



2018-07-01

Current State of Online Teaching Evaluation Processes in Post-Secondary Institutions

Jon E. Thomas
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Educational Psychology Commons](#)

BYU ScholarsArchive Citation

Thomas, Jon E., "Current State of Online Teaching Evaluation Processes in Post-Secondary Institutions" (2018). *All Theses and Dissertations*. 7000.

<https://scholarsarchive.byu.edu/etd/7000>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Current State of Online Teaching Evaluation Processes in Post-Secondary Institutions

Jon E. Thomas

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Charles R. Graham, Chair
Randall Spencer Davies
Anthony Alan Piña
Richard Edward West
Stephen C. Yanchar

Department of Instructional Psychology and Technology

Brigham Young University

Copyright © 2018 Jon E. Thomas

All Rights Reserved

ABSTRACT

Current State of Online Teaching Evaluation Processes in Post-Secondary Institutions

Jon E. Thomas
Department of Instructional Psychology and Technology, BYU
Doctor of Philosophy

This is a multi-article dissertation that seeks to address the current state of online teaching evaluation processes in post-secondary institutions. The last two decades have seen a dramatic increase in enrollment in online courses at post-secondary institutions. Unfortunately, evaluating online instructors has been a neglected field of research leaving many post-secondary institutions to develop their own evaluation systems. A deeper analysis of the current practices of online instructor evaluation will help administrators to strengthen their evaluation processes, thereby providing more effective online teaching. The first article is a literature review that explores common practices of post-secondary institutions. By performing an extensive review of the literature, it is clear that very little research has been done to address online instructor evaluation beyond student evaluations. The second article compares different approaches to online instructor evaluation in various post-secondary institutions. By performing interviews with administrators, we found that many institutions are using a variety of types of evaluations and not just student evaluations to evaluate online teaching. The third article is a study that explores how well institutions that utilize a master course model evaluate online teaching competencies. This is done by performing a content analysis of their observational rubrics.

Keywords: virtual universities, online courses, online faculty evaluation, teaching evaluation methods, evaluation research

ACKNOWLEDGEMENTS

I never imagined myself capable of accomplishing something as expansive as writing this dissertation. I marvel as I reflect back on the journey and see how this would never have happened without the support of so many. I must first recognize the unfailing support of Dr. Charles Graham. He has proven to be an exceptional mentor. The weekly meetings with him and high expectations he reiterated helped draw out the best in me. Other members of the IP&T faculty have also each had a unique influence on me. Every course and interaction contributed in a profound way to my development. I also recognize the members of my committee whom I carefully selected for their unique contributions and experiences.

To my parents, who pray regularly for me and my family. They see more than I do. After I finished my master's degree, I was quick to declare that I would never, EVER pursue a doctorate. My father immediately quipped, "We'll see." Their unfailing support, prayers, blessings, and belief in me has buoyed me up in moments of discouragement.

To my amazing wife who always knew I would do this. She carried far more than a wife should ever have to in order to provide me with both the time and the energy to do this. There were many times where I could not bear to see all that she was doing alone and suggested that maybe it would be best to quit. It did not feel that it was worth the sacrifices we were making. She urged me on, "You have to finish this!" I love her dearly and am so grateful for her unflinching support. Her strength has become mine.

To my children who have never known their dad when he wasn't a student. They regularly prayed both personally and as a family that I could finish my dissertation. I am eager to give them all of my excess time.

Finally, to the Lord. It is truly for Him that I am doing this. He was the one that directed me here. I do not see what comes after this but know that He does. I have felt His help throughout the experience and am quick to acknowledge that anything good that has come from my hand is because of Him. I dedicate to Him all that I have, do, and am!

TABLE OF CONTENTS

TITLE	i
ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	v
LIST OF FIGURES	viii
LIST OF TABLES.....	ix
DESCRIPTION OF RESEARCH AGENDA AND STRUCTURE OF THE DISSERTATION ..	x
ARTICLE 1: Common Practices for Evaluating Post-Secondary Online Instructors	1
Abstract.....	2
Methodology.....	4
Exclusion/Inclusion Criteria	4
Analysis.....	4
Discussion.....	6
How are Online Instructors Evaluated?	6
When and Why are Online Instructors Evaluated?.....	17
What are Institutions Evaluating?.....	19
Conclusion and Implications.....	24
References.....	27
ARTICLE 2: Current Practices of Online Instructor Evaluation in Higher Education	34
Abstract.....	35
Literature Review.....	37
Traditional and Online Course Evaluations.....	37

Evaluations Emphasize Course Design	38
The Master Course Model	39
Methodology.....	40
Data Collection	41
Data Analysis	42
Findings.....	43
How do Institutions Evaluate Online Instructors and Why?	43
Discussion.....	50
Triangulation.....	51
Course Observations Using Rubrics	51
Formative Evaluation.....	52
Metrics	53
Future Research	53
Conclusion	54
References.....	55
Appendix.....	59
Qualitative Data	59
ARTICLE 3: Online Teaching Competencies in Post-Secondary Observational Rubrics:	
What Are Institutions Evaluating?.....	96
Abstract.....	97
Review of Literature	99
Similarities and Differences of Evaluating Online and Traditional Courses.....	100
Focusing Evaluations on Instructional Behaviors Instead of Course Design.....	101

Online Instructor Competencies	102
Research Question	103
Methodology	103
Sample.....	104
Data Collection, Coding, and Analysis.....	104
Findings and Discussion	110
How Strongly are Each of the Online Teaching Competencies and the Associated Tasks Represented in the Rubrics?.....	115
Are There any Rubric Items that do not Address the Tasks in the Bigatel et al. (2012) Model?	118
Conclusion and Recommendations.....	120
References.....	124
Appendix A.....	130
Appendix B.....	132
DISSERTATION CONCLUSION	134
DISSERTATION REFERENCES.....	138

LIST OF FIGURES

Article 3

<i>Figure 1.</i> Percentage of total coding decisions (N=421) in eight online teaching observation rubrics coded to competencies outlined by Bigatel et al. (2012).....	113
<i>Figure 2.</i> Percentage of coding decisions (N=49) for Administration/Leadership organized by type of institution.	117

LIST OF TABLES

Article 1

Table 1. *Literature Review Results* 5

Table 2. *A Comparative Analysis of Four Student Evaluations of Online Instructors* 23

Article 2

Table 1. *Types of Institutions and a Generalized Title for Interviewees* 42

Table 2. *Types of Evaluation by Institution Type* 45

Article 3

Table 1. *Peer and Administrative Evaluation Rubrics Collected by Type of Institution* 105

Table 2. *Competencies for Online Teaching Success* 106

Table 3. *Examples of Coded Rubric Items* 109

Table 4. *The Number and Percentage of Coding Decisions in Each Rubric that Address Each
Competency* 112

Table 5. *Other Category Coding Groups and Rubric Item Examples* 114

Table 6. *All Coding Decisions Organized by Tasks, Including Percentage of Total Coding
Decisions, Percentage of Competency, and Rank among All Tasks* 130

Table 7. *Percentage of All Coding Decisions by Task Divided into Three Categories of Post-
Secondary Institution Rubrics; For Profit, Private, and Public* 132

DESCRIPTION OF RESEARCH AGENDA AND STRUCTURE OF THE DISSERTATION

This dissertation focuses on current practices of online instructor evaluation. Online programs have become a common fixture of the current landscape of post-secondary education. Some have suggested that online instructor evaluation is lagging behind the development of online programs (Berk, 2013; Rothman, Romeo, Brennan, & Mitchell, 2011). This is alarming considering that evaluating online instructors brings benefits to institutions, instructors, and students as a result (DeCosta, Bergquist, & Holbeck, 2015; Dziuban & Moskal, 2011; Mandernach, Donnelly, Dailey, Schulte, 2005; Palloff & Pratt, 2008; Roberts, Irani, Telg, & Lundy, 2005; Stanišić Stojić, Dobrijević, Stanić, 2014). This research seeks to explore the current state of online instructor evaluation and identify specific problems and challenges that post-secondary institutions face when evaluating online instructors as well as possible solutions to those problems.

Article 1: Common Practices for Evaluating Post-Secondary Online Instructors

This literature review presents a synopsis of what types of evaluation post-secondary institutions utilize when evaluating online instructors. This article addresses the following questions: (1) How are online instructors evaluated? (2) When and why are online instructors evaluated? (3) What are institutions evaluating? As a result of this research, we found that the literature suggests that online institutions largely use student evaluations to evaluate online instructors (Delaney, Johnson, Johnson, & Treslan, 2010; Piña & Bohn, 2014). Those that use other types of evaluation (peer, administrative, and self) are far less common. Many of these institutions only focus on course design as opposed to observable teaching behaviors during the course (Drouin, 2012; Schnitzer & Crosby, 2003; Schulte, 2009). Additional research could help

inform better practices of online instructor evaluation. This article was published in the winter 2017 edition of the *Online Journal of Distance Learning Administration*.

Article 2: Current Practices of Online Instructor Evaluation in Higher Education

This study seeks to further clarify the current state of online instructor evaluation in post-secondary institutions. We interviewed 10 administrators from four-year, degree-granting institutions that have more than 10,000 online students. Each of these institutions utilized more than one type of evaluation (student, administrative, peer, self, metrics). These interviews helped to answer the following questions: (1) How do institutions evaluate online instructors and why? (2) How do current practices of post-secondary institutions match up with what they feel are best practices of online instructor evaluation? The results of this study helped to identify challenges post-secondary institutions face in evaluating online instructors as well as solutions to those problems. These results help to inform more effective evaluation practices. The appendix includes some of the qualitative data that helped inform the discussion and conclusion sections. This article was published in the summer 2018 edition of the *Online Journal of Distance Learning Administration*.

Article 3: Evaluating Post-Secondary Online Instructors Using an Observation Rubric

From the previous study, we found that many institutions utilize an observation rubric when evaluating online instructors. These rubrics guide evaluators in what observable online teaching behaviors they look for in determining teaching effectiveness. As part of data collection for article 2, we collected 8 rubrics from institutions that utilize a master course model. Institutions that utilize this model employ instructors to teach a course they did not design. This study addressed the following question: To what extent do online teaching evaluation rubrics from institutions with a master-course model address the online teaching

competencies identified by Bigatel, Ragan, Kennan, May, and Redmond (2012)? We broke this research question down into two other questions: (1) How well are each of the online teaching competencies and the associated tasks represented in the rubrics? (2) Are there any items found in the rubrics that do not address the tasks within the Bigatel et al. (2012) model? As a result of this study, we found that institutions are focusing on similar online teaching competencies in their observational rubrics. Their emphasis can also help to inform a clearer model of online teaching competencies. The appendix includes tables to which we briefly refer in the article.

There are several journals to which I am considering submitting this article. These include: *Distance Education*, which has a 25% acceptance rate and a Google Scholar h5-index of 25; *The Internet and Higher Education*, which has a 22% acceptance rate and a Google Scholar h5-index of 43.

ARTICLE 1

Common Practices for Evaluating Post-Secondary Online Instructors

Jonathan E. Thomas

Charles R. Graham

Brigham Young University

Thomas, J. E., & Graham, C. R. (2017). Common practices for evaluating post-secondary online instructors. *Online Journal of Distance Learning Administration*, 20(4). Retrieved from

https://www.westga.edu/~distance/ojdla/winter204/thomas_graham204.html

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.

Keywords: virtual universities, online courses, online faculty evaluation, teaching evaluation methods, evaluation research

Common Practices for Evaluating Post-Secondary Online Instructors

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 (Allen, Seaman, Poulin, & Straut, 2016). This means that more than 5.8 million students at post-secondary institutions enrolled in online education during 2014. 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 focusing on online instructor evaluation. Therefore, the purpose of this literature review is to 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 we coupled in our database search with either faculty evaluation or teacher evaluation (both ERIC thesaurus items). We initially limited the search to more recent articles published in the last decade. The search returned 51 results.

