Common Practices for Evaluating Post-Secondary Online Instructors

Jonathan E. Thomas
Brigham Young University
jonathanethomas2002@gmail.com

Charles R. Graham
Brigham Young University
charles.graham@byu.edu

Abstract

This literature review explores current post-secondary practices for evaluating online instructors. As enrollment of students in online courses has steadily increased over the last few decades, instructor evaluation has lagged behind. Through a thematic analysis of existing literature, this review seeks to answer these questions: (1) How are online instructors evaluated? (2) When and why are online instructors evaluated? (3) What are institutions evaluating? This review reveals that many unresolved problems still exist among online instructor evaluations. One of the more significant problems raised in the research is whether evaluation instruments used to evaluate traditional face-to-face instructors are appropriate to evaluate online instructors. Another significant finding of this review is that current practices of post-secondary institutions evaluate instructors based on course design. These and other findings indicate that evaluation of online instructors is a field that requires additional research.

Introduction

Online learning is meeting a legitimate need for students, as enrollment in online courses at universities and colleges continues to grow. A recent report based on data collected by the National Center for Education Statistics' Integrated Postsecondary Education Data System (IPEDS) found that more than 28% of all enrolled students in 2014 were taking at least one course online (Pearson, 2016). That means that more than 5.8 million students at post-secondary institutions are enrolling from online education. These results confirm that there continues to be a steady increase in student enrollment in online courses. Online learning has become a permanent fixture of the post-secondary education landscape and will likely continue as an essential way to provide education to millions of students.

This rapid growth of online learning requires careful measures to ensure that courses are designed and facilitated according to quality standards. Evaluation is a critical component of education to ensure these careful measures. Unfortunately, several studies have alarmingly pointed out that the systematic evaluation of online courses and instructors is surprisingly underdeveloped considering the rapid growth of online education (Berk, 2013; Rothman, Romeo, Brennan, & Mitchell, 2011). This indicates that post-secondary institutions are grappling with how to address online instructor evaluation.

To date, there is not a comprehensive review of literature that addresses how post-secondary institutions are addressing online instructor evaluation. Therefore, the purpose of this literature review is address this problem by answering these three questions: (1) How are post-secondary online instructors evaluated? (2) When and why are institutions evaluating their instructors?, and (3) What are they evaluating?

Methodology

Utilizing the thesaurus in ERIC, we identified terms related to online learning (e.g. virtual universities, asynchronous communication, online courses, virtual classrooms, web based instruction), which I coupled in my database search with either faculty evaluation or teacher evaluation, (both ERIC thesaurus items). I initially limited the search to more recent articles published in the last decade. The search returned 51 results.

Exclusion/Inclusion Criteria

I excluded any articles of the initial search from the final analysis that did not directly address online instructor evaluation in post-secondary institutions. I examined the reference lists of the remaining articles and identified additional, relevant studies that were not included in the initial search. These were also added. This included some articles that were outside the original search parameters. As a result of these criteria, the final analysis included 43 articles.

Analysis

In coding and analyzing these articles, I utilized the method of thematic analysis as outlined by Braune and Clarke (2006). I allowed the research questions to drive the data collection. As I examined each article, I looked for descriptions of evaluation practices and coded them based on which of the following questions they addressed: (1) How are online instructors evaluated? (2) When and why are online instructors evaluated? (3) What are institutions evaluating? The results are included in Table 1.

Table 1.
Literature Review Results

Research Question	Pertinent Research
How are online instructors evaluated?	*Student Evaluations:* Bangert, 2008; Bangert, 2004; Benton & Cashin, 2012; Darling, 2012; Delaney, Johnson, Johnson, & Treslan, 2010; Drouin, 2012; Dziuban & Moskal, 2011; Flynn, Maiden, Smith, Wiley, 2013; Loveland, 2007; Moore, 2014; Moskal, Dziuban, & Hartman, 2013; Piña & Bohn, 2014; Rhea, Rovai, Ponton, Derrick, & Davis 2007; Roberts, Irani, Telg, & Lundy, 2005; Rothman, Romeo, Brennan, & Mitchell, 2011; Stanišic´ Stojic´, Dobrijevic´, Stanišic´, & Stanic´, 2014; Terry, 2007. *Administrative Evaluation:* Dana, Havens, Hochanadel, & Phillips, 2010; Tobin, 2004; Weschke & Canipe, 2010. *Peer Evaluation:* Berk, 2005; Colleges, 2013; Cordeiro & Muraoka, 2015; DeCosta, Bergquist, & Holbeck, 2015; Drouin, 2012; Hathorn & Hathorn, 2010; Mandernach, Donnelli, Dailey, & Schulte, 2005; Palloff & Pratt, 2008; Piña et al., 2014; Schulte, 2009; Schulte, Dennis, Eskey, Taylor, & Zeng, 2012. *Self Evaluation:* Berk, 2005; Delaney et al., 2010; Drouin, 2012; Kennedy, 2015; Piña et al., 2014; Schulte et al., 2012; Weschke et al., 2010.
When and why are online instructors evaluated?	*End of Course:* Dziuban et. al., 2011 Palloff et al., 2008; Piña et al., 2014; Schulte et al., 2012. *During the Course:* Colleges, 2013; Dana et al., 2010; DeCosta et al., 2015; Palloff et al., 2008; Mandernach et al., 2005; Schulte et al., 2012; Tinoca & Oliveira, 2013; Weschke et al., 2010. *To Promote Professional Development:* Colleges, 2013; Dana et al., 2010; DeCosta et al., 2015; Mandernach et al., 2005; Palloff et al., 2008; Schulte et al., 2012; Tinoca et al., 2013; Weschke et al., 2010. *To Inform Administrative Decisions:* Colleges, 2013; Darling, 2012; Donovan, 2006; Dziuban et. al., 2011; Roberts et al., 2005; Stanišic´ Stojic´ et al., 2014.
What are institutions evaluating?	*Course Design:* Bangert, 2004; Drouin, 2012; Roberts et al., 2005; Rothman et al., 2011; Stewart, Hong, & Strudler, 2004; Ternus, Palmer, & Faulk, 2007. *Instructor Competencies:* Coll, Rochera, Gispert, Díaz-Barriga, 2013; Darabi, Sikorski, & Harvey 2006; Eskey & Schulte, 2012; Gaytan & McEwen, 2007; Gorskey & Blau, 2009; Graham, Cagiltay, Lim, Craner, & Duffy, 2001; Moore, 2014; Nandi, Hamilton, & Harland, 2012; Schulte et al., 2012; Weschke et al., 2010.

