e., for finding ways to support and treat even part-time instructors as valued and recognized members of the community of teaching and learning, even in conditions where they may not be able to meet in person. It could be said that our findings affirm both the need for creating a sense of community online both inside and outside of the courses, for faculty members who teach them. The experiences of our faculty members suggest that peer review can be an integral part of departmental culture that supports faculty peer to peer engagement, leading to a host of intangible benefits including trust, reciprocity, belonging, and, indeed, respect.

References

- Baran, E., Correia, A. P., & Thompson, A. (2011). Transforming online teaching practice: Critical analysis of the literature on the roles and competencies of online teachers. *Distance Education, 32*(3), 421-439.
- Barnard, A., Nash, R., McEvoy, K., Shannon, S., Waters, C., Rochester, S., & Bolt, S. (2015). LeAD-in: a cultural change model for peer review of teaching in higher education. *Higher Education Research & Development, 34*(1), 30-44.
- Barnett, D. E. (2019). Full-range leadership as a predictor of extra effort in online higher education: The mediating effect of job satisfaction. *Journal of Leadership Education, 18*(1).
- Bennett, S., & Barp, D. (2008). Peer observation—a case for doing it online. *Teaching in Higher Education, 13*(5), 559-570.
- Blackmore, J. A. (2005). A critical evaluation of peer review via teaching observation within higher education. *International Journal of Educational Management, 19*(3), 218-232.
- Calsamiglia, C., & Loviglio, A. (2019). Grading on a curve: When having good peers is not good. *Economics of Education Review, 73*(C).
- Cavanagh, R. R. (1996). Formative and summative evaluation in the faculty peer review of teaching. *Innovative higher education, 20*(4), 235-240.
- Chism, N. V. N. (1999). Peer review of teaching. A sourcebook. Bolton, MA: Anker.
- Drysdale, J. (2019). The collaborative mapping model: Relationship-centered instructional design for higher education. *Online Learning, 23*(3), 56-71.
- Edkey, M. T. & Roehrich, H. (2013). A faculty observation model for online instructors: Observing faculty members in the online classroom. *Online Journal of Distance Learning Administration, 16* (2).
http://www.westga.edu/~distance/ojdla/summer162/eskey_roehrich162.html
- Edström, K., Levander, S., Engström, J., & Geschwind, L. (2019). Peer review of teaching merits in academic career systems: A comparative study. In *Research in Engineering Education Symposium*.
- Esfijani, A. (2018). Measuring quality in online education: A meta-synthesis. *American Journal of Distance Education, 32*(1), 57-73.
- Gosling, D. (2002). Models of peer observation of teaching. Report. LTSN Generic Center.
https://www.researchgate.net/profile/David_Gosling/publication/267687499_Models_of_Peer_Observation_of_Teaching/links/545b64810cf249070a7955d3.pdf

- Graham, C., Cagiltay, K., Lim, B. R., Craner, J., & Duffy, T. M. (2001). Seven principles of effective teaching: A practical lens for evaluating online courses. *The Technology Source*, 30(5), 50.
- Hanbing, Y., & Mingzhuo, L. (2012). Research on master-teachers' management model in online course by integrating learning support. *Journal of Distance Education*, 5(10), 63-67.
- Hutchings, P. (1996). Making teaching community property: A menu for peer collaboration and peer review. AAHE Teaching Initiative.
- Hutchings, P. (1996). The peer review of teaching: Progress, issues and prospects. *Innovative Higher Education*, 20(4), 221-234.
- Hyland, K. M., Dhaliwal, G., Goldberg, A. N., Chen, L. M., Land, K., & Wamsley, M. (2018). Peer review of teaching: Insights from a 10-year experience. *Medical Science Educator*, 28(4), 675-681.
- Johnson, G., Rosenberger, J., & Chow, M. (October 2014) The importance of setting the stage: Maximizing the benefits of peer review of teaching. *eLearn*, 2014 (10).
<https://doi.org/10.1145/2675056.2673801>
- Jones, M. H., & Gallen, A. M. (2016). Peer observation, feedback and reflection for development of practice in synchronous online teaching. *Innovations in Education and Teaching International*, 53(6), 616-626.
- Keig, L. (2000). Formative peer review of teaching: Attitudes of faculty at liberal arts colleges toward colleague assessment. *Journal of Personnel Evaluation in Education*, 14(1), 67-87.
- Keig, L. W., & Waggoner, M. D. (1995). Peer review of teaching: Improving college instruction through formative assessment. *Journal on Excellence in College Teaching*, 6(3), 51-83.
- Kell, C., & Annetts, S. (2009). Peer review of teaching embedded practice or policy-holding complacency? *Innovations in Education and Teaching International*, 46(1), 61-70.
- Knowles, E., & Kalata, K. (2007). A model for enhancing online course development. *Innovate: Journal of Online Education*, 4(2).
- Lomas, L., & Nicholls, G. (2005). Enhancing teaching quality through peer review of teaching. *Quality in Higher Education*, 11(2), 137-149.
- Mayes, R. (2011, March). Themes and strategies for transformative online instruction: A review of literature. In *Global Learn* (pp. 2121-2130). Association for the Advancement of Computing in Education (AACE).
- McGahan, S. J., Jackson, C. M., & Premer, K. (2015). Online course quality assurance: Development of a quality checklist. *InSight: A Journal of Scholarly Teaching*, 10, 126-140.