Exclusion/Inclusion Criteria

We excluded any articles of the initial search from the final analysis that did not directly address online instructor evaluation in post-secondary institutions. We 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, we utilized the method of thematic analysis as outlined by Braune and Clarke (2006). We allowed the research questions to drive the data collection. As we examined each article, we 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?	<p><i>Student Evaluations</i>: Bangert, 2004; Bangert, 2008; 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 et al., 2011; Stanišić' Stojic', Dobrijevic', Stanišić', & Stanic', 2014; Terry, 2007.</p> <p><i>Administrative Evaluation</i>: Dana, Havens, Hochanadel, & Phillips, 2010; Tobin, 2004; Weschke & Canipe, 2010.</p> <p><i>Peer Evaluation</i>: Berk, 2005; ASCCC, 2013; Cordeiro & Muraoka, 2015; DeCosta, Bergquist, & Holbeck, 2015; Drouin, 2012; Hathorn & Hathorn, 2010; Mandernach, Donnelly, Dailey, & Schulte, 2005; Palloff & Pratt, 2008; Piña & Bohn, 2014; Schulte, 2009; Schulte, Dennis, Eskey, Taylor, & Zeng, 2012.</p> <p><i>Self-Evaluation</i>: Berk, 2005; Delaney et al., 2010; Drouin, 2012; Kennedy, 2015; Piña & Bohn, 2014; Schulte et al., 2012; Weschke & Canipe, 2010.</p>
When and why are online instructors evaluated?	<p><i>End of Course</i>: Dziuban & Moskal, 2011 Palloff & Pratt, 2008; Piña & Bohn, 2014; Schulte et al., 2012.</p> <p><i>During the Course</i>: ASCCC, 2013; Dana et al., 2010; DeCosta et al., 2015; Palloff & Pratt, 2008; Mandernach et al., 2005; Schulte et al., 2012; Tinoca & Oliveira, 2013; Weschke & Canipe, 2010.</p> <p><i>To Promote Professional Development</i>: Colleges, 2013; Dana et al., 2010; DeCosta et al., 2015; Mandernach et al., 2005; Palloff & Pratt, 2008; Schulte et al., 2012; Tinoca & Oliveira, 2013; Weschke & Canipe, 2010.</p> <p><i>To Inform Administrative Decisions</i>: ASCCC, 2013; Darling, 2012; Donovan, 2006; Dziuban & Moskal, 2011; Roberts et al., 2005; Stanišić' Stojic' et al., 2014.</p>
What are institutions evaluating?	<p><i>Course Design</i>: Bangert, 2004; Drouin, 2012; Roberts et al., 2005; Rothman et al., 2011; Stewart, Hong, & Strudler, 2004; Ternus, Palmer, & Faulk, 2007.</p> <p><i>Instructor Competencies</i>: 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 & Canipe, 2010.</p>

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. For example, several different evaluation measures are used to provide different information and perspectives on the effectiveness of an online instructor. These include 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 & Bohn, 2014). For many years, performing student evaluations of online instructors was overlooked (Darling, 2012). This may have been due to other administrative constraints 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 (ASCC, 2013; Cordeiro & Muraoka, 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, that 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, Loveland 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. Loveland found that many of the measures were statistically significant and accurately described the variation in the global variables. However, of the 18 measures, 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 fail to evaluate things that are relevant to online courses.

A significant study by Dziuban and Moskal (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 was 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 this by 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 et al. (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) and 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. Stanišić' Stojic' et al. (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. 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 according to 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) listed 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 identified 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 (1987). Additionally, he identified several instruments that could guide administrators in their evaluations that follow the principles delineated by Graham et al. (2001). 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 they 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) described a model at Walden University where they sought to make evaluations less punitive and more helpful for instructors. They explained that "program administrators attempt to be assistant problem-solvers rather than a cudgel bearer to 'beat' faculty into compliance" (2010, p. 46). Faculty were 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 also received a follow-up phone call to ensure that the information is clear and to review it together. This helped avoid misunderstandings that may have resulted from the evaluation data.

Dana et al. (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. (2010) 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 is peer evaluation. Peer evaluation, coupled with student evaluation, provides complementary evidence of teaching effectiveness (Berk, 2005; Palloff & Pratt, 2008; Hathorn & Hathorn, 2010). Unfortunately, not many post-secondary institutions perform them (Piña & Bohn, 2014).

The Academic Senate for California Community Colleges, an organization that represents all community colleges in California, seeks to uphold quality standards of education. 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” (ASCCC, 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) recommended peer evaluation as a way to encourage professional development. They even suggested 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) described 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 instructor's 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 et al. (2005) described an innovative approach to online instructor evaluation called the Online Instructor Evaluation System (OIES). Each instructor received five formative evaluations, one of which happened before the semester began. These evaluations were performed by another faculty member with experience in online teaching. The reports from each visit were not reported to the academic department unless there were patterns of behavior, either positive or negative, about which the administrators should be informed. The objective for these visits was to either begin or continue a conversation on professional development and improvement. These discussions were meant to be collaborative between the evaluator and the

instructor that was evaluated. Mandernach et al., (2005) explained, “The low-stakes formative assessments promoted dialogue and sharing of best practices among instructor and evaluator as peers” (p. 5). The evaluator acted more as a counselor that asked questions or provided suggestions. The instructor, likewise, asked questions and proposed solutions. Plans were then made for follow-up visits and discussions.

This system required additional faculty that devoted 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 sought to improve the OIES educator evaluative process by implementing a “Quick Check” (Schulte, 2009, p. 110) evaluation wherein 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 established best practices that guided their system from both a review of literature and their own experience. They sought 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 discussed is an objective checklist of items that evaluators utilized to quickly evaluate an instructor. These objective measures decreased the amount of time it took 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 members regularly reflected on their efforts and seek to improve.

Self-Evaluation. Some post-secondary institutions use self-evaluation extensively. Delaney et al. (2010) found that 82% utilized self-evaluation as part of their evaluation process. Additionally, a survey administered among the Academic Senate of California Community Colleges confirmed that most evaluations included an opportunity for an instructor to evaluate his or her own efforts (ASCCC, 2013). These evaluations provided instructors with an opportunity to report on their teaching efforts and accomplishments. Sometimes instructors answered specific questions in narrative form, or questions may be more broad allowing an instructor to discuss what they felt was pertinent (Berk, 2005). On other occasions, the evaluations were a rebuttal to student evaluations (ASCCC, 2013).

Schulte et al. (2012) described a “self-review” as part of the OIES. Instructors filled these out every two weeks and they coincided with the peer evaluations that were already performed. These self-reports were not shared with the evaluators unless an instructor chose to but were ultimately shared at the end of the semester with the instructor’s academic department. These reviews provided instructors with additional opportunities to reflect on and improve their performance.

Weschke and Canipe (2010) described self-evaluations as one part of a “360-degree view” of an instructor’s performance. These self-evaluations encouraged instructors to share some of their concerns or challenges they were facing as well as their successes and achievements. They hoped that this along with many other aspects of evaluation would encourage a “self-initiated process” that would lead to improvement (p. 46). Used in this way, self-evaluations may motivate greater desires for improved performance.

Although some post-secondary institutions use self-evaluations extensively, it appears that generally, not many institutions use them. Piña and Bohn (2014) found that less than 3% of instructors and administrators (from a sample of 140) reported that the institution they represented employed self-evaluation of instructors. The instructor’s own assessment of his or her teaching effectiveness together with peer and student evaluations help to highlight discrepancies among other evaluation results 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 & Moskal, 2011; Palloff & Pratt, 2008; Schulte et al., 2012). In a survey of 140 online education

administrators and instructors, Piña and Bohn (2014) found that 89% of the institutions represented by the sample utilized end-of-course student surveys.

However, some institutions also utilized formative student evaluations to assess effective online teaching. Flynn et al. (2013), shared an example of student evaluations that were collected halfway through the course, in addition to the typical end-of-course student evaluation. These midpoint evaluations opened a dialogue between the instructor and students while instructors could quickly incorporate feedback during course delivery. In fact, Flynn et al. (2013) explained that the instructors were encouraged to address the feedback with students as a way of acknowledging that they had received it and affirmed 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 correlative to 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 (ASCCC, 2013; Dana et al., 2010; DeCosta et al., 2015; Mandernach et al., 2005; Palloff & Pratt, 2008; Schulte et al., 2012; Tinoca & Oliveira, 2013; Weschke & Canipe, 2010).

Currently, many institutions perform online instructor evaluations primarily to inform decisions concerning tenure and promotion with little additional efforts to collect other data to assist in administrative decisions (promotions, employment, etc.) (ASCCC, 2013; Darling, 2012; Donovan, 2006; Dziuban & Moskal, 2011; Roberts et al., 2005; Stanišić' Stojic' 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 et al. (2004) were 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 an exploratory factor analysis.

Similarly, Roberts et al. (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 targeted 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 et al. (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 asserted 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 identified strengths and weaknesses of each (see Table 2). We grouped the various competencies into seven categories that one or more of the instruments address. We engaged in peer debriefing to ensure the trustworthiness of these categories. As a result, we made minor adjustments to the categories as they are now presented. We then assessed the percentage of the items that addressed each of the categories. The first grouping was the effective use of technological tools, including the effective use of media, chat rooms, and hyperlinks. The second was the visual design and function of the course. This category addressed visual aesthetics: consistent fonts, images, and external links. The third item was how well technical concerns were addressed. Were there links to resources that provided students with the technical support necessary to succeed in the course? The fourth addressed the clarity of expectations and instructions of course assignments. This also included clearly outlining general course objectives. The fifth grouping focused on assignments; in particular, it

addressed how well the assignments engaged students and helped them to better understand the subject. The sixth item emphasized learning opportunities that encouraged student-student interactions. The seventh and eighth categories focus on specific behaviors of an instructor during course delivery to personalize the instruction, demonstrate expertise in the field, and interact individually with students.

Each of these student evaluations devoted considerable attention to course design. Of the seven categories, five specifically addressed course design (effective use of technological tools, visual design and function of the course, technical concerns, clear expectations and instructions, student-student interaction, and meaningful assignments). The only categories that addressed specific instructor behaviors that were separate from course design were learner-instructor interaction and instructor expertise. By organizing all items on the instrument as percentages of the entire instrument, the heavy emphasis these rubrics place on course design becomes more apparent. Stewart et al. (2004) and Roberts et al. (2005) devoted 70% and 75% respectively of the items on their instrument to course design. Bangert's (2008) instrument applied 69% to course design and Rothman et al. (2011) was the highest with 88%.

The use of these instruments is 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 & Bohn, 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 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 et al. (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 addressed course design. They

divided 29 items into four groupings (1) structure, (2) content, (3) processes, and (4) outcomes. Of these items 17% (5 of 29) addressed specific behaviors of online instructors. Researchers pilot-tested the instrument at two different universities with six different evaluators. They revised the instrument based on their review but made no mention of what changes they made and why. No conclusive results were included as to the instrument's effectiveness. Their instrument was 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 more accurate representation of an online instructor's effectiveness. This can inform one of the main reasons that institutions perform evaluations, namely, to help administrators 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 used for face-to-face instructors can effectively evaluate online instructors. Are the behaviors of an online instructor unique enough to warrant a different instrument to measure their efficacy? Additional research can help address this question.

Second, these articles raise the question as to 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 discussion 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 as follows:

- 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 are still 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: Author. Retrieved from https://asccc.org/sites/default/files/publications/Principles-Faculty-Evaluation2013_0.pdf.
- Allen, I. E., Seaman, J., Poulin, R., & Straut, T. (2016). *Online report card: Tracking online education in the United States*. Babson Park, MA: Babson Survey Research Group and Pearson.
- 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. (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 attitudes toward televised courses. *The American Journal of Distance Education* 7(1), 62–73.

- Braun, V., & Clarke, V. (2006). 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](http://doi.org/10.1016/0307-4412(89)90094-0)
- Coll, C., Rochera, M., Gispert, I. D., & Diaz-Barriga, F. (2013). Distribution of feedback among teacher and students in online collaborative learning in small groups. *Digital Education Review*, 2013, num. 23, p. 27-46. <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(1), 105–122. <http://doi.org/10.1080/01587910600654809>
- Darling, D. D. (2012). *Administrative evaluation of online faculty in community colleges*. Fargo: 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., & Wood, G. (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., Craner, J., & Duffy, 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 TPACK as a scaffold to self-assess the novice online teaching experience. *Distance Education*, 36(1), 148–154.
<https://doi.org/10.1080/01587919.2015.1019964>
- 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., Donnelly, 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 <http://www.westga.edu/~distance/ojdla/fall83/mandernach83.htm>.
- 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*(1), 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. Paper presented at the Annual Conference on Distance Teaching & Learning, Madison, WI, 2008. University of Wisconsin.
- 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. <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, *International Journal for e-Learning Security, 1*(1-2), 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>
- Stanišić Stojic', S. M., Dobrijevic', G., Stanišić', 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 course 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/ojdl/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.

ARTICLE 2

Current Practices of Online Instructor Evaluation in Higher Education

Jonathan E. Thomas

Charles R. Graham

Brigham Young University

Anthony A. Piña

Sullivan University

Thomas, J. E., Graham, C. R., & Piña, A. A. (2018). Current practices of online instructor evaluation in higher education. *Online Journal of Distance Learning Administration*, 21(2).

Retrieved from

https://www.westga.edu/~distance/ojdla/summer212/thomas_graham_pina212.html

Abstract

As enrollment of students in online courses has steadily increased over the last few decades, very little attention has been given to online instructor evaluation. This is an area of online education that needs additional research to better ascertain the current state of online instructor evaluation as well as discover ways to improve its effectiveness. The purpose of this study is to identify how institutions evaluate online instructors and why. Findings indicated that the post-secondary institutions studied utilized many types of evaluation including student evaluations, administrative evaluations, peer evaluations, self-evaluations, and metrics in their evaluations. Recommendations for the use of triangulation, course observation rubrics, formative evaluations, and metrics as part of an online instructor evaluation system are provided.