Discussion

In this section we discuss some of the results and general findings of our research that can inform better practices in evaluating online instructors.

How Are Online Instructors Evaluated?

The evaluation of online instructors in some cases follows a similar system as that which is utilized in traditional courses. This includes several different evaluation measures to provide different information and perspectives on the effectiveness of an online instructor. This includes evaluations performed by students, administrators, peers, and the instructor.

Student evaluations. Student evaluations of instructors are the most common form of evaluation in online higher education courses. Although a few institutions have not yet established procedures to evaluate their online instructors, the majority at least perform student evaluations, if nothing else (Delaney et al., 2010; Piña et al., 2014). For many years, performing student evaluations of online instructors was overlooked (Darling, 2012). This may have been due to other administrative demands that were far more urgent such as designing, staffing, and maintaining online courses that were in high demand. Before long, most institutions recognized the need for students' perspectives and determined to collect this information in online courses.

As administrators began developing processes for performing student evaluations, some utilized existing student evaluations of instructors in traditional, face-to-face courses (Colleges, 2013; Cordeiro et al., 2015; Drouin, 2012). They assumed that there was little difference between the competencies and skills necessary to be an effective face-to-face instructor and those necessary to be an effective online instructor. A review of research done by Benton and Cashin (2012) specific to student evaluations in both traditional and online courses concluded that there is little difference between the two. It is important to note, however, their conclusions were limited to specific aspects of course design (learning objectives, teaching methods, etc.) as opposed to behaviors specifically associated with the instructor.

A study by Loveland (2007) called into question the conclusions that student evaluation instruments can be the same regardless of whether they are used in face-to-face or online courses. Loveland (2007) made minor adjustments to a Student Evaluation of Teaching (SET) that had been widely tested and deemed as a valid and reliable instrument to evaluate instructors in a face-to-face classroom and used it to evaluate online instructors. These minor adjustments included changing things from "oral" communication skills to ask about "written" communication skills. After using the instrument in online courses, they grouped the results from the 18 items of the instrument into five global variables and utilized linear regression to determine if the five variables were accurately represented by the 18 independent variables. They found that many of the measures are statistically significant and accurately describe the variation in the global variables. Of the 18 measures, however, three of the items did not fit within the five global variables.

Two of these (clarity of course objectives and clarity of student responsibilities and requirements) were not statistically significant in any of the models. In other words, these two items were not related to the students' views of instructor effectiveness in online courses. Another interesting finding was that "user friendliness of course materials" actually had a negative effect on student evaluations. The higher a student evaluated the user friendliness of course materials, the lower the rating of the course and instructor tended to be. These findings suggest that there may be aspects of traditional student evaluations that either evaluate things that are irrelevant to online courses or may fail to evaluate things that are relevant to online courses.

A significant study by Dziuban et. al. (2011) countered the assumption that traditional face-to-face student evaluation instruments do not adequately measure online instructor effectiveness. Using a large data set of over one million student responses to an end-of-course evaluation, they found that their student evaluation instrument is measuring the same aspects of instruction, regardless of modality. Additional research from Moskal et. al. (2013) found that regardless of the course modality, "if the instructor facilitates learning, communicates well, and respects his or her students then [the instructor] will be rated excellent" (p. 19). Both of these studies utilized the student evaluation instrument developed at the University of Central Florida. The instrument is included in the appendix of both articles. It is important to note that this particular instrument was designed to address face-to-face, online, and blended modalities and so treats each question in a general way. It may be that the instrument Loveland used (2007) was designed primarily for face-to-face courses and may have been too specific to that particular modality to be useful for other modalities.

It is critical that student evaluation instruments accurately address instructor effectiveness, regardless of modality. If SET instruments designed for traditional classrooms fail to accurately measure teaching effectiveness in online courses, online instructors may receive ratings that are inaccurate measures of their teaching effectiveness.

Loveland (2007) drew attention to this possibility. She reported that student evaluation scores for the online instructors she studied were 20% lower than the scores of instructors in a traditional course. She also reported that instructors who teach the same class in both a traditional format as well as an online format receive lower scores from their online students, sometimes a full point lower on a scale of six.

Lower student evaluation scores for online instructors is not an uncommon finding. Terry (2007) also found in his study comparing data collected from traditional, online, and blended formats. In this study, the sample consisted of MBA students, broken down into three groups: 366 in traditional courses, 312 in online courses, and 198 in blended courses. The instructors sought to ensure that content and course requirements were as consistent as possible across the different mediums. Of the three mediums, online instructors had the lowest faculty and course evaluation scores. It appears that they utilized the same evaluation instrument, regardless of medium, which may have negatively affected the online evaluation scores. This may be related to what Rhea, Rovai, Ponton, Derrick, and Davis (2007) found when they discovered that online students tend to provide feedback that is far more negative than face-to-face students.