- McManus, D. A. (2001). The two paradigms of education and the peer review of teaching. *Journal of Geoscience Education*, 49(5), 423-434.
- Nilson, L. B. (2012). 14: Time to raise questions about student ratings. *To improve the academy*, 31(1), 213-227.
- Nilson, L. B. (2013). 17: Measuring student learning to document faculty teaching effectiveness. *To Improve the Academy*, 32(1), 287-300.
- Nogueira, I. C., Gonçalves, D., & Silva, C. V. (2016). Inducing supervision practices among peers in a community of practice. *Journal for Educators, Teachers and Trainers*, 7, 108-119.
- Parscal, T., & Riemer, D. (2010). Assuring quality in large-scale online course development. *Online Journal of Distance Learning Administration*, 13(2).
- Peel, D. (2005). Peer observation as a transformatory tool? *Teaching in Higher Education*, 10(4), 489 - 504.
- Roy, R., & Ashburn, J. R. (2001). The perils of peer review. *Nature*, 414(6862), 393-394.
- Schneider, G. (2013, March). Student evaluations, grade inflation and pluralistic teaching: Moving from customer satisfaction to student learning and critical thinking. In *Forum for Social Economics* 42(1),122-135.
- Seldin, P. (1993). The use and abuse of student ratings of professors. *Chronicle of Higher Education*, 39(46), A40-A40.
- Seldin, P., Miller, J. E., & Seldin, C. A. (2010). *The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions*. John Wiley & Sons.
- 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.ms: a comparative study. In *Research in Engineering Education Symposium*.
- Shattuck, K., Zimmerman, W. A., & Adair, D. (2014). Continuous improvement of the QM Rubric and review processes: Scholarship of integration and application. *Internet Learning Journal*, 3(1).
- Shelton, K. (2011). A review of paradigms for evaluating the quality of online education programs. *Online Journal of Distance Learning Administration*,4(1), 1-11.
- Smith, S. L. (2014). Peer collaboration: Improving teaching through comprehensive peer review. *To Improve the Academy*, 33(1), 94-112.
- Swinglehurst, D., Russell, J., & Greenhalgh, T. (2008). Peer observation of teaching in the online environment: an action research approach. *Journal of Computer Assisted Learning*, 24, 383-393.

Taylor, A. H. (2017). *Intrinsic and extrinsic motivators that attract and retain part-time online teaching faculty at Penn State* (Doctoral dissertation, The Pennsylvania State University).

Thomas, S., Chie, Q. T., Abraham, M., Jalarajan Raj, S., & Beh, L. S. (2014). A qualitative review of literature on peer review of teaching in higher education: An application of the SWOT framework. *Review of Educational Research, 84*(1), 112-159.

Tobin, T. J., Mandernach, B. J., & Taylor, A. H. (2015). *Evaluating online teaching: Implementing best practices*. San Francisco, CA: John Wiley & Sons.

Trautmann, N. M. (2009). Designing peer review for pedagogical success. *Journal of College Science Teaching, 38*(4).

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22-42.

Wood, D., & Friedel, M. (2009). Peer review of online learning and teaching: Harnessing collective intelligence to address emerging challenges. *Australasian Journal of Educational Technology, 25*(1).

Zipser, N., & Mincieli, L. (2018). Administrative and structural changes in student evaluations of teaching and their effects on overall instructor scores. *Assessment & Evaluation in Higher Education, 43*(6), 995-1008.

Appendix A

Anonymous Survey Questions

Likert Questions [1-8]

Answer Options

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

1. The peer-review process was beneficial to my teaching.
2. The peer-review process was beneficial to my career development.
3. The peer-review process was not worth the time spent on doing it.
4. The peer-review process was collegial.
5. The peer-review process provided me with new insight into my teaching practice.
6. The peer-review process inspired me to try new things related to my teaching.
7. The steps of the peer-review process were clear.
8. I have little to no prior experience with peer review of online teaching.

Open-ended Questions [9-11]

9. Did you make (or do you plan to make) changes to your instruction based on your participation in this peer-review process (e.g. feedback you received, conversations with your peers, rubrics, etc...)? Y/N
 - a. If Y, please describe the change(s) you plan to make to your instruction based on the feedback you received through the peer-review process.
10. Please describe at least two insights gained from your participation in the peer-review process.
11. What changes, if any, would you suggest should be made to enhance the benefits of the peer-review process?

Appendix B

Instructor Intake Form Questions

Your information

1. What is your name?
2. What is your e-mail address?
3. Who is your assigned peer reviewer?

Your Online Course

4. What is your course name, number & section (e.g., STAT 500 001)?
5. What is the title of your course (e.g., Applied Statistics)?
6. What is the Canvas link to your course?
7. What is the link to the online notes in your course?

Context

8. How many semesters have you taught this course? Choose: (0-3) (4-6) (6 or more)
9. Does your course have multiple sections?
10. If yes, are all sections based on a single master (or another instructor's) course?
11. If yes, roughly what percentage of the course do you change or personalize from the master?
12. How do you know if students are meeting the learning outcomes of your course?
13. Is there a specific part of the course content or design for which you would like the reviewer to provide feedback?
14. Please describe the nature and purpose of the communications between students and instructors in this course.
15. Are you trying anything new this semester based on prior student or peer feedback, professional development, or your own experiences?
16. If yes, please explain.

Canvas Communication

17. Please identify **other communications among students and instructors** about which the Reviewer should be aware, but which are not available for review at the sites listed above.
18. Does the course require any **synchronous activities** (same time, same place)?
 Yes No
19. If yes, please describe: _____
20. Is there any other information you would like to share with your peer before they review your course?