Keywords: virtual universities, online courses, faculty evaluation, evaluation methods, evaluation research

Current Practices of Online Instructor Evaluation in Higher Education

The 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 to attain these standards. It is also critical to ensure that evaluations provide accurate information (Rothman, Romeo, Brennan, & Mitchell, 2011; Tobin, Mandernach & Taylor, 2015). Through evaluations of online courses and the instructors that teach them, important information can be conveyed to instructors, instructional designers, and administrators to improve course quality and facilitate learning objectives. These evaluations inform administrative decisions like tenure and promotion (ASCCC, 2013; Darling, 2012; Donovan, 2006; Roberts, Irani, Telg, & Lundy, 2005; Stanišić ' Stojic ' , Dobrijevic ' , Stanišić ' , & Stanic ' , 2014) as well as professional development (ASCCC, 2013; Dana, Havens, Hochanadel, & Phillips, 2010; DeCosta, Bergquist, & Holbeck, 2015; Mandernach, Donnelly, Dailey, & Schulte, 2005; Palloff & Pratt, 2008).

Unfortunately, the systematic evaluation of online courses and instructors is surprisingly underdeveloped, considering the rapid growth of online education (Thomas & Graham, 2017; Berk, 2013; Rothman et al., 2011). Berk (2013) admitted that “evaluation of these online courses and the faculty who teach them lags far behind” course production, especially “in terms of available measures, quality of measures, and delivery systems” (p. 141). Some institutions still do not perform any evaluation of online instructors, while others perform evaluations that do not measure unique aspects of online instructor performance (Piña & Bohn, 2014). Many instructor evaluations tend to focus more on instructional decisions reflected in course design (Drouin, 2012) rather than specific behaviors of the instructor. This is problematic because not all instructors are responsible for course design and should, therefore, be evaluated separately

(Piña & Bohn, 2014; Schnitzer & Crosby, 2003; Schulte, 2009). More research is needed to inform better practices of online instructor evaluation separate from course design.

The purpose of this study is to inform improved evaluation practices of online instructors by examining current practices of instructor evaluation at post-secondary institutions.

Literature Review

The current landscape of online instructor evaluation includes many of the measures recommended by Berk (2005) as ways to evaluate teaching effectiveness. These recommendations include student, administrative, peer, and self-evaluations. In a survey sent out to attendees of the Distance Learning Administration Conference and to members of the Association for Educational Communications and Technology (AECT), Piña and Bohn (2014) found that the most commonly used method for measuring online instructor effectiveness was student evaluations (89%), followed by supervisor evaluations (47%), peer evaluations (32%), and other (3%). That 92% of institutions performed some kind of evaluation of their online instructors is a positive sign of the maturing nature of online learning. This is significant since the previous decade saw little attention given to evaluating online courses and instructors, while online courses and enrollments increased at a feverish pace (Bangert, 2004; Compora, 2003).

Traditional and Online Course Evaluations

In their review of research on student evaluations, Benton and Cashin (2012) affirmed that traditional and online courses are similar enough that there is no need to develop new instruments. Their conclusion was based on an earlier work of Benton, Webster, Gross, and Pallett (2010). In this study, Benton et al. (2010) tested the use of a student evaluation instrument in both online and traditional courses focused on student's views of learning objectives and whether instructors used a variety of methods in their teaching. They found

minimal differences between the results of students in the two modalities. Dziuban and Moskal (2011) found that traditional face-to-face student evaluation instruments can also measure online instructor effectiveness. Moskal, Dziuban, and Hartman (2013) added further that effective teaching, including providing feedback, answering questions, etc., is the same, regardless of modality.

Berk (2013) agreed that a student evaluation instrument used in a traditional course could also be used in an online course if it will be used to inform summative personnel decisions. Perhaps this is because traditional face-to-face student evaluations tend to address instructor effectiveness in a very broad way, independent of modality.

Berk also suggested that peer evaluations of online teaching ought to be different than those performed in face-to-face courses. He recommended that a new instrument that is specific to online instructors ought to be developed for peer or self-evaluation. Many feel that colleagues are better equipped to evaluate teaching effectiveness than students (Darling, 2012) and this may be even more true in online courses where additional competencies are necessary to be an effective instructor.

Evaluations Emphasize Course Design

In response to the question, “Do you use a rubric to measure online instructor quality?” Piña and Bohn (2014) found that the sample was almost evenly divided among those that use a rubric developed by Quality Matters (33.6%), a rubric that the institution developed on its own (32.9%), and those that did not use any rubric at all (32.9%). Almost one third of those that developed their own rubric admitted that they based it on the Quality Matters rubric, which focuses on course design and not teaching. This is consistent with what Drouin (2012) asserted.

Post-secondary institutions typically use some kind of general course rubric “as checklists” for peer evaluations (p. 61).

These general course rubrics, along with the student evaluations developed specifically for online courses, focus heavily on course design (Tobin, et al., 2015). The student evaluations developed by Stewart, Strong, and Strudler (2004), Roberts et al. (2005), Bangert (2008), and Rothman et al. (2011) devote 70%, 75%, 69%, and 88% respectively of the items on their instrument to evaluate elements of course design (Thomas & Graham, 2017). This heavy emphasis on course design is not ill-placed. Course design is made up of a series of instructional decisions often made by the instructor that teaches the course. In these instances, it is reasonable to evaluate an instructor on the instructional decisions that make up the course design. However, not all online instructors are responsible for the design of the course they teach.

The Master Course Model

A prominent model utilized in online education, but strangely absent from the research, is the master course model (Cheski & Muller, 2010; Hill, 2012). In this model, a team is responsible for course design. This team may include one or more instructional designers and one or more subject matter experts that may or may not be faculty members. When the course design is complete, the course is then duplicated into as many sections of the course as are necessary to accommodate student enrollment. Instructors, usually content matter experts, are then assigned to facilitate a course that they did not design. Consequently, utilizing instruments to evaluate online instructors that focus heavily on course design in these instances would be inappropriate. In these circumstances, it is important to ensure that the evaluative instruments effectively evaluate the online instructor separate from course design.

Methodology

The purpose of this research study is to explore evaluation practices of online instructors at a variety of post-secondary institutions. It will address the following research question: How do post-secondary institutions evaluate online instructors and why?

To accomplish this purpose and answer this research question, this study investigated multiple institutional cases. We utilized purposive sampling to identify all institutions in the United States that are in each of the following three categories: 4-year for profit, 4-year not for profit, and 4-year public. Using a recent report based on data collected by the National Center for Education Statistics' Integrated Postsecondary Education Data System (IPEDS) and data available from a report entitled, *Online Report Card: Tracking Online Education in the United States* (Allen, Seaman, Poulin & Straut, 2016), we identified all institutions in the U.S. that are four year degree-granting institutions offering baccalaureate degrees and above and have more than 10,000 enrolled distance education students (some of which are first-time, full-time undergraduate students). All data is from the 2015 calendar year. By doing this, we identified 15 for profit, 9 private, and 24 public universities.

From these 48 institutions, we sought representation of at least two institutions from each category for inclusion in the study. This is consistent with previously published research that identified three categories and sampled two institutions from each (Graham, Woodfield, & Harrison, 2012). We felt that this gave us the variety of perspectives we needed within each category.

We utilized a network of professionals in online learning to identify individuals at these institutions we could contact as potential interview subjects. We reached out to all those for whom we obtained contact information. Although we were only seeking 2 institutions from each

type, a few more than we anticipated were eager to participate. The final sample included 2 for-profit institutions, 5 private institutions and 3 public institutions. The sample along with the interviewees and their generalized titles are included in Table 1.

Data Collection

This study relied primarily on interviews. The interviews helped to identify instruments used in student, peer, self, or some other kind of instructor evaluation. These instruments provided important data to help answer the questions this study seeks to address. We collected any other documents that informed historical changes in evaluations, previous instruments, and other forms.

To establish credibility and trustworthiness, we performed member checks both during data collection via email as well as after the analysis so that interviewees could confirm our conclusions and ensure that our analyses were accurate. Additionally, we engaged other colleagues with a strong research background to employ peer debriefing. They were invited to review the methodology and conclusions of the study to also help ensure the study's accuracy. The various sources of data collection helped to employ triangulation. By collecting data from interviews, artifacts, and relevant literature, the validity of the study's conclusions was confirmed through multiple data points. Additionally, researchers sought to attain data redundancy in order to ensure that data is adequate to provide meaningful analysis and conclusions.

Table 1

Types of Institutions and a Generalized Title for Interviewees.

Institution	Generalized Title of Interview Participants
For profit 1	Director of Research and Teaching
For profit 2	Department Level Chair
Private 1	Manager of Faculty
Private 2	Director of Faculty
Private 3	Department Level Supervisor
Private 4	Manager of Online Department
Private 5	Assistant Director of Faculty
Public 1	Online Administrator
Public 2	Director of Research
Public 3	Director of Faculty

Data Analysis

This is exploratory research. The transcribed interviews were systematically coded and compared in order to perform a thematic analysis. In coding and analyzing these interview transcripts, we utilized a modified application of thematic network analysis to identify relevant themes (Attride-Stirling, 2001). This began by coding the data into global themes that seek to “[encapsulate] the principal metaphors in the text as a whole” (p. 388). These global themes

were established by the literature. Global themes were then broken down into organizing themes and then into basic themes.

We recognize that as researchers, our work can be influenced by our own biases and experiences. One of the authors has practical experience with performing online instructor evaluations and has preconceived notions about what constitutes effective evaluation. In order to moderate this potential researcher bias, we sought to only describe evaluation practices rather than evaluating them.

Findings

As a result of this research, we found that evaluation of online instructors at the sampled post-secondary institutions shows signs of improvement compared to recent findings in the literature (Thomas & Graham, 2017, Piña & Bohn, 2014). The sampled institutions utilized many different types of evaluation in assessing the effectiveness of online instructors, which include student, administrative, peer, self, and metrics that measure different aspects of teaching effectiveness as part of the evaluation.

How do Institutions Evaluate Online Instructors and Why?

The institutions that participated in this research utilized great variety in the types of evaluation they employ. Rather than depending on only a few types of evaluation, these institutions sought to incorporate many types, as indicated in Table 2. All institutions in this sample utilized student evaluations to evaluate online instructors. Only one institution did not use an administrative evaluation of its online instructors. Of the 10 institutions in this sample, four of them utilized all five of the identified types of evaluation. However, the public institution admitted that it focused primarily on course design in its evaluations.

One discovery of this research was the growing trend of post-secondary institutions utilizing metrics to inform online instructor evaluation. Metrics include performance metrics that address instructor behaviors, activity metrics that address student behaviors, and other metrics that may be indicative of effective teaching. Of the 10 institutions sampled, five did not use metrics as part of their evaluation of online instructors. Some use metrics more than others, but all those that use it have found that it can be a helpful resource in performing online instructor evaluation.

We will describe each of these types of evaluations below as well as the institutions' reason for their use.

Student evaluation. Student evaluations are the only form of evaluation that every institution in this sample used and all reported that it occurred at the end of every course. This type of evaluation allows students to have some influence in their instructional experience. Six of the 10 institutions in this sample reported using student evaluation instruments in online courses that are exactly like the instruments that their institution uses in traditional courses. Those that did mention differences, admitted that there are agreed upon similarities regardless of modality. It is also interesting to note that several institutions utilized formative mid-course student evaluations as part of their process. A formative student evaluation may provide valuable information to an instructor to make improvements during the course, thereby providing a better experience for students.

Table 2

Types of Evaluation by Institution Type

Institution	Student	Administrative	Peer	Self	Metrics
For profit 1	X	X	X	X	X
For profit 2	X	X	X	X	
Private 1	X	X	X	X	X
Private 2	X	X	X	X	X
Private 3	X	X		X	X
Private 4	X	X	X		
Private 5	X	X	X	X	
Public 1	X	X	X	X	X
Public 2	X		X		
Public 3	X	X	X	X	

After performing student evaluations, every institution reported the results back to the instructor. Instructors were then encouraged to use the information provided to make any necessary adjustments. This information was also communicated back to their supervisors or department chairs. On occasion, administrators shared the results of a student evaluation with a faculty support center to provide additional help for a struggling instructor.

Many institutions acknowledged that student bias is a clear problem with student evaluations; however, they also capture a critical perspective on online instructor behaviors that may be missed otherwise. Recognizing the importance of the student perspective, administrators made sure that student evaluation questions addressed things such as instructor communication, feedback and responsiveness that students are capable of answering and evaluating.

Administrative evaluation. Among the institutions sampled, administrative evaluations were performed at least once a year to determine how instructors performed. Administrators may have utilized a variety of other types of evaluations conducted previously to inform their evaluations, including student and peer evaluations or evaluation performed by a separate institutional entity, such as an online support department. Although administrative evaluations were performed at least annually in a summative way, many institutions also performed formative evaluations.