There are a variety of factors that may have caused the drop in SET scores as reported by Terry (2007), Loveland (2007). It is probable that many instructors that transition from teaching face-to-face struggle to adjust to the new modality and are initially less effective in an online setting than they are in a traditional classroom. Stoji?, Dobrijevi?, Staniši?, and Stani? (2014), who also found a similar drop in online instructor ratings compared to traditional instructors, suggested that the drop may be due to infrequent interaction between students and their online instructors. This dissatisfaction could be because students rarely interacted with their online instructor. It is easier to be critical of someone, with whom we do not have a relationship. Other possibilities include that the online courses may have been poorly designed or that instructors struggled to interact in meaningful ways with students in online courses. Terry (2007) attributed the lower scores to the fact that online students also had a lower average grade than other mediums and that the lower levels of interaction in online courses may have contributed to the lower rating scores. These lower scores provide some evidence that the instruction of the course failed, either in the design or the facilitation of the course. By failing to disambiguate evaluation of the course design from the instructor, it is difficult to know what really needs to be improved.

Closer scrutiny of an online instructor's roles reveals that there are significant differences between what makes a face-to-face instructor effective and what makes an online instructor effective (Darabi et al., 2006; Tallent-Runnels, Cooper, Lan, Thomas, & Busby, 2005). Berk (2013) argued that because of this, face-to-face evaluations miss items that are unique to online courses. He proposed seven different approaches to identifying or creating a student evaluation instrument that an institution might use to evaluate online instructors. He determined that the most efficient and cost-effective recommendation was to add several items to the traditional face-to-face instrument that address unique aspects of an online classroom. His recommendation, however, falls short of identifying what these items should be, leaving it up to individual institutions to decide in ways that meet their specific needs.

Where many feel that a student evaluation instrument used in face-to-face courses could be adapted to meet the unique circumstances of online courses, others feel that an entirely new instrument ought to be created and adopted (Bangert, 2008; Roberts et al., 2005; Rothman et al., 2011; Stewart et al., 2004). In these cases, researchers systematically developed new instruments to evaluate online teaching effectiveness. Each of these instruments will be discussed in greater detail later. These researchers recognized that in an online course, instruction is largely encapsulated in the course design and can happen independent from the teacher.

Although student evaluations are helpful in evaluating online instructors, depending on them as the only measure of online teaching effectiveness is problematic (Shao, Anderson, & Newsome, 2007). Moreover, many feel that students are ill-equipped to evaluate an instructor's teaching effectiveness (Darling, 2012). This underscores the importance of incorporating other measures of teaching effectiveness into an institution's evaluation procedures.

Administrative evaluations. Administrative evaluation is another piece of the evaluation process for post-secondary institutions. Tobin (2004) explained that many administrators have never taught an online course before and consequently suspect that they can approach the evaluation similar to a traditional course. However, this assumption may result in an inaccurate assessment of instructor effectiveness. Tobin (2004) lists several questions that administrators may have about instructor evaluations including the following:

- "How do I 'visit' the classroom for a set period of time if the classroom is asynchronous?
- "What should I look at to prepare myself for the discussions that the class will have?
- "How can I evaluate the instructor's classroom presence in an online course?
- "In order to say I have evaluated the instructor, where should I visit in the course shell, how often, and why?

- "How can I ascertain the quality of the class discussion and whether the instructor is taking enough of a part?
- "Should the online instructor use more multimedia than a classroom instructor?
- "A lot of the questions from my classroom-visit rubric don't seem to apply. What questions are cognate?
- "How can I evaluate an online course if I've never taught online, myself?" (Tobin, 2004, p. 1).

In order to address these questions, Tobin identifies principles outlined by Graham, Cagiltay, Lim, Craner, and Duffy (2001) to guide best practice in online instruction. These principles are based on Chickering and Gamson's principles that guide best practice in traditional courses. Additionally, he identified several instruments that could guide administrators in their evaluations that follow the principles delineated by Graham et al. Of these, he recommended the Checklist for Online Interactive Learning (Sunal, Sunal, Odell, & Sundberg, 2003), primarily because of its objective nature.

It is interesting to note that although Tobin (2004) recommended a different approach to administrative evaluations of online instructors, he stressed that it is unnecessary to create a new instrument for student evaluations. This argument is based on the assumption that if the outcomes of a course are the same regardless of modality, then the way that students evaluate their instructor should be no different. It stands to reason that if online courses require a different approach to administrative evaluation, that it would also demand a change in approach to other forms of evaluation, including those performed by students.

Some researchers have expressed concerns about the current practices of administrative evaluations. Weschke and Canipe (2010) describe a model at Walden University where they seek to make evaluations less punitive and more helpful for instructors. They explain that "program administrators attempt to be assistant problem-solvers rather than a cudgel bearer to 'beat' faculty into compliance" (2010, p. 46). Faculty are informed that administrators may or may not inform them about their course visit, but are also assured that these visits will be primarily informal and formative. In addition to the written evaluation, instructors will also receive a follow-up phone call to ensure that the information is clear and to review it together. This can help avoid misunderstandings that may result from the evaluation.

Dana, Havens, Hochanadel, and Phillips (2010) described administrative evaluations as "faculty coaching" where department chairs perform an evaluation once per term for each faculty member seeking to help instructors improve (p. 29). Dana et al. expressed concern that all communication as part of these administrative evaluations was static and communicated via text. They recommended the use of screen recording technology as administrators visited online courses to overcome communication barriers that tend to exist when communicating solely through text. Communication can be facilitated by also being able to see body language and hear voice inflections. By doing this, they argued that administrators could strengthen relationships with remote faculty, and consequently provide coaching that would be more widely accepted.

Peer evaluations. Another form of evaluation that some institutions utilize are peer evaluations. Peer evaluation coupled with student evaluations provides complementary evidence of teaching effectiveness (Berk, 2005; Palloff, 2008; Hathorn & Hathorn, 2010). Unfortunately, not many post-secondary institutions perform them (Piña et al., 2014).

The Academic Senate for California Community Colleges is an organization that represents all community colleges in California. They seek to uphold quality standards of the education that they provide. Among these standards they affirm that all of their online courses need to abide by the same standard as their traditional courses. Having regular peer reviews is included as one of these standards. However the Academic Senate for California Community Colleges explains that "due to such issues as the interaction with students through technology and the opportunity for direct observation of the instructor's performance, many colleges have established different or supplemental processes for the evaluation of faculty who teach online" (Colleges, 2013, p. 9). Highlighting these issues seems to suggest that there is still some confusion about how exactly to perform peer evaluations in an online environment, how often peer evaluations should happen, and what they should entail, but also indicates that, in some form, they should be occurring.

Palloff and Pratt (2008) recommend peer evaluation as a way to encourage professional development. They even suggest that a new instructor should have a mentor that provides regular formative evaluation, or an ongoing discussion of successes and areas of improvement. This can shift the emphasis of a peer review from monitoring an instructor's behavior to actively helping each other improve. These kinds of peer reviews are rarely used to inform administrative decisions, but certainly help an instructor to become more effective at facilitating a course.

Cordeiro and Muraoka (2015) describe peer evaluations as a "classroom visitation" (p. 6). The observations they

discuss take place once a year and include a two-hour visit to the course, where the evaluator will visit discussion boards or other online communication tools to evaluate an instructors interactions with students. The idea of a classroom observation can help provide a snapshot of an instructor's facilitation skills. However, doing this once a year for a two-hour period may provide an inaccurate picture of an instructor's effectiveness.

Mandernach, Donnelli, Dailey, and Schulte (2005) described an innovative approach to online instructor evaluation called the Online Instructor Evaluation System (OIES). Each instructor receives five formative evaluations, one of which happens before the semester begins. These evaluations are performed by another faculty member with experience in online teaching. The reports from each visit are not reported to the academic department unless there are patterns of behavior, either positive or negative, about which the administrators should be informed. The objective for these visits is to either begin or continue a conversation on professional development and improvement. These discussions are meant to be collaborative between the evaluator and the instructor that is evaluated. Mandernach et al. explained, "The low-stakes formative assessments promoted dialogue and sharing of best practices among instructor and evaluator as peers" (2005, p. 5). The evaluator acts more as a counselor that may ask questions or provide suggestions. The instructor, likewise, asks questions and proposes solutions. Plans are then made for follow-up visits and discussions.

This system requires additional faculty that devote half of their teaching load to evaluating other instructors. This is a luxury that is not widely available (Piña & Bohn, 2014). Research from Park University has sought to improve their process by implementing a "Quick Check" (Schulte, 2009, p. 110) evaluation where an evaluator checked mid-week to see two questions related to instructor standards:

1. Are they posting in discussion boards at least three days per week?
2. Are they providing timely feedback and grades on student assignments?

By implementing the "Quick Check" evaluation, they found that 70% of the instructors involved in the sample (n=57) improved in these particular aspects of instructor behaviors over the course of two semesters (Schulte, 2009). With continued efforts to revise and improve their system, they have established identified best practices that guide their system from both a review of literature and their own experience. They seek to develop effective online instructors by (1) Encouraging community in the classroom by posting introductions, (2) Establishing strong instructor presence in the course, especially on discussion boards, (3) Providing clear and individualized feedback to students about their performance, and (4) Facilitating a conducive learning environment by clarifying assignment expectations, calling students by name, and providing timely responses to student questions (Schulte et al., 2012). In this most recent study they outlined some of the changes that have occurred in the OIES to overcome some of the weaknesses of the system. These weaknesses included the amount of time the evaluations took to perform, having a standardized language in evaluations, and knowing how to balance institution expectations with instructor adaptations. Among the items that they discuss is an objective checklist of items that evaluators utilize to quickly evaluate an instructor. These objective measures decrease the amount of time it takes to perform the evaluations (Schulte, 2009).

Mandernach et al., (2005) provided anecdotal evidence for the system. They reported that the response of online instructors to the system was mixed (Mandernach et al., 2005; Schulte, 2009). Newer instructors were far more receptive, but the more experienced instructors approached the evaluations suspiciously, wondering about the evaluation's purpose (Mandernach et al., 2005; Schulte et al., 2012). Nevertheless, Mandernach et al. (2005) concluded that the benefits far outweighed the drawbacks. In particular, they noted that as a result of the OIES, faculty regularly reflect on their efforts and seek to improve.

Self Evaluation. Self-evaluation is widely used in post-secondary institutions. Delaney et al. (2010) found that 82% utilize self-evaluation as part of their evaluation process. Additionally, a survey administered among the Academic Senate of California Community Colleges confirms that most evaluations include an opportunity for an instructor to evaluate their own efforts (Colleges, 2013). These evaluations often provide instructors with an opportunity to report on their teaching efforts and accomplishments. Sometimes instructors answer specific questions in narrative form, or they may be more broad allowing an instructor to discuss what they feel is pertinent (Berk, 2005). Other occasions the evaluations may be a rebuttal to student evaluations (Colleges, 2013).

Schulte et al. (2012) describe a "self-review" as part of the OIES. Instructors fill these out every two weeks and they coincide with the peer evaluations that are already performed. These self-reports are not shared with the evaluators unless an instructor chooses to, but are ultimately shared at the end of the semester with the instructor's academic department. These reviews provide instructors with additional opportunities to reflect on and improve their performance.

Weschke et al. (2010) describe self-evaluations as one part of a "360-degree view" of an instructor's performance. These self-evaluations encourage instructors to share some of their concerns or challenges they are facing as well as their successes and achievements. The hope is that this along with many other aspects of evaluation will encourage a "self-initiated process" that will lead to improvement (2010, p. 46). Used in this way self-evaluations may motivate greater desires for improved performance.

Unfortunately, self-evaluations among online instructors teaching in post-secondary institutions are far more rare. Piña et al. (2014) found that less than 3% of instructors and administrators (from a sample of 140) reported that the institution they represented employs self-evaluation of instructors. The instructor's own assessment of their teaching effectiveness together with peer and student evaluations may highlight discrepancies that ought to be noted and addressed. Together these three forms of observation and evaluation provide a more complete picture of teaching effectiveness.

When and Why Are Online Instructors Evaluated?