Institutions faced a variety of challenges in performing administrative evaluations. These included the time and logistics of performing evaluations for all instructors and the lack of a common standard among institutions for evaluating online instructors. In addition to this challenge of sufficient resources, many also faced both faculty and department resistance to performing regular evaluations of online instructors. There are a variety of reasons for why faculty and departments resist evaluation. Some attributed this resistance to feeling mistrusted. Others feel like it limits their academic freedom.

Institutions performed administrative evaluations to reveal areas in which instructors could improve teaching. Administrators could then tailor training to meet personalized needs. Additionally, these evaluations identified teachers who excelled and could be rewarded with

promotion or tenure-like benefits. This data also allowed administrators to schedule the best instructors as often as possible.

Peer evaluation. Of the types of online instructor evaluation described in this study, there was the greatest amount of variability among how institutions approached peer evaluations. This was certainly the case with how often the institutions in the sample performed these evaluations, with some doing reviews only during the first year of teaching, and others doing them annually. Still others performed reviews only upon departmental request, while others conducted them during each teaching term.

There is some variation in whom each institution selected to perform the peer evaluation. Those chosen included close associates of the instructor, other faculty who teach the same course, or a departmental supervisor. In all cases, these were individuals that could provide valuable feedback to an instructor because of their own experience and/or training. This evaluation typically involved the peer “visiting” an online course and observing the teaching activities of the instructor.

Six institutions in this sample utilized a rubric to facilitate the peer evaluation. These rubrics addressed very similar things including the kind of feedback instructors gave and the timeliness of grading. Many also addressed the regular posting of useful, course-specific announcements by the instructor and regular, positive interactions with students that encouraged participation and dialogue through email or discussion boards. Each rubric also had some variability in institutional goals and other items that a peer evaluator might consider during his or her course visit.

Institutions identified several challenges in performing peer evaluations. Orchestrating a process for effective peer-to-peer evaluation sometimes required more resources to organize and

implement than were currently at an institution's disposal. The time and resources necessary to perform peer evaluations required careful consideration as to whether the benefits would be worth the cost.

There was also some concern about the subjective nature of these evaluations, which may have been a result of a rubric that was vague, an evaluator's inattention to detail, or even the evaluator's reluctance to provide any critique of a colleague's performance. These "love letters," as one institution called them, did little to improve teaching or identify high quality instructors. Without a clear standard to measure by, peer evaluations may continue to vary based on who is evaluating.

Several institutions identified three major reasons why they perform peer evaluations. First, they helped to identify effective teaching practices among their faculty. These instructors were then encouraged to share these practices with their peers. Second, peer evaluation provided a safe environment for feedback because the results were often not reported to administrators. Administrators felt that when instructors were evaluated, they tended to feel exposed and vulnerable. When the evaluation was performed by someone that an instructor knows and feels comfortable with, it helped to lower any defensiveness that would otherwise result in an unwillingness to take feedback. The final reason administrators used peer evaluation is to provide an avenue whereby instructors can pursue and demonstrate professional development to make their case for institutional benefits like promotion or tenure.

Self-evaluation. The most common approach to self-evaluation among the sample was to use an unstructured format. Instructors were invited to write whatever they would like to about their goals, personal improvement plans, or how they feel that they have demonstrated excellence. Some institutions had instructors fill out the same form or rubric that is filled out

during a peer or administrative evaluation. These were then used to inform the subsequent peer or administrative evaluation. It allowed instructors to spot areas of weakness in anticipation for the evaluation of others. It can also lead to instructors making a case for why their assessment of their performance is more accurate than a peer or administrator in case there are inconsistencies among reviewers. Among the institutions that utilized self-evaluation the main purpose was to provide a reference point for other types of evaluations, primarily the peer evaluation.

Metrics evaluation. Half of the institutions in this sample utilized metrics to evaluate their instructors. The institutions that were utilizing metrics in their evaluations were at varying levels of development and use. Some had developed programs that automatically retrieve and aggregate data. Aggregated data can be retrieved from the learning management system and student information system and used to populate dashboards for administrators and instructors. Others retrieved various types of data to analyze and discover useful statistical patterns that informed faculty, administrators, or other faculty support departments about effective teaching behaviors and student indicators of teaching effectiveness.

Some examples of student indicators of teaching effectiveness included student engagement, satisfaction, and success. Metrics can help identify instructors whose students regularly do more than what is expected of them for a good grade. This suggests that students care and are engaged in what they are learning. Student satisfaction is another indicator of teaching effectiveness that metrics can help faculty and administrators to more clearly see. Some institutions used metrics to evaluate instructor effectiveness through their students' success rates. This was defined in a variety of ways, but often includes student retention and success in subsequent courses.

The main challenge that institutions have faced by incorporating the use of metrics into their evaluation is the resistance of faculty members. Some faculty were worried about how automated metrics might infringe on their academic freedom when their own pedagogical approach was different from institutional philosophies. They may feel forced to comply to the institutional policies rather than risk being labeled as an ineffective instructor. Other faculty expressed concerns about being accountable for the success or satisfaction of their students. This is a difficult obstacle to the effective use of metrics as part of an online instructor evaluation process.

A unique reason institutions used metrics is to identify teachers that are not meeting baseline standards in a very fast and effective way. It did not address the quality of instruction, only the lack thereof. This was a more efficient and precise way to monitor instructor behavior as well as student engagement, satisfaction, and even success. Although there are obstacles to effective use of metrics at institutions, it provides a possible solution to the challenge of an unwieldy and large evaluation system.

Discussion

In comparing and analyzing the online instructor evaluation processes at 10 different institutions, there are several implications for online instructor evaluation. These include the importance of triangulation in providing a clear representation of instructor teaching effectiveness, employing course observations using a rubric, incorporating formative evaluations into the process, and capitalizing on the use of metrics.

Triangulation

A pattern we discovered in this research was that all institutions relied on multiple sources of data and types of evaluation in their process. We conclude from this that triangulation (i.e., using a variety of sources) is an important aspect of an effective evaluations process. By utilizing a variety of approaches to evaluation, different types of useful data can be acquired to better inform faculty and administrators of effective teaching. It is important to include the insights of students, skilled peers, and the instructors themselves to provide a more complete picture of the instructor's efforts to be effective in his or her virtual classroom.

Course Observations Using Rubrics

The majority of the institutions in this study performed online course observations as part of either an administrative or peer evaluation. This allowed them to observe specific online teaching behaviors and more accurately assess online teaching effectiveness. We recommend that this be a part of evaluation systems of online programs. These observations allow administrators to make more accurate evaluations of instructors by focusing on teaching behaviors and not only on course design. This is particularly important when instructors did not design the course they are teaching. Course observations also provide opportunities to provide feedback to improve or commend effective teaching. All institutions that utilized this type of evaluation found this information to be among the most useful in determining instructor effectiveness.

In every case where institutions used course observations as part of an evaluation, observers used an observation rubric. These institutions explained that the use of a rubric helps establish standards of instructor behavior and clear directions to observers about what to look for in their evaluation. They have regularly revised their rubrics constantly seeking to be more clear

and simple in order to avoid inconsistency among observers. Therefore, we suggest that institutions develop and use rubrics to guide observers as they perform course visits. These rubrics may take time to revise and improve. They help provide clear standards of performance and contribute to more trustworthy evaluations of teaching behaviors.

Formative Evaluation

All institutions in this study supported the use of formative evaluations, but not all performed them institution-wide. In most cases where the institution did not have an established, formative evaluation process, means were available for faculty to perform their own formative evaluations either with students or another peer. These are valuable evaluations that ought to be a part of evaluation systems institution-wide. Based on our findings, we recommend that formative evaluations should only be communicated with the instructor. Observations, either performed by a peer or an administrator, can be an important way to provide formative feedback. Peer observations, in particular, provide a safe environment for instructors to seek and receive feedback, especially when they know that the results will not be communicated to supervisors. Formative student evaluations during a course can also provide great feedback to instructors. This mid-course feedback allows instructors to make immediate adjustments to their teaching to better serve students. Instructors are not as responsive to end-of-course student evaluations as they could be to mid-course student evaluations.

Although formative evaluations can have positive effects on teaching, performing them too often can also negatively affect instructor morale. Instructors, generally, do not like evaluation. Some faculty may assume that increased evaluations mean mistrust in their ability as an instructor or assume that an administrator worries there may be a problem. Determining a balanced approach to formative evaluations may vary based on the institution. What works for

one institution, may not necessarily work for another. Be prepared to implement a plan and revise accordingly.

Metrics

Institutions that are effectively using metrics, have found that it helps an institution monitor instructor behavior in an efficient way without the need of large numbers of online course observers. As a result of these findings, we conclude that institutions should make efforts to incorporate the use of metrics in their online instructor evaluations. Automated metrics can meet the needs of an institution by allowing regular monitoring of behavior without the intrusiveness of peer or administrative observations. They perform an important role in helping an evaluation system become efficient and scalable. By implementing an automated system of producing metrics that populate a dashboard, administrators can regularly have a pulse on faculty and ensure that they are meeting baseline standards for faculty. The use of metrics can allow course observations to focus more on quality rather than simply a baseline standard of performance. Other metrics can also monitor student behaviors that provide additional insight into the quality of an instructor. This kind of data may be difficult to use without establishing a system to not only retrieve the data, but also to employ statistical analyses on the data. This will help to translate the data into clear indicators of effective instructional behaviors.

Future Research

Additional research could focus on the specifics of online course evaluations to evaluate instructor performance. In particular, which information is included on the course observation rubrics institutions use to help guide evaluators. What teaching behaviors are similar and which are different? It would also be helpful to know what institutions based their decisions on as they

developed their instrument. Was it largely based on research or were there items that they added after their own observations and experience?

Other research that would be extremely useful could focus on more specific details on how institutions utilize metrics to help evaluate instructors. Which metrics are meaningful to collect and utilize in the regular process of evaluation to really inform effective practices of instructors? Case study examples of how institutions developed their use of metrics, including their use of the LMS (whether it was developed by a third party or by the institution itself), could also help inform best practices of online instructor evaluation.

A final suggestion for future research includes establishing a consensus among post-secondary institutions regarding online teaching competencies. This could facilitate the development of rubrics of online teaching observations. Related to this, research could also explore current rubrics used by institutions to evaluate online teaching. These rubrics could be compared to help identify criteria being used across institutions and better inform online teaching competencies.

Conclusion

This study confirms that online instructor evaluation at post-secondary institutions has made tremendous improvements in recent years. It is apparent that online programs are aware of the importance of online instructor evaluation as well as the challenges. Many institutions have been grappling with challenges for many years and have found important solutions to difficulties with which others are still struggling. Additional research can help share these solutions and thereby continue to improve practices of online instructor evaluation.

References

- Academic Senate for California Community Colleges. (2013). *Sound Principles for Faculty Evaluation*. Sacramento, California: Author. Retrieved from https://asccc.org/sites/default/files/publications/Principles-Faculty-Evaluation2013_0.pdf.
- Allen, I. E., Seaman, J., Poulin, R., & Straut, T. (2016). *Online report card: Tracking online education in the United States*. Babson Park, MA: Babson Survey Research Group and Pearson.
- Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research, 1*(3), 385–405. <http://doi.org/10.1177/146879410100100307>
- 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. (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.
- Benton, S. L., Webster, R., Gross, A. B., & Pallett, W. H. (2010). IDEA technical report no. 15: An analysis of IDEA student ratings of instruction in traditional versus online courses, 2002–2008 data. *The IDEA Center, (IDEA Paper #15)*, 1–38.
- 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.
- 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.
- Compura, D. (2003). Current trends in distance education : An administrative model. *Online Journal of Distance Learning Administration*, 6(2), 1–17.
- 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>
- 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.
- 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>

- Hill, P. (2012). Online educational delivery models: A descriptive view. *EDUCAUSE Review*, 47(6), 84–86. Retrieved from <http://ezproxy.lib.utexas.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ996668&site=ehost-live%5Cnhttps://net.educause.edu/ir/library/pdf/ERM1263.pdf>
- Graham, C. R., Woodfield, W., & Harrison, J. B. (2012). A framework for institutional adoption and implementation of blended learning in higher education. *The Internet and Higher Education*, 18(1), 4–14. <http://doi.org/10.1016/j.iheduc.2012.09.003>
- Mandernach, B. J., Donnelly, 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 <http://www.westga.edu/~distance/ojdla/fall83/mandernach83.htm>.
- Moskal, P., Dziuban, C., & Hartman, J. (2013). Blended learning: A dangerous idea? *Internet and Higher Education*, 18(1), 15–23. <http://doi.org/10.1016/j.iheduc.2012.12.001>
- Palloff, R. M., & Pratt, K. (2008) Effective course, faculty, and program evaluation. Paper presented at the Annual Conference on Distance Teaching & Learning, Madison, WI, 2008. University of Wisconsin.
- 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.
- 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://doi.org/10.1207/s15389286ajde1901>

- Rothman, T., Romeo, L., Brennan, M., & Mitchell, D. (2011). Criteria for assessing student satisfaction with online courses, *International Journal for e-Learning Security*, 1(1-2), 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>
- Schnitzer, M., & Crosby, L. S. (2003). Recruitment and development of online adjunct instructors. *Online Journal of Distance Learning Administration*, 6(2), 1–7. Retrieved from http://www.westga.edu/~distance/ojdl/summer62/crosby_schnitzer62.html
- 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>
- Stanišić' Stojic', S. M., Dobrijevic', G., Stanišić', 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://doi.org/10.1207/s15389286ajde1803>
- Thomas, J. E., & Graham, C. R. (2017). Common practices for evaluating post-secondary online instructors. *Online Journal of Distance Learning Administration*, 20(4). Retrieved from https://www.westga.edu/~distance/ojdl/winter204/thomas_graham204.html
- Tobin, T. J., Mandernach, B. J., & Taylor, A. H. (2015). *Evaluating online faculty: Implementing best practices*. San Francisco, CA: Jossey-Bass.