Most institutions perform summative evaluations of online instructors (Dziuban et. al., 2011; Palloff et al., 2008; Schulte et al., 2012). In a survey of 140 online education administrators and instructors, Piña et al. (2014) found that 89% of the institutions represented by the sample utilize end-of-course student surveys.

However, some institutions also utilize formative student evaluations to assess effective online teaching. Flynn, Maiden, Smith, and Wiley (2013), shared an example of student evaluations that were collected halfway through the course, in addition to the student evaluation at the end. These midpoint evaluations opened a dialogue between the instructor and students while corrective measures could still be made. In fact, Flynn et al. explained that the instructors were encouraged to address the feedback with students as a way of acknowledging that they had received it and affirm that they would incorporate it (insomuch as it was feasible) into the remainder of the course. This feedback was also reviewed by the instructor's supervisors, not to make tenure or promotional decisions, but simply to maintain a minimal standard of teaching effectiveness.

The timing of an evaluation is indicative of the purpose for the evaluation (Roberts et al., 2005). When an evaluation of an online instructor occurs during the course, the objective is to receive data that will help to improve teaching effectiveness. Utilizing formative evaluation in this way generally promotes professional development (Colleges, 2013; Dana et al., 2010; DeCosta et al., 2015; Mandernach et al., 2005; Palloff et al., 2008; Schulte et al., 2012; Tinoca et al., 2013; Weschke et al., 2010).

Currently many institutions perform online instructor evaluations primarily to inform decisions concerning tenure and promotion with little other efforts to collect other data to assist in administrative decisions (promotions, employment, etc.) (Colleges, 2013; Darling, 2012; Donovan, 2006; Dziuban et. al., 2011; Roberts et al., 2005; Staniši? Stoji? et al., 2014).

What are institutions evaluating?

Many post-secondary institutions utilize a general course rubric to evaluate online instructors (Drouin, 2012) to help identify any performance concerns that may exist. Some of these rubrics include the following: Quality Matters, Quality Online Course Initiative (QOCI), Online Course Evaluation Project (OCEP), Online Course Assessment Tools (OCATs), the self-assessment Rubric for Online Instruction (ROI) (Drouin, 2012). The use of these rubrics to evaluate online instructors, however, is problematic as they are designed to measure online course *design*. In fact, the creators of the Quality Matters course rubric admitted that it was never intended to evaluate online instructors (Quality Matters, n.d.), so, using them to evaluate online instructors is inadvisable.

Some institutions utilize student evaluation instruments of instructor performance to address the unique nature of online courses compared to traditional courses. Researchers developed these instruments without using a traditional face-to-face instrument as a baseline. These four instruments are listed and compared below and can be found within each of the published studies.

In two different studies, researchers created and tested a student evaluation instrument using Biner's model (1993). Stewart, Hong, and Strudler (2004) was the first of these. They followed Biner's pattern by initially surveying 111 students and three instructors of distance education courses. The survey asked the participants to identify as many items as they could that they felt addressed the effectiveness of a web-based course. The items were then assembled into a tentative instrument. A review of literature confirmed their findings from the initial survey and introduced additional items that they added to the instrument. The final instrument included 44 items organized into seven dimensions. These dimensions included (1) appearance and structure of web pages, (2)

hyperlinks and navigation, (3) technical issues, (4) class procedures and expectations, (5) content delivery, (6) quality of communication, and (7) the presence of instructor and peers. The instrument was then tested for reliability (Chronbach's alpha greater than .70 for each of the measures) and performing an exploratory factor analysis.

Similarly, Roberts, Irani, Telg, and Lundy (2005) modified Biner's pattern by following these steps: (1) having students identify individual items related to course satisfaction, (2) defining dimensions underlying items, (3) selecting essential items, and (4) writing and pretesting the instrument. A sample of 214 students enrolled in a distance education course identified 85 items that they felt could affect the quality of the course. A panel of experts grouped and reduced the 85 items to 20 Likert-type items. These were then organized into 9 dimensions: (1) learner–instructor interaction, (2) learner–learner interaction, (3) learner–content interaction, (4) instructor, (5) course organization, (6) support services/administrative issues, (7) facilitator, (8) technical support, and (9) delivery method. They tested the instrument, sought feedback, revised it, and tested it again. Cronbach's alpha was .95 for the final version. By following Biner's model, both of these instruments target specific aspects that are unique to online courses. These help to highlight some aspects of online instruction that may be overlooked by using a student evaluation instrument that is designed for face-to-face courses.

One of the more widely cited student evaluation instruments developed for online teaching is one developed by Bangert (2008). The Student Evaluation of Online Teaching Effectiveness (SEOTE) was created in 2004 and tested through a series of validation studies (2004, 2006, 2006, 2008). It is based on Chickering and Gamson's seven principles of effective teaching (1987). Using exploratory analysis they created a four factor solution that identified four of the seven principles that have bearing on an online classroom, namely (1) student faculty interaction, (2) cooperation among students, (3) active learning, and (4) time on task. They then performed confirmatory factor analysis and found that only 23 of the original 35 items on the instrument address factors of online teaching effectiveness.

Another instrument used for students to evaluate online courses was developed by Rothman, Romeo, Brennan, and Mitchell (2011). They sought to validate and test the reliability of an instrument they were already using as a student evaluation of instruction. The student sample included 281 students enrolled in 34 online graduate courses for two years. Using a principal components analysis, they identified a six-factor solution that they assert needs to be included for there to be effective instruction in online courses: (1) appropriateness of readings and assignment, (2) technological tools, (3) instructor feedback and communication, (4) course organization, (5) clarity of outcomes and requirements, and (6) content format.

A comparative analysis of these instruments identifies strengths and weaknesses of each (see Table 2). I have grouped the various competencies into seven categories that one or more of the instruments address. I engaged in peer debriefing to ensure the trustworthiness of my categories. As a result, I made minor adjustments to the categories as they are now presented. I then assessed the percentage of the items that addressed each of the categories. The first grouping is the effective use of technological tools, including the effective use of media, chat rooms, and hyperlinks. The second is the visual design and function of the course. This category addresses visual aesthetics: consistent fonts, images, and external links. The third item is how well technical concerns are addressed. Are there links to resources that provide students with the technical support necessary to succeed in the course? The fourth addresses how clear expectations and instructions of course assignments are. This also includes clearly outlining general course objectives. The fifth grouping focuses on assignments. In particular, how well the assignments engaged students and helped them to better understand the subject. The sixth item addresses learning opportunities that encourage student-student interactions. The seventh and eight categories focus on specific things an instructor does to personalize the instruction, demonstrate their expertise in the field, and interact individually with students.

Each of these student evaluations devotes considerable attention to course design. Of the seven categories, five specifically address course design (effective use of technological tools, visual design and function of the course, technical concerns, clear expectations and instructions, student-student interaction, and assignments are meaningful). The only categories that address specific instructor behaviors that are separate from course design are learner-instructor interaction and instructor expertise. The heavy emphasis these instruments place on course design becomes more apparent when all the items of each instrument are represented as a percentage of the total number of items. Stewart et al. (2004) and Roberts et al. (2005) devote 70% and 75% respectively of the items on their instrument to course design. Bangert's (2008) instrument applies 69% to course design and Rothman et al. (2011) is the highest with 88%.

These instruments are appropriate as long as the instructor is also responsible for course design. However, many institutions are adopting a master course model (Cheski & Muller, 2010; Piña et al., 2014). This model involves

designing a course with a team of instructional designers and content experts and then duplicating it into many sections. Administrators then assign as many instructors as enrollment numbers require to facilitate the course. These instructors are usually limited with what aspects of course design they can adjust and which ones they cannot. Institutions that follow this model would be wise to avoid using these instruments to evaluate their instructors. It could make it more difficult to separate the effectiveness of the instructor from the effectiveness of the course design.

Table 2.
A Comparative Analysis of Four Student Evaluations of Online Instructors

	Stewart et al. (2004)	Roberts et al. (2005)	Bangert (2008)	Rothman et al. (2011)
Learner-instructor interaction	25%	10%	31%	12%
Instructor expertise	5%	15%		
Student-student interaction	2%	10%	12%	
Assignments are meaningful	9%	15%	34%	32%
Clear expectations and instructions	14%	10%	4%	16%
Technical concerns	11%	20%		4%
Visual design and function of the course	20%		4%	20%
Effective use of technological tools	14%	20%	15%	16%

Note. The number of items for each instrument are as follows: Stewart et. al. (2004) included 44 items; Roberts et al. (2005) included 20 items; Bangert (2008) included 26 items; and Rothman et. al. (2011) included 25 items.

Ternus, Palmer, and Faulk (2007) also created an instrument specifically for online courses, but that could be utilized by administrators, peers, or individual instructors to evaluate course design. Similar to those mentioned above, this instrument primarily addresses course design. They divide 29 items into four groupings (1) structure, (2) content, (3) processes, and (4) outcomes. Of these items 17% (5 of 29) address specific behaviors of online instructors. They pilot-tested the instrument at two different universities with six different evaluators. They revised the instrument based on their review, but make no mention of what changes they made and why. No conclusive results are included as to the instruments effectiveness. Their instrument is included as part of their study.

Conclusion and Implications

The current landscape of online instructor evaluation is hopeful. The past 10 years of research indicate that post-secondary institutions have done a great deal to address evaluation in online learning. From existing research, it is clear that most institutions evaluate online instructors through end-of-course student evaluations. Unfortunately, far fewer institutions use other measures to evaluate online instructors. By using peer, administrative, or self evaluations, administrators can obtain a far clearer and accurate representation of an online instructor's effectiveness. This can inform one of the main reasons that institutions perform evaluations, namely, to help administrators to make better decisions regarding the hiring, promotion, or firing of instructors. Another reason is to encourage professional development. Evaluations that focus on professional development are often done formatively during a course rather than at the end. Unfortunately, very few institutions utilize formative evaluations during a course in this way.

Another discovery of this review is the heavy emphasis of online instructor evaluation on course design. This is evident in the instruments that have been developed to evaluate online instructors. Instructors are often evaluated as a subset of a general

course evaluation. This may make it difficult to isolate and evaluate specific instructor behaviors. It may be necessary to establish separate evaluations of instructors from course design.

As a result of this review, some unanswered questions still remain. First, much of the literature raises questions about whether or not a general student evaluation of instructor effectiveness can effectively evaluate face-to-face as well as online instructors. Are the behaviors of an online instructor unique enough to warrant a different instrument to measure online instructor effectiveness? Additional research can help address this question.

Second, these articles raise the question of how institutions can address online evaluation without the expense of new staff or faculty. Student evaluations of online instructor effectiveness are sometimes the only measure of instructor effectiveness. Student, peer, administrative, and self evaluations each provide unique information that gives a more accurate picture of an instructor's performance and ought to be included as a comprehensive approach to evaluating online instructors. There is a great deal of research that addresses complex evaluation systems that require several full-time positions devoted to evaluation. This is a luxury that not all institutions can afford. How are institutions addressing these other important measures of instructor effectiveness when resources are limited? This is a discussion that could benefit the research community and could fill a void in the literature.

Third, most instructor evaluations occur as a course concludes. However, some post-secondary institutions perform mid-course student evaluations or peer evaluations. What are some of the benefits and challenges of performing these mid-course evaluations of instructors for students, instructors, and institutions? It may be that these could be performed with minimal costs and maximum gains. This is another area of research that could guide current practices.

Fourth, an important discovery of this review is that post-secondary institutions are evaluating online instructors primarily based on course design. This fails to account for institutions that utilize a master course model where instructors are not responsible for course design. This model is gaining traction and can inform improved practices of instructor evaluation. Some of the research questions that this setting could help answer are these:

- How do these institutions evaluate their instructors separate from course design?
- In what ways can an instructor influence the effectiveness of a course that they did not design?
- What observable behaviors demonstrate online teacher effectiveness?