Appendix

Qualitative Data

Here we have included rich data compiled from interview transcripts that informed the conclusions made in this article. The transcripts are not included in their entirety. In this section we address how institutions are evaluating online instructors. We discuss the use of student, administrative, peer, self, and metric evaluation.

Student evaluation. One administrator elaborated that there are shared items on student evaluations regardless of modality: “We have some shared, agreed upon, best practices in general like the faculty are responsive, the faculty are timely in their feedback, are good communicators, but then the way it is worded on the student evals is unique for face-to-face and online. The concepts are similar, but how they are operationalized depends on mode” (For private 1). By focusing on these kinds of concepts, the instrument could look very similar for any modality.

One institution reported that while they use a campus wide, end-of-course student evaluation that is not specific to modality, they have a mid-course student evaluation that is not only specific to online but also addresses teaching behaviors of the online instructor. This formative, mid-course evaluation is a part of the formalized evaluation process. This institution is not the only one that reported doing student evaluations during a course in addition to the typical end-of-course/term evaluation. One reported that there are formative student evaluations following any interaction between a teacher and student. These interactions occur in a formal way, often by phone, on a regular basis. Following these interactions, students are randomly selected to fill out an evaluation on how helpful the instructor or student mentor was during the interaction. One institution also reported that although mid-course student evaluations are not

formally a part of its instructor evaluation process, instructors “have the option of doing their own midterm evaluation if they choose, but those are considered formative. They do not feed into any overarching database” (For profit 1).

Some supervisors are not really instructed to do anything with the results of student evaluations. One administrator explained that it is possible that one of these supervisors might look through the results of the evaluation of an instructor they supervise, “see red flags,” and approach the instructor to ask, “how can I help you more with this?” but for the most part, the instructor’s supervisor would do little else with the results of the student evaluation (Private 1).

Another institution explained that the results of the student evaluations are fed into the system and represented in a dashboard that allows supervisors to “see every day what the students' answers are to any of those surveys” (Private 2). This then allows the supervisor to take any corrective action they deem necessary. Instructors and supervisors can receive near real time feedback on how instructors are performing with students.

One institution reported that the results of student evaluations can be accessed by a variety of people and are often used to inform program specific questions. These include the college “dean, associate dean, the chairs, or department chairs” (Private 4). They then usually access the data as a summary report “by course, by instructor, by degree level, by college” in order to “see what their students are saying about multiple facets of the program” (Private 4). Another institution also explained that department chairs receive the results “in a summative kind of way” to help them make decisions regarding faculty. “It's very high stakes here. These ratings, you know, have impact on financial incentives, raises, awards, you know, all kinds of things, tenure decisions are, to some degree, based on them. So, they are, they have a lot of gravitas on this campus, you know, and I think on other campuses as well” (Public 2). By

making the information clearly accessible, it can be obtained with ease to inform decisions administrators may need to make regarding faculty.

Although data collected from student evaluations “are intended to be formative” one administrator countered, “they wind up being summative, the way our departments use them... We have tried to discourage them from doing that with very little luck” (Public 2). Typically, departments will “look at the average of an instructor compared to the average for the department... and make some evaluative judgments” (Public 2). Thus, the results of the student evaluation are typically used in a summative way.

Many institutions have acknowledged that student bias is a clear problem with student evaluations. “All the [students] who really don't like you are going to [complete and submit a student evaluation], all the ones who really like you are going to do it and then the rest of 'em you know fall in” (Private 3). Another admitted that “a lot of the students will answer it in an extreme. Either they really, really liked the instructor or they had a very bad experience and so now they have to go in and have to make themselves heard, so we realize that those types of issues usually happen at least for us” (Private 4). Another agreed and explained “we hate to use a current end-of-term survey score in isolation, because it could be biased in that way” (For profit 2). Student evaluations must be cautiously considered because of the potential for bias in the results, but still must be a consideration in the evaluation of an instructor. “When [a student], even on Facebook, says ‘this guy is so boring, my pillow needs a pillow,’ I think there is something we need to pay attention to in a course. I know Rate My Professors gets trashed, but it does capture a voice. And students want their voices heard” (Public 2). Another institution explained that even though students do not understand effective pedagogy, “they know that if

they had a problem and they went to you whether you helped them or not” (Private 3). Some aspects of teaching simply cannot be evaluated “without hearing from students” (Public 3).

Recognizing the importance of the student perspective, administrators have made sure that student evaluation questions address things students “are capable of answering and evaluating. So rather than asking them broad questions about the instructor’s pedagogical variability or those things, we asked, “Did the instructor communicate effectively?”, “Was the instructor responsive?”, “Could you understand the feedback?”, “Did they provide useful feedback?” So, we really tailored down that evaluation, so it was stuff students could tell us about” (For profit 1). One institution admitted that they “look more at the comments that the students make than just the scores... we do look at the comments very closely” (For profit 2). They felt that this can help to identify clear patterns that suggest reason for concern. “If we see a comment repeated by several students or, or you know, the same flavor of a comment repeated by several students, that’s something we go back and look at” (For profit 2). Another administrator agreed and explained that they felt that “the most important part of the course student survey... is a thing... for comments. Those comments are validators to me as to really what’s going on... that is where you can really get um good ammunition to either support that [instructors] are better than sliced bread or they are sliced bread, or they didn’t even figure out that we have a loaf of bread” (Private 3). The open response aspect of a student evaluation provides a valuable avenue for students to provide feedback on their instructor.

Administrative evaluation. Of the institutions in the sample, three reported using a peer evaluation to inform their administrative evaluation. In one case, they explained that it may be “full time faculty members that are approved to teach the same course or even the course developer [that] will go in and do a review of the course to provide the feedback over to the

department chair before finalizing the annual evaluation” (Private 4). One institution reported that even though it takes a tremendous amount of resources to perform effective peer evaluations, “where it pays off and where we gain the time back” is during the administrative evaluation. They store all of the peer evaluations in a database. When it is time for the administrative evaluation, it “dumps the information to an Excel spreadsheet, that I then read in to the database, it uses [the peer evaluation] scores and writes all the comments for me for the adjunct evaluations... So, all I have to do when I do the adjunct eval is cut it, paste it, done. ‘Cause it actually averages their score on each of those items for the whole year and then uses a ranking system to write which comment is appropriate” (For profit 2). By leveraging the peer evaluation as part of the administrative evaluation, it can reduce the amount of time it takes to complete administrative evaluations

Other evaluations that may be used in administrative evaluations include self-evaluations and student evaluations of instructor performance. Three institutions reported using self-evaluations as part of their administrative review of instructor performance. Although it is probable that more do, only two institutions reported referring to student evaluation results to inform the administrative evaluation (Private 4).

Some institutions have a department organized to improve online learning. As a part of this effort, they focus on improving online teaching. They evaluate online instructors and communicate the results to whomever will perform the administrative evaluation. These support entities use a rubric in their evaluation. They “do a thorough review of the course looking at the quality of the interactions.” Then, they “fill out that form providing not only a rating, but then also comments, anything that was exceeds or below I require that there be comments added, uh to justify or substantiate that particular rating” (Private 4). This process seeks to determine

whether instructors are “meeting certain criteria” and not necessarily if “the instructor is engaging” (Private 5).

Some institutions perform more than one review before submitting it to the department administrator. One institution explains that following the initial evaluation, it is submitted to the online operations manager who performs a “a cursory review making sure that it looked like all the comments aligned with the ratings provided.” This may also include “a random audit” where they will also look at the course and the review to make sure that there is “an alignment there” (Private 4). It would then be submitted “to the assigned department chair. That triggered it over to being the evaluation” (Private 4). The department chair also would visit the course to ensure that all the ratings were accurate. If necessary, they would request edits on items that felt were too subjective along with “other things that we may not necessarily have access to within the faculty quality management team; what conversations or improvement plans have they put that instructor on previously? Had there been noted improvement in those areas?” (Private 4). They then complete the evaluation and send it to the instructor.

As evidenced by this example, some of these support institutions are limited in their ability to provide thorough evaluations. One explained that they do not do anything more than evaluate online instructor behaviors “because of the fact that we are not exactly subject matter experts in all of the disciplines” (Private 5). At this particular institution, the faculty is “more of a matrix system” where the online department addresses some aspects of the online instruction and the dean or chair addresses other aspects. “If it starts to get specific,” in terms of a particular subject matter, then the online department will “reach out to the actual department” (Private 5). One institution explained that they try to ameliorate this problem by assigning reviewers that are at least in related departments. So, if it is a math course “it's not necessarily gonna be a math

person, but it might be a chemistry person, or a physics person, or something like that” (Public 3). This can help to address both online teaching behaviors as well as subject-specific feedback and instruction.

Many post-secondary institutions that offer online courses utilize adjuncts to help teach courses. One institution asserted that its process for evaluating adjuncts is no different than how they evaluate full-time faculty. One administrator explained that “there are a lot of universities out there that differentiate how adjuncts are evaluated from full-time... there is this general belief in the academy that it is OK to put adjunct faculty under this microscope and look at everything they do and be right on top of them, but somehow we should be much more hands off if you are full-time... I always call it the full-time bias, the belief that full-timers don’t have the same issues that the adjuncts do” (For profit 2). In contrast, another institution does have different evaluation processes for full-time faculty compared to full or part-time adjuncts, but the process includes more evaluation for the full-timers. “Our full-time adjuncts do not do any type of evaluation other than the evaluation that I do for them yearly and the [peer evaluation] every term that they teach” (For profit 2) whereas full-time faculty have a much more extensive evaluation process.

One institution described the effort they make in communicating these results to the instructor: “There's a lot of times that the department chair and the adjunct faculty member would actually have a live phone conversation, Skype conversation, uh something like that, to go through especially when there's concerns noted, before they actually sign the evaluation form” (Private 4). As noted in this particular account, the instructor has the opportunity to express his or her agreement with the assessment of the evaluation. Another institution explained that by leveraging the database that includes all of an instructor’s evaluation, the administrator is able to

provide feedback that is “appropriate and relevant and individual.” For example, “it actually tells ‘em, your announcements are good because... you know, your grading feedback meets all of the expectations of the department and you know, or it actually might say it exceeds the department expectations because, and then it lists all of the reasons why it’s each of those categories... And it’s, because the evaluation is for the whole year, it uses the average score for each line item for deciding exactly which comment goes into the thing” (For profit 2). This data is provided largely by the peer evaluation that had been performed during each teaching term in the year. In this way, instructors receive very personalized feedback about their performance.

Although, all of the institutions that perform administrative evaluations do them at least once every year, some perform them more frequently than this. Two institutions perform them twice a year. One institution performs them at the end of every teaching semester. These are summative evaluations that are recorded and used to make personnel decisions.

Six institutions use administrative evaluations for formative purposes. These evaluations usually are not recorded and serve to help instructors improve. These formative evaluations are often performed by the online support institution. One institution described those that perform this evaluation as a “concierge, ensuring that the instructor is really meeting their obligations preterm, during the term, and after the term, uh but also monitoring the instructors to make sure they are doing the things we asked them to do” (Private 4). One institution performs these evaluations mid-course every teaching semester. Another performs these every quarter. One does it every three years.

An additional type of formative evaluation occurs by request. The request may come from a department chair, a student advisor, or the faculty member. These are usually in instances where either something in the instructor's performance appears to be amiss or the instructor

wants to demonstrate quality teaching to a chair or department head. In these instances, the online support institution may act as “air traffic controller” connecting people for different reasons (Private 4). One example where someone from an online support institution may perform this role is in the case that a student complains to an advisor that an instructor is not meeting his or her needs in some way. In this case, someone from online support will perform an initial formative administrative evaluation. “If it's something that can easily be addressed, you know, that's worked with, directly with the instructor. Other times they have to escalate things up to the actual college whether that be the department chair or associate dean or dean, to say... I need your assistance to help get this issue addressed” (Private 4). There is no other pattern that is the same among the institutions for formative administrative evaluation. One institution of those that perform this formative administrative evaluation pointed out that they do not communicate the results of this evaluation with the instructor’s supervisors. The instructor may choose to share it if they want to bring it to the supervisor’s attention, but the online faculty support institution will not.