By studying online instructors that work within a master course model, we can isolate instructor performance separate from course design and learn more about what constitutes a quality online instructor. This knowledge could guide the development of better measures to evaluate online instructors.

Although there has been great progress in the evaluation of online instructors, there is still much more research that needs to be done to improve current practices. Researchers have spent decades trying to answer similar questions about evaluation in traditional courses and still continue seeking answers. It is reasonable to assume that evaluation of online instructors will also require extensive research to continue to improve practice.

References

Academic Senate for California Community Colleges. (2013). *Sound Principles for Faculty Evaluation*. Sacramento, California.

Bangert, A. W. (2004). The seven principles of good practice: A framework for evaluating on-line teaching. *Internet and Higher Education*, 7(3), 217–232. http://doi.org/10.1016/j.iheduc.2004.06.003

Bangert, A. W. (2006). Identifying factors underlying the quality of online teaching effectiveness: an exploratory study. *Journal of Computing in Higher Education*, 17(2), 79–99.

Bangert, A. (2006). The development of an instrument for assessing online teaching effectiveness. *Journal of Educational Computing Research*, 35(3), 227–244. http://doi.org/10.2190/B3XP-5K61-7Q07-U443

Bangert, A. W. (2008). The development and validation of the student evaluation of online teaching effectiveness. *Computers in the Schools*, 25(1/2), 25–47. http://doi.org/10.1080/07380560802157717

Benton, S. L., & Cashin, W. E. (2012). Student ratings of teaching: A summary of research and literature. *The IDEA Center*, (IDEA Paper #50), 1–22.

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. Retrieved from http://www.isetl.org/ijtlhe/

Berk, R. A. (2013). Face-to-face versus online course evaluations: A "consumer' s guide" to seven strategies. *Journal of Online Learning and Teaching*, *9*(1), 140–148.

Biner, P.M. 1993. The development of an instrument to measure student at- titudes toward televised courses. *The American Journal of Distance Education 7*(1), 62–73.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. http://doi.org/10.1191/1478088706qp063oa

Cheski, N. C., & Muller, P. S. (2010). Aliens, adversaries, or advocates? Working with the experts (SMEs). Proceedings from the Conference on Distance Teaching & Learning. Madison WI: University of Wisconsin Extension.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, *Mar*, 3–7. http://doi.org/10.1016/0307-4412(89)90094-0

Coll, C., Rochera, M. J., Gispert, I. D., & Diaz-Barriga, F. (2013). Distribution of feedback among teacher and students in online collaborative learning in small groups. *Digital Education Review*, *2*(23), 27–46. Retrieved from https://files.eric.ed.gov/fulltext/EJ1013721.pdf

Cordeiro, W. P., & Muraoka, D. (2015). Lessons learned: Creating an online business degree from a successful on-campus business degree. *Research in Higher Education Journal*, *27*(1), 1–9.

Dana, H., Havens, B., Hochanadel, C., & Phillips, J. (2010). An innovative approach to faculty coaching. *Contemporary Issues in Education Research*, *3*(11), 29–34. Retrieved from http://www.journals.cluteonline.com/index.php/CIER/article/view/244

Darabi, A. A., Sikorski, E. G., & Harvey, R. B. (2006). Validated competencies for distance teaching. *Distance Education*, *27*, 105–122. http://doi.org/10.1080/01587910600654809

Darling, D. D. (2012). *Administrative Evaluation of Online Faculty in Community Colleges*. Fargo, North Dakota: North Dakota State University.

DeCosta, M., Bergquist, E., & Holbeck, R. (2015). A desire for growth: Online full-time faculty's perceptions of evaluation processes. *Journal of Educators Online*, *12*(2), 73–102.

Delaney, J., Johnson, A., Johnson, T., & Treslan, D. (2010). *Students' perceptions of effective teaching in higher education*. Retrieved from http://www.mun.ca/educ/faculty/mwatch/laura_treslan_SPETHE_Paper.pdf

Donovan, J. (2006). Constructive student feedback: Online vs. traditional course evaluations. *Journal of Interactive Online Learning*, *5*(3), 283–296.

Drouin, M. (2012). What's the story on evaluations of online teaching? In M. E. Kite (Ed.), Effective evaluation of teaching: A guide for faculty and administrators (pp. 60-70). Washington, DC: Society for the Teaching of Psychology. Retrieved from http://www.teachpsych.org/Resources/Documents/ebooks/evals2012.pdf

Dziuban, C., & Moskal, P. (2011). A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments. *Internet and Higher Education*, *14*(4), 236–241. http://doi.org/10.1016/j.iheduc.2011.05.003

Eskey, M. T., & Schulte, M. (2012). Comparing attitudes of online instructors and online college students: Quantitative results for training, evaluation and administration about administrative student support services? *Online Journal of Distance Learning Administration*, *15*(4).

Flynn, M., Maiden, R. P., Smith, W., & Wiley, J. (2013). Launching the virtual academic center: Issues and challenges in innovation. *Journal of Teaching in Social Work*, *33*(4-5), 339–356. http://www.tandfonline.com/doi/abs/10.1080/08841233.2013.843364

Gaytan, J., & McEwen, B. C. (2007). Effective online instructional and assessment strategies. *American Journal of Distance Education*, *21*(3), 117–132. http://doi.org/10.1080/08923640701341653

Gorskey, P., & Blau, I. (2009). Online teaching effectiveness: A tale of two institutions. *The International Review of Research in Open and Distance Learning*, *10*(3), 1–27.

Graham, C., Cagiltay, K., Lim, B.-R. B., Craner, J., & Duffy, T. T. M. (2001). The technology source archives - seven principles of effective teaching: A practical lens for evaluating online courses. *The Technology Source Archives*. Retrieved from http://technologysource.org/article/seven_principles_of_effective_teaching/

Hathorn, L., & Hathorn, J. (2010). Evaluation of online course websites: Is teaching online a tug-of-war? *Journal of Educational Computing Research*, *42*(2), 197–217. http://doi.org/10.2190/EC.42.2.d

Kennedy, J. (2015). Using TPCK as a scaffold to self-assess the novice online teaching experience, *36*(1), 148–154.

Lack, K. a. (2013). Current status of research on online learning in postsecondary education, 73. http://doi.org/10.18665/sr.22463

Loveland, K. A. (2007). Student evaluation of teaching (SET) in web-based classes?: Preliminary findings and a call for further research. *The Journal of Educators Online*, *4*(2), 1–18.

Mandernach, B. J., Donnelli, E., Dailey, A., & Schulte, M. (2005). A faculty evaluation model for online instructors: Mentoring and evaluation in the online classroom. *Online Journal of Distance Learning Administration*, *8*(3), 1–28. Retrieved from https://www.westga.edu/~distance/ojdla/fall83/mandernach83.htm

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning*. *Structure*. Washington, D.C.: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

Moore, J. (2014). Effects of online interaction and instructor presence on students' satisfaction and success with online undergraduate public relations courses. *Journalism & Mass Communication Educator*, *69*(3), 271–288. http://doi.org/10.1177/1077695814536398

Moskal, P., Dziuban, C., & Hartman, J. (2013). Blended learning: A dangerous idea? *Internet and Higher Education*, *18*, 15–23. http://doi.org/10.1016/j.iheduc.2012.12.001

Nandi, D., Hamilton, M., & Harland, J. (2012). Evaluating the quality of interaction in asynchronous discussion forums in fully online courses. *Distance Education*, *33*(1), 5–30. http://doi.org/10.1080/01587919.2012.667957

Palloff, R. M., & Pratt, K. (2008). Effective course, faculty, and program evaluation, 1–5.

Pearson. (2016). Online report card: Tracking online education in the United States, 1. Retrieved from http://www.onlinelearningsurvey.com/reports/2015SurveyInfo.pdf

Piña, A. A., & Bohn, L. (2014). Assessing online faculty: More than student surveys and design rubrics. *The Quarterly Review of Distance Education*, *15*(3), 25–34.

Rhea, N., Rovai, A., Ponton, M., Derrick, G., & Davis, J. (2007). The effect of computer-mediated communication on anonymous end-of-course teaching evaluations. *International Journal on E-Learning*, *6*(4), 581–592. Retrieved from https://eric.ed.gov/?id=EJ771828

Roberts, G., Irani, T. G., Telg, R. W., & Lundy, L. K. (2005). The development of an instrument to evaluate distance education courses using student attitudes. *American Journal of Distance Education*, *19*(1), 51–64. http://www.tandfonline.com/doi/abs/10.1207/s15389286ajde1901_5

Rothman, T., Romeo, L., Brennan, M., & Mitchell, D. (2011). Criteria for assessing student satisfaction with online courses, *1*(June), 27–32. Retrieved from http://infonomics-society.org/wp-content/uploads/ijels/published-papers/volume-1-2011/Criteria-for-Assessing-Student-Satisfaction-with-Online-Courses.pdf

Schulte, M. (2009). Efficient evaluation of online course facilitation: The "quick check" policy measure. *Journal of Continuing Higher Education*, *57*(2), 110–116. http://doi.org/10.1080/07377360902995685

Schulte, M., Dennis, K., Eskey, M., Taylor, C., & Zeng, H. (2012). Creating a sustainable online instructor observation system:

A case study highlighting flaws when blending mentoring and evaluation. *International Review of Research in Open and Distance Learning*, *13*(3), 83–96.

Shao, L. P., Anderson, L. P., & Newsome, M. (2007). Evaluating teaching effectiveness: where we are and where we should be. *Assessment & Evaluation in Higher Education*, *32*(3), 355–371. http://doi.org/10.1080/02602930600801886

Stanišic´ Stojic´, S. M., Dobrijevic´, G., Stanišic´, N., & Stanic´, N. (2014). Characteristics and activities of teachers on distance learning programs that affect their ratings. *International Review of Research in Open and Distance Learning*, *15*(4), 248–262.

Stewart, I., Hong, E., & Strudler, N. (2004). Development and validation of an instrument for student evaluation of the quality of web-based instruction. *American Journal of Distance Education*, *18*(3), 131–150. http://www.tandfonline.com/doi/pdf/10.1207/s15389286ajde1803_2

Sunal, D. W., Sunal, C. S., Odell, M. R., & Sundberg, C. a. (2003). Research-supported best practices for developing online learning. *Learning*, *2*(1), 1–40.

Tallent-Runnels, M. K., Cooper, S., Lan, W. Y., Thomas, J. A., & Busby, B. (2005). How to Teach Online?: What the Research Says. *Distance Learning*, *2*(1), 21–27.

Ternus, M. P., Palmer, K. L., & Faulk, D. R. (2007). Benchmarking quality in online teaching and learning: A rubric for coruse construction and evaluation. *The Journal of Effective Teaching*, *7*(2), 51–67.

Terry, N. (2007). Assessing instruction modes for master of business administration (MBA) Courses. *Journal of Education for Business*, *82*(4), 220–225. http://doi.org/10.3200/JOEB.82.4.220-225

Tinoca, L., & Oliveira, I. (2013). Formative assessment of teachers in the context of an online learning environment. *Teachers and Teaching: Theory and Practice*, *19*(2), 221–234. http://doi.org/10.1080/13540602.2013.741836

Tobin, T. J. (2004). Best practices for administrative evaluation of online faculty. *Online Journal of Distance Learning Administration*, *7*(2), 1–12. Retrieved from http://www.westga.edu/~distance/ojdla/summer72/tobin72.html

Weschke, B., & Canipe, S. (2010). The faculty evaluation process?: The first step in fostering professional development in an online university. *Journal of College Teaching & Learning*, *7*(1), 45–57.

Back to the Online Journal of Distance Learning Administration Contents