Institutions face a variety of challenges in performing administrative evaluations. One institution lamented that there is no common standard among institutions for evaluating online instructors. “I think that's the trickiest part for me and for my group... it's not like there's some wonderful faculty evaluation rubric available to go and observe teaching practices” (Public 3). As a result, they had to grapple with difficult questions alone and rely on their best judgment. Developing and performing administrative evaluations is difficult because it requires a lot of time and people to effectively support and carry it out. “I can only ask for so much time,” one administrator explained, “that's my biggest challenge, uh, my other big challenge is I am a very small department. So, I have, so it's me and it's three other people... I don't have a lot of, a lot of

staff to, to really handle some of this” (Public 3). Another institution described its circumstances, “we are only a team of 3 people and we have approximately almost 300 hundred faculty right now” (Private 5). One institution has one department with over 90 online instructors (For profit 2). Only one administrator performs the evaluations. It can be very difficult to perform effective administrative evaluations when there are so many instructors to evaluate and so few resources to complete the evaluations.

One institution explained that part of the problem is that “a lot of the [department] chairs do not teach online, don't want to know about it, don't care, uh, and that's kind of my next hurdle” (Public 3). Additionally, they found that many faculty were unwilling to engage in the evaluative process. Where both faculty and department leadership do not want to receive and perform evaluations, it becomes an uphill battle for those that see the value of it.

There are also conflicting values that make it difficult for online instructors to satisfy administrative expectations. One institution outlined some of its expectations of online instructors, including timeliness of grading. “[Faculty] have 72 hours to get all of their grades in every week so when you are having that much of a time crunch that's when it starts to get, you know, it gets a little bit hairy” (Private 5). Because of this “time crunch,” instructors may feel the need to cut corners in the kind of feedback they give their students. “We want expediency, but then also [instructors] are being evaluated on how substantive you are... It's kind of a catch 22” (Private 5). Instructors may struggle to meet competing expectations of the institution.

There are a variety of reasons why institutions perform administrative evaluations. Although some feel that the administrative evaluation is nothing more than “paperwork” (For profit 2), others feel that it can be a worthwhile experience that benefits the institution, the instructor, and students. The most obvious reason for administrative evaluation is to provide a

summative evaluation on an instructor's performance. These summative decisions help identify exceptional instructors “to make sure we get them in the classroom as frequently as possible” (For profit 1). Exceptional instructors can be rewarded with raises and promotions. These evaluations are often “what raises are based on” (Private 2). It can also lead to “extended term contracts. So, if you have been there X amount of time, you might get a three-year contract instead of a one-year contract” (For profit 1). Even “the adjuncts can still go up for promotion even if it's not, like, tenure” (Public 1). These instructors may also be invited to lead “faculty development workshop and to do those kinds of things because we have a much higher rate of the faculty attending if it's their peer who is in the same boat as they are” (For profit 1). Identifying instructors who are excelling can be a positive experience for the instructor as well as extremely beneficial for the institution.

Administrative evaluations can also identify instructors “who are struggling and may need additional development, support, resources” (For profit 1). At some institutions, these instructors that are not meeting institutional expectations “are actually mandated to go through faculty development training, a guided training module and they have to do that to get back on the schedule, if they don't meet expectations, they're forced to do training” (For profit 1).

Another institution explained that instructors that are falling below expectations in an administrative evaluation “need to go through our refresher training courses before they're assigned our next course to teach with us. Uh, so typically that's when there's multiple areas identified as being below expectations, and then making sure that the instructor receives a refresher on how to better address those student needs before they teach their next course” (Private 4). They explained further that “there's kind of a hidden fifth option” that instructors may experience if they fall below expectations, “which is where the observation was so poor that

an instructor is no longer allowed to teach with us. So, then an HR decision is made” (Private 4). Even if the results of an administrative evaluation identify a faculty member who is struggling, it can lead to greater support for improvement as well as improved experiences for students.

Peer evaluation. One institution reported that the frequency of peer review is “entirely up to the discretion of individual departments” (Public 2). Three institutions perform peer evaluations by request of either the department chair or the instructor. Another does it only during the first year that an instructor begins teaching online. Two institutions reported that they do peer reviews annually as part of their process. Another two institutions perform peer reviews as frequently as every teaching term or semester.

Most of these evaluations include someone visiting an active course and observing the teaching activities of the instructor. One of these institutions, however, has determined to perform these peer evaluations retroactively. They do this “so the faculty member does not know what course it is going to be in” (For profit 1). By doing this, they feel that it will lead to the instructor doing his or her best in every course rather than only the course he or she knows will be evaluated. “We want people giving their best effort all the time” (For profit 1). By doing this they can see all of the activity of the instructor throughout the course. This retroactive approach encourages an instructor to do her or his best right up through the end of the course.

Others use evaluators that teach the same course, an assistant department chair or even the person who designed the course to performs the peer evaluations. The department chair explained that in the event that assistant department chair evaluates, this individual “is considered to be a peer because she has no oversight of [faculty] officially... And she does periodically teach classes, so it’s a peer review because she is on their level” (For profit 2). One

institution uses a “trained faculty evaluator” that will visit the course and “do a deep dive in one of the courses that you taught” (For profit 1).

Rubrics allows evaluators to “give a holistic analysis where they look at what you're saying, their impression. How are you providing feedback? What are the things you are actually doing? What do your announcements look like?” (For profit 1). Whereas most institutions that have developed rubrics are fairly rigid in the way they use the rubric, one institution tells academic departments to “adapt it to your needs, you know, if this doesn't work for your clinicals, adapt it for your clinicals, you know, decide what evidence is important to you, you know, if this doesn't work for your lab course, that's fine, add the evidence that's appropriate to you” (Public 3). This is done assuming that a department chair may have no experience with online learning. In this way, they at least have guidelines provided in the rubric of what to look for in a quality online instructor.

At some institutions, peer evaluators use the self-evaluation the instructor already completed to “compare with their own assessment and evaluation. They have talking points with the instructor about how they feel they're doing” (Private 1). Using the self-evaluation in this way can create dialogue between the evaluator and the instructor. For example, one institution explained that a peer evaluator sends the instructor a list of questions that address how they are performing in the classroom to answer before they perform the evaluation. Following the self-evaluation, the peer evaluator “will write their own notes based on those [questions]... and then they'll send it back to [the instructor] for any additional comments and to, you know, ask any questions” (Private 5). This dialogue helps clarify and possibly rectify any apparent deficiencies in an instructor's behaviors in the course. It can also lead to personalized goals or improvement plans.

When peer evaluations are complete, the institutions in the sample communicate the results in different ways. The majority of these institutions (6/10) reported that instructors receive the results of the peer evaluation. So essentially “it's supposed to be just between the two of them, right. They just fill out that they've done it” (Private 2). Some of these institutions also explained that the instructor has the prerogative to determine who will and will not receive the results for the evaluation. Only one institution reported that these evaluations are shared with the department chair as a matter of course who “will incorporate that feedback into the faculty evaluation before submitting it over to the faculty member” (Private 4). Thus, the majority of these peer evaluations are intended to be kept between the instructor and the peer.

However, if the evaluation revealed unsatisfactory teaching behaviors, it may have to be communicated with the chair who may then do her or his own course visit (Private 5). A poor evaluation may also be communicated to other institutional departments. One institution will communicate poor evaluations “to coaching. So, coaching will take that to come up with kind of a bottom sort of list of, 'ok, who are our poorest performers,' and then they'll do their own course visits” (Private 1). Additionally, the results may also be communicated to employment and scheduling to ensure that the best teachers are in the classroom as often as possible. “Our scheduling system gives highest scheduling priority to our more effective faculty. So, if you are a struggling instructor you probably are not going to get scheduled as often as an instructor who is excelling” (For profit 1). So, although peer evaluations are intended to be kept only between the instructor and the peer, if the results are too poor, they may need to be communicated to other departments.

Institutions identified several challenges in performing peer evaluations. One felt that they simply not “given it serious thought to... investing the time and resources that would be

necessary for effective peer to peer evaluations” (Public 2). One institution explained that it indeed requires a lot of resources to complete. “[Peer evaluation] takes an incredible amount of time... It really does. ‘Cause [the peer] has to visit all these [instructors] and not just, you know, do a quick review, and close the class and walk away, [the peer] has to fill out this form which is all of these score-able items and then [the peer] also has a notes section that [the peer] can type in and then [the peer] writes up this big detailed email to the instructors and sends that to ‘em, and so we invest a lot of time in this” (For profit 2).

There is also some concern about the subjective nature of these evaluations, which may be a result of a rubric that is vague, an evaluator’s inattention to detail, or even the evaluator’s effort to not provide any critique of a colleague’s performance. These “love letters” as one institution called them, do little to improve teaching or identify high quality instructors. This same institution is looking for ways of “actually beefing up training and providing more resources to help faculty peer reviewers to write up something a little more useful” (Public 1). Without a clear standard to measure by, peer evaluations may continue to vary based on who is evaluating.

They perform peer evaluations to encourage and identify effective teaching. “This is less about dinging [instructors] for doing bad things,” one institution explained, “it’s more about supporting them for better teaching” (For profit 1). Peer evaluation provides a place where ineffective teaching can be identified and improved. One institution remarked that peer evaluations have led to remarkable results. “I have the data to prove that [peer evaluation] has been incredibly effective for improving instructor performance, and improving student satisfaction with the courses, and most importantly, improving student performance in the courses” (For profit 2). Essentially, the institution found that its peer evaluation is evaluating

teaching behaviors that lead to higher student satisfaction with and better performance in a course. If the rating on the peer evaluation improves, so do the other measures related to students.

This same institution also reported that peer evaluations help it identify effective teaching practices among its faculty. These instructors are then encouraged to share these practices with their faculty during their yearly conference. For example, “an instructor who does an exceptional job in the discussion board” was encouraged to submit a proposal to present in the yearly conference “based on feedback that she received from the [peer evaluation]” (For profit 2). Peer evaluation can be one way to help encourage and identify effective practices of online teaching.

Instructors at one institution in this sample request peer evaluations so “they can put it in their tenure and promotion binders” (Public 1). This is in the best interest of the instructor because she or he can demonstrate evidence of both teaching excellence and efforts to actively build skills. This is all in an effort “to encourage [instructors] to measure, to self-monitor, to report” (Public 1).

Self-evaluation. Most of the institutions in this sample (8/10) also include self-evaluations as part of their online instructor evaluation process. Just as there was wide variability among institutions for how they do peer evaluations, there is also not a consistent approach to the way these institutions perform self-evaluations. Only one institution has instructors fill out a self-evaluation towards the end of every semester they are teaching. Two institutions in the sample have instructors complete self-evaluations twice each year. The next most frequent use of self-evaluation was once per year by one institution. Two other institutions

from the sample include self-evaluations as part of their process, however, they are optional. Although instructors are encouraged to complete a self-evaluation, it is not compulsory.

One institution where the self-evaluation is optional will not perform any other evaluation unless the instructor first completes a self-evaluation. The administrator explained that the self-evaluation is “the first thing that we request, um, we do have faculty who decline doing that, and uh, going forward, we're just not going to do their review Because it's so hard to identify the evidence of whatever those items are in the course for my reviewers but it, you know, if they don't want to do a self-review, they're clearly not involved in the process, and I don't want to waste my time and I don't want to waste the time of my faculty reviewers” (Public 3). In this case, the self-evaluation indicates that an instructor is ready and willing to receive feedback and wants to engage in the process.

A self-evaluation can also provide an opportunity for instructors to set goals for improvement, but sometimes it works better in theory than in reality. One institution encourages instructors to do this each year but finds that many instructors do not do it well. “Some people really do it and they do really well and they're like, yeah I wanna do this, but a lot of people, they show up and do their evaluations when they absolutely have to which is twice a year, right, so I would say, yeah, that's pretty weak” (Private 2). Without additional encouragement or incentives, some instructors may not be very motivated to use a self-evaluation this way.

One institution has found that a self-evaluation can do more than help an instructor improve; it can also help improve course curriculum. One administrator explained, “at the end of every course, we give instructors an option, they don't have to do it, but they can submit something about the course” to explain why they may have “struggled teaching this course” (For profit 1). Instructors can then include some of their reflections about how they feel the course

could be improved. That information can then inform course designers the next time the course is adjusted to incorporate any suggestions they feel are valid.

Self-evaluations are largely unstructured opportunities for instructors to reflect on their efforts in the previous year and their goals for the upcoming year. Other institutions provide structured self-review of an instructor's efforts that a peer considers as part of his or her evaluation. It provides a good indicator as to the willingness of an instructor to receive feedback on her or his performance.

One explains that they will take the self-evaluation of the instructor "into consideration when they do that peer eval" (For profit 1). Another explains that it helps a peer to prepare "for their own assessments of the instructor" (Private 1). In some ways it provides the first half of a conversation by presenting the instructor's perspective first.

One institution, where the self-evaluation is optional, will not perform any other evaluation unless the instructor first completes a self-evaluation. The administrator explained that the self-evaluation is "the first thing that we request, um, we do have faculty who decline doing that, and uh, going forward, we're just not going to do their review because it's so hard to identify the evidence of whatever those items are in the course for my reviewers but it, you know, if they don't want to do a self-review, they're clearly not involved in the process, and I don't want to waste my time and I don't want to waste the time of my faculty reviewers" (Public 3). In this case, the self-evaluation indicates that an instructor is ready and willing to receive feedback and wants to engage in the process.

A self-evaluation can also provide an opportunity for instructors to set goals for improvement, but sometimes it works better in theory than in reality. One institution encourages instructors to do this each year but finds that many instructors do not do it well. "Some people

really do it and they do really well and they're like, yeah I wanna do this, but a lot of people, they show up and do their evaluations when the absolutely have to which is twice a year, right, so I would say, yeah, that's pretty weak" (Private 2). Without additional encouragement or incentives, some instructors may not be very motivated to use a self-evaluation this way.

One institution has found that a self-evaluation can do more than help an instructor improve. It can also help improve course curriculum. One administrator explained, "at the end of every course, we give instructors an option, they don't have to do it, but they can submit something about the course" to explain why they may have "struggled teaching this course" (For profit 1). Instructors can then include some of their reflections about how they feel the course could be improved. That information can then inform course designers the next time the course is adjusted to incorporate any suggestions they feel are valid.

Self-evaluations are largely unstructured opportunities for instructors to reflect on their efforts in the previous year and their goals for the upcoming year. Other institutions provide structured self-review of an instructor's efforts that a peer considers as part of his or her evaluation. It provides a good indicator as to the willingness of an instructor to receive feedback on her or his performance.

Metrics evaluation. One institution explained that after students have filled out surveys and evaluations on teachers, the dashboard of administrators displays "every day what the students' answers are to any of those surveys and I can see every day what they said about interacting with a certain mentor on a certain call" (Private 2). This data then becomes easy to retrieve and instructor behavior becomes easy to monitor.

One institution developed an automated program that can alert administrators and instructors when teaching behaviors fall below baseline standards of teaching effectiveness. The

institution described its dashboard as “a big program that is only going to give you the names that don't meet that threshold or whatever it might be... training and coaching, reach out to X they are struggling on this, but it is actually done through a centralized unit as opposed to in a department” (For profit 1). This same institution described an example of “one of [their] built in metrics” regarding timeliness of grading. They explained:

The moment 7 days hits, you have now moved to a dashboard that will say they have not graded their papers. The minute day 9 hits... that dashboard automatically generates an email to the instructor that says... your grades were due 2 days ago, the following is a list of students who you have not submitted a grade on, this is how many days it has been since submission... That email will then generate every single day until the faculty addresses it... By the tenth day, so three days after, you are getting an individual reach out from a faculty support staff person. And that will be the case on any expectation that we have for teaching that they haven't met. So, the same thing would happen if you have not logged into your classroom and posted for two days. (For profit 1)

They used this kind of automated evaluation on any of the items where they have established baseline standards. By using automated alerts based on metrics, instructors can receive real time feedback when they fall below standards and administrators can be aware of instructors that may need additional attention.

One institution identifies evidence of engagement when a student does more than is expected using an example of discussion board participation: “Like many universities, our students have to respond to each discussion question and then they have to put X number of replies, whatever it is on that course. Well what we look at is, how much does that instructor get

students to post beyond that baseline minimum” or another way to look at it is whether or not “students are willing to go beyond what they have to do for their grade” (For profit 1).

One institution explains that every month, faculty members “get a scorecard and it gives them the data points about the outcomes of the students they've worked with” (Private 2). This scorecard includes many types of data, including how satisfied students are with the instructor and the help they provide.

Metrics can also help to determine in a master course model whether or not student dissatisfaction is a result of the instructor’s efforts and behaviors or the course design. When several sections of a course that also use the same course design are offered, institutions can average student evaluation scores for all students taking that course and can consider how “that instructor’s specific evaluation varies from that courses average,” being sure to only compare courses with the same course design (For profit 1). In this way, institutions “don’t penalize those instructors nor does it impact their evaluation negatively because... we look at that variability from course average for those factors” (For profit 1). This is a helpful way to evaluate instructors separate from course design.

Many institutions also use metrics of student success to help evaluate an instructor’s effectiveness. One institution calls it a “success rate” and states that a faculty member must have “at least a 75% success rate for the students” in any given course (Private 3), referring to how many students pass the course. Another institution even said that evaluating instructors based on the success of their students is “one of the most important ways that we evaluate” (Private 2). They admitted that faculty may be very uncomfortable with “the fact that we... hold them accountable for their students' success” (Private 2). The scorecard faculty receives each month includes real time data about how well their students are succeeding in their courses. “If students

aren't learning, you're a part of the problem. And so, I actually think that's really important though, because there needs to be some level of expectation and accountability for faculty members to care about the students that they teach... they're successful when their students are successful” (Private 2).

One institution reasoned that it is important to be careful to use grades students receive as a way of evaluating instructors “because it was just too tempting to set up an environment where instructors would feel the need to inflate grades and we didn't want them to inflate grades, we want them to inflate learning and hopefully that will correspond with grades” (For profit 1). Alternatively, they decided to look at grade distribution instead. Rather than looking at whether grades are high or low, they decided to look to see if there is a “distribution for any given course” (For profit 1). If so, then “we should be able to predict fairly well how that distribution will fall and if you are a particularly hard instructor your distribution should be similar but lower. If you are a more generous instructor, your distribution should be similar, but higher so rather than focusing on their grades we focused on grade distribution as a component” (For profit 1). This would help an institution avoid inadvertently encouraging instructors to provide higher grades to students in hopes of improving their success rates with students.

Some institutions have determined to not only look at student success in the current course, but also on how students perform in later courses. One institution calls it “throughput” which is where a student “actually enrolls in the very next term and does not take a term or two off” (Private 4). A poor instructor may lead to a student choosing to not enroll in the next term. Another institution also felt that it was important to consider “to what extent are those students persisting and going on to the next course and then when they do go on to the next course, how successful are they?” (For profit 1). This is easier to do when the curriculum is designed by a

centralized body at the institution. Curriculum designers can then map out “what skills and abilities in one course should lead to [student] success in the next course or should impact different components of the next course... if we looked at those same metrics and we saw wide instructor variability, then we started diving into OK, what kinds of behaviors and things are the instructors doing that are allowing their student to succeed and be more successful in the ongoing courses” (For profit 1). In this way, student success in later courses can be an indicator of teaching effectiveness, but this requires careful course design and some fairly advanced metrics.

Many of the teaching behaviors that lead to better student outcomes in an online course can be automatically observed and reported through the use of metrics. Only one institution in this sample uses metrics in this way. This administrator reports “we can look at how many times an instructor posts an announcement, we can look at what days of the week they post the announcement, we can look at the times of the day when they are in the course, how many times they post in a discussion thread, the length of their discussion thread posting. The time between a student posting and an instructor responding” (For profit 1). Essentially, this type of process can help to identify when important baseline standards of behavior are not being met without requiring a peer or administrator to constantly visit a course to ensure baseline standards are met.

Challenges that institutions have faced by incorporating the use of metrics into their evaluation system primary concern faculty members. Some faculty are worried about how metrics may infringe on their “academic freedom to teach how [instructors] want and you can't use those metrics” or else you will “[reduce] them to teaching robots” (For profit 1). This same administrator explained that “people will get really hostile about how we use data analytics” (For profit 1). In an effort to calm these concerns, she countered that faculty “can do whatever they want as long as they meet baseline behaviors” (For profit 1). Another institution acknowledged

the value of using metrics to evaluate online faculty, but then quickly admitted that “we haven't gone down that political road yet” acknowledging that there would be some opposition to it (Public 1). “Traditionally, faculty don't like this idea” (For Profit 1).

This same administrator has found that another challenge of using metrics this way is that some instructors “game the system.” She explained:

So, faculty will go into their gradebook, they'll post a grade of 0 for every student with a comment that says, ‘Please forgive me, I'm late on grading, I'll update it as soon as I can.’ So, then the system sees that feedback has been given and the system is off their back. So, faculty will game it. Faculty will also pop into the discussion threads and they will post a discussion thread that literally says, you know, ‘Out for a few days, I'll try to pop in’ and then 3 days later they will pop in and say ‘Out for a few days...’ so the system only catches it if they fail to pop in.

And so, there are faculty that game the system. (For profit 1)

Whether the institution faces resistance from faculty or faculty try to trick the system into not flagging their behavior, these are difficult obstacles to the effective use of metrics as part of an online instructor evaluation process.

Institutions use metrics for similar reasons as other types of evaluation, namely, to obtain information about effectiveness of instructors. Data collected through metrics informs different departments about faculty. “All of these data points feed into our faculty scheduling system and so they are all filled in, and I don't even know the actually weighting and metric and math rubric, so any given faculty member has a number associated with their name and our scheduling system gives highest scheduling priority to our more effective faculty” (For profit 1). Faculty development also receives some of this data and will use it to “change our training as a function

of the aggregate data we are getting” (For profit 1). This can help to tailor services to the needs of faculty in both general and specific terms.

“So, our metrics, I always say, are... valuable for many things from a research and institutional deal, but in terms of faculty evaluation, they're really only valuable for identifying the really really bad faculty... Our metrics are not enough to actually improve teaching which is really the goal of the evaluation, it's not going to improve anything, it's just going to catch the slackers” (For profit 1). Another institution confirmed this idea that metrics don't “get at maybe the most important point which is the quality piece, right?” (Private 2).

The institutions in this study demonstrated many strengths in their efforts to effectively evaluate online instructors. They also did not shy away from recognizing some weaker aspects of their process. There was a general consensus among these institutions regarding what are the best practices of online instructor evaluation. These best practices of evaluation can be grouped into: 1) those that focus on administrative needs and, 2) those that focus on the needs of instructors and their efforts to teach. Both of these different areas will be discussed below utilizing what institutions feel are their weaknesses and any changes they have made approximate these best practices.

Administrative focus. In terms of operating an evaluation process at an institution, there are a variety of best practices that institutions in this sample identified. The most discussed best practice of institutions was the need to have some structure in their evaluation process but not too much. Structure suggests a regular approach to how and when they perform evaluations. In most cases, the institutions that said they had too much structure also admitted that there were other areas where structure was lacking.

Six of the institutions in this sample felt that there were aspects of their evaluation where there was too much structure. Where some felt that they needed to find a way to reduce the frequency of their evaluations, one institution actually did. This institution changed from observing faculty behavior once every two weeks to only when there was a problem. The administrator explained that they made this change because regular evaluations provide “pretty redundant information.” She continued, “the faculty typically are very consistent in how they are... once we have a good faculty member, there's no reason for us to continuously go into the classroom and hassle them with that, it just doesn't make sense” (Private 5). Although others have, likewise, wondered “maybe we could reduce frequency,” they have been more hesitant to make the change in fear that they will lose access to important data about instructor performance (Private 1).

Flexibility was another aspect of evaluation where some institutions felt that there was too much rigidity in their approach. One worried that faculty felt “a little smothered or big brothered” by the inflexibility. One administrator explained, that “for instructors that have valid pedagogical reasons to want the variability, I do think we are a little rigid and so I see that as a weakness.” She continued:

a faculty member cannot say, 'I'm going to take ten days because I have really long papers and I want to give better feedback.' There is no exception. The rule is 7 days. If you violate the rule, then you are starting to get those automated emails, and somebody is going to follow up with you. If an instructor says, you know in this particular discussion, I don't want to post because I want to hear the student's voice without me swaying them, um, there is no exception. The policy says you

post every two days and that you post at this ratio to the students post, and there isn't that individual variability. (For profit 1)

Regardless of whether or not instructors have “good pedagogical reasons to not want to follow the policy,” the institution struggled to find a way to allow for greater flexibility in the evaluation process (For profit 1).

Another institution has sought to accommodate some flexibility in how they evaluate by incorporating notes in each course directed to the evaluator. In this case however, they are allowing human evaluators to make the adaptations to the policy rather than metrics. These “course visit considerations” address unique aspects of course design to allow for some flexibility in how instructors were evaluated. These considerations include what “course-specific differences” in teaching behavior standards still “meet expectations” compared to typical courses. “For instance, some courses have approved on their course counsel certain assignments that can be graded within ten days” instead of the 7-day policy. Another course may include situations where “most of the work is done in some other McGraw-Hill tool and that's why they're not in the course as much... Or this course only has discussion boards this week and this week that the instructor needs to be part of, these discussion boards aren't part of that week” (Private 1). Therefore, one of the advantages of human evaluators compared to metrics is that it allows the use of some flexibility in evaluation.

Some institutions demonstrated flexibility by making regular changes to different aspects of the evaluation process. One reported that these changes have led to substantial benefits, “especially to our students” (Private 4). One administrator explained, “I feel like we are always changing because we are very data driven so as we discover things that are more or less predictive, we will shift specific questions, specific metrics. So, in that sense, we are always

shifting a little” (For profit 1). Even if some of these changes may be small, they can lead to big differences in the experience students can have and the improvement instructors can make.

Another aspect of excessive structure can be manifest in the rubrics that institutions develop and use. These rubrics can introduce complications into the evaluation process by being unnecessarily complicated. In developing an instrument to evaluate course design, one institution found that they “had an inter-rater reliability problem on our rubric, it was really very subjective” (Public 1). One reviewer would rate an instructor differently than another. In an effort to rectify this, administrators identified which items in their rubric could be a binary choice, either yes or no and made this change in their rubric. This helped to address the problem of subjectivity among reviewers. Similarly, another institution also made its rubric a list of questions with binary responses “for reliability purposes” (For profit 2). The department chair explained the reason for the change:

At the time that we did create that, I actually had two assistant chairs because the department was over 200 instructors at the time. And with two different people scoring on the instrument, I wanted to keep the inter-rater reliability as high as possible. And so not only is it a binary, but it also has very specific criteria that they can see, like, you know, for example when they check the discussion board it actually says that, you know, this number of items, or during unit three, or you know, it's very specific and they answer yes or no, yes or no, yes or no, all the way down. And that way, it was to try and keep [inter-rater reliability] as high as possible. Again, going back to, you know, the statistical data analysis person, I was trying to minimize variables and keep it reliable... those are the kind of things you have to be aware of, because if it's a 0 - 5 scoring system, most people

are gonna put a 3, um, and, you just, what's the difference between a 3 and a 4?...

It could be an opinion. Whereas, did it or didn't do it, that's pretty clear. (For profit 2)

Simplifying an instrument can help increase inter-rater reliability by making it a simple decision for each item on the rubric. It can make the process more streamlined and clear.

Another institution sought to rectify this problem by adjusting its own observation instrument. The instrument included items about different teaching behaviors that were evaluated on a 7-point scale. This created problems both in terms of how instructors responded as well as how supervisors evaluated them. "It seemed like everyone was getting really, uh, I don't know how to say it nicely, uh, 'hyper' about points on a scale that didn't really seem to matter, that didn't really seem to make that much difference, but from the instructor perspective was like super paranoid about it" (Private 1). They felt that the best way to improve it was to simplify it by reducing the 7-point scale to either "exceeds, meets, or below" expectations. For everybody involved, "it just seemed clearer" (Private 1).

While most of the institutions acknowledged that too much structure is a problem, five institutions also expressed that they felt that that there was not sufficient structure in aspects of their current processes to administer effective evaluation. One in particular felt that its evaluations were far too reactive and not proactive enough to identify problems. The administrator felt that because of the absence of a formalized process for evaluating that they were "relying more so on outside sources to get the information instead of us being able to go in ourselves and see that our faculty members are doing a really good job" (Private 5). Without the structure, administrators may struggle to have a good sense of how instructors are performing.

One institution felt that they were not performing evaluations frequently enough to provide instructors with formative feedback. They felt that “the annual evaluation... is not always acceptable in my mind if you will, especially when there's identified challenges with an instructor, I think it needs to be a more frequent evaluation, whether that's a semi-annual or a quarterly if they teach throughout the entire year... even if it's an informal evaluation” (Private 4). He hastened to add that this particular informal evaluation would “not go into their permanent record, uh but at least identify and make sure that those things that have been identified as a deficiency are being addressed and that we are seeing the gains in that” (Private 4). Thus, greater frequency of informal evaluation can help instructors regularly work on and seek improvement.

A number of the institutions in this study felt that utilizing triangulation of evaluation was something that they were doing well. They felt that each type of evaluation provided a different perspective and different information regarding an instructor's performance. This was the only way in which some institutions felt there was sufficient structure in their process.

Others felt that they were not doing enough triangulation in their evaluation. They felt they needed “more peer evaluations” as well as “a self-evaluation” (Private 4). By utilizing these two types of evaluations, one administrator felt they might better answer questions like, “what is the perception of the instructor and how they're doing in the classroom? How is that really aligning with what the students are saying and what we are observing from the administrative side of the operations?” (Private 4). Another felt that a weakness was that there was no way to observe “faculty behavior on a day-to-day basis within their course... we don't have a solid rubric to evaluate teacher activity once the courses began, nor do we have a regular, consistent process for having a peer evaluation of that” (Public 1). By including more data, retrieved from

different sources, administrators can obtain a clearer picture of the effectiveness of their instructors.

Efficiency was another best practice of institutions. One institution talked about its struggle to be efficient. They considered many other possible ways they could approach evaluating online instructors. They wondered whether or not they should just utilize the results from the student evaluation “and maybe some other flags of stuff we needed to worry about, flags from the actual LMS and deal with hot button issues and not, you know, maybe we don't need to be so evaluative” (Private 1). They wanted to find an efficient way “to see what's going on in a classroom relative to an instructor's performance without it taking a ton of time” (Private 1). It seems that some of the greatest obstacles to efficiency is how long it takes to evaluate instructors and how many people are necessary to implement the process.

Another institution explained that they feel that efficiency has become one of the strengths of its current evaluation process. As a result, instructors “are getting feedback all the time... it really doesn't make sense oftentimes to wait until the end of a semester... faculty aren't waiting 8 weeks for a semester to get info on how they are doing. They can get a lot of here and now information, and we have the support there to help them when they struggle” (For profit 1). By its own admission, a lot of this institution's success has come from learning to harness the data available in the LMS and other sources to inform instructor effectiveness. It has been a 7-year process that included:

building dashboards, building systems so that data isn't meaningless... we have been very intentional with trying to go beyond evaluation data simply to evaluate that individual faculty member and use it to inform us. So, we really... look at the data not just from a person, but from all the people and see how we can provide

better training, how can we provide better support, how can we meet our students' needs better. So, we have really done a lot with making sure we don't have meaningless data that is collected, but not used for anything. And kind of streamlining too, deciding what kinds of things we don't need to collect anymore because nobody is using it. If no one uses it, why are we collecting it and storing it, so we have also eliminated some data points that just weren't overly meaningful. (For Profit 1)

Evidently, leveraging metrics can increase efficiency, but will require extensive work and data analytics to determine what data is useful and worth collecting and what is not.

One institution felt this extensive process would be worth it. They wanted to do more to harness the LMS data available at any given time. They hoped “to do more data analysis. Building it whole from the back end of the system... to make sure that our students are getting the quality experience that they expect and they deserve in our online classes” (Private 4). Metrics may be able to help to not only make an evaluation process more efficient but can also lead to a better experience for students.

Closely tied to efficiency was the idea that the evaluation process should be scalable, meaning that the process could easily accommodate the growth of the institution. Most institutions that mentioned scalability addressed it as a weakness. One institution described its evaluation system as “a really complicated process that's not so scalable... so we've been talking about both simplifying that process and trying to find more efficient, more effective ways to provide that evaluation (Private 1). In order to accommodate growth, it may be important to simplify the process and make it more efficient.

It seems the biggest obstacle to being scalable is finding out how to evaluate a lot of online faculty without needing a lot of people to do it. One institution admitted, “we are so large” that the limited resources make it very difficult to roll out an evaluation process for the entire institution (Public 2). In referring to the system of evaluation, another administrator lamented, “it takes a lot of manpower to do that kind of evaluations” (Private 1). One institution has to leverage volunteers in order to have enough people to perform the evaluations. The administrator explained that “the big challenge is just getting faculty to serve on my committee. So, I have to beg people to be on this committee” (Public 3). Another big problem facing institutions, therefore, is having enough people who are sufficiently motivated to help improve teaching effectiveness through evaluation.

Only one institution felt justified in declaring “We are scalable and consistent” (For profit 1). The administrator largely felt that the evaluation process was scalable because of how it leveraged metrics in evaluations. Clear and defensible standards helped too. “It is very laid out for [instructors] what they need to do. What we expect of them is absolutely supported in best practice research, we can show them data that's why we want them to do X” (For profit 1). Utilizing metrics and providing clear, evidence-based standards can help an evaluation system become more scalable and consistent.

From an administrative focus, in order to have perform online instructor evaluations effectively, the evaluation system need to have some structure, but too much structure can also lead to problems. This structure needs to include semi-frequent evaluations as well. Using triangulation institutions can provide more accurate information regarding instructor’s effectiveness using a variety of evaluation sources. This process ought to be efficient as well,

leveraging any resources available to reduce the amount of time and people required. This can also lead to a system that is more scalable that can adapt to an institution's growth.

Teaching focus. Other best practices of evaluation processes that institutions utilized addressed instructors and their teaching. Of these, several institutions emphasized the importance of instructors feeling trusted. It requires a delicate balance between having an appropriate amount of structure to evaluate instructors and an excessive amount that leads to instructors feeling like they are not trusted. One institution worries that its current evaluation process is too demoralizing for the instructors. The administrator worried that it has impacted "instructor morale. It's just like, 'I'm always being assessed. I'm always being evaluated.'... It's still all seen as 'evaluation, evaluation, evaluation, evaluation,' all the time. So, I feel like at the beginning when we started out, campus was asking for all of this because there was kind of a lack of trust in remote instructors you never see, so they want all this data, and yet I don't think it's been good for instructor morale" (Private 1). An inordinate amount of structure appears to lead instructors to feel like they may not be trusted.

One administrator who valued instructors feeling trusted more than having a structured system for regular evaluation explained "that the faculty feel like their classroom is like, 'this is my domain' you know so to speak, and it feels very intrusive when administrators are constantly needing to go into the classroom, to know everything's fine. I think at some point they need to feel like that there is trust between the administrators and the faculty. We need to have some type of trust there" (Private 5). This same administrator also emphasized the importance of striving to bring attention to the great things instructors are doing with their students. This can also help to demonstrate trust in instructors.

Two institutions felt that when they started developing their online evaluation system that there had been a lack of focus on what instructors were doing well. They determined that they “were gonna let the instructors know, this is what you’re doing really well, this is what you need to, you know ha! These are opportunities, we didn’t even want to say to them, this is what you need to fix, we didn’t want to approach it that way. We wanted to instead approach it from a very positive reinforcement way” (For profit 2). This focus on positive behaviors rather than negative behaviors can help improve morale as well as lead instructors to look forward to their evaluations, rather than fearing or dreading them.

Most institutions felt that evaluations should focus on teaching behaviors within the instructor's control and not only on course design. One institution felt that its evaluations overemphasized the course design. The administrator admitted that the current process for online instructor evaluation “falls a little short... in being pretty general and not always giving us an idea of faculty behavior on a day-to-day basis within their course” (Public 1). If the evaluation focuses on course design, an instructor that did not design the course will be evaluated on the merits and quality of the course rather than his or her own efforts to engage and interact with students.

Other institutions made changes to their evaluation practices to focus more on teaching behaviors. One institution explained how it identified what aspects should be included in the observation rubric with, “It actually started as more of a checklist and then it evolved into more of a rubric observation so that we could identify what are some of the things that our really stellar teachers are doing to be successful with students and then we put it in a rubric” (Private 2). Another institution focused on teaching behaviors that lead to student engagement. They sought to incorporate “effective pedagogy in the online environment” (Private 4). Some of these

things included “the timeliness of the feedback, the quality of the feedback... We also put a little bit more of a focus on participation in the online discussion forums so that we were getting that student faculty engagement piece taking place that way” (Private 4). One institution made adjustments to the evaluation rubric, which was only focused on course design, and added twelve additional items that focused on teaching behaviors using existing literature. They “took from a variety of the other rubrics that were out there at the time and looked at the research and went through and we figured out all these items and put it together” (Public 3). By observing teaching practices and consulting research, administrators were able to improve the rubric.

In order to ensure that the rubric really did focus on teaching behaviors that led to greater student engagement, one institution utilized the data collected using the peer evaluation rubric and performed a longitudinal study. Administrators at this institution sought to compare the teaching behaviors they were evaluating with the performance of students in those classes. They were very pleased with the results. They found that they “were able to correlate [the teaching behaviors from the rubric] to improved student performance in the classroom which was incredible... with over a thousand data points. And, uh, 99-point-something percent correlation” (For profit 2). Additionally, they found that if they improved the monitoring score, which was the result of the peer evaluation, “the end of term survey score did also go up” (For profit 2). In other words, if an instructor performed well in a peer evaluation, that score was positively correlated with the end-of-course student evaluation score. They were able to confirm with metrics what others had found through observation and research.

As an institution increases its focus on teaching behaviors, it may lead to instructors feeling concerned about their performance, especially if they are not used to receiving feedback. One institution reported that when they first started “we got a lot of push back... the instructors

freaked out. And I got 10,000 emails saying, what is, what's going wrong? Why am I getting an email? Is my performance not what it's supposed to be?" (For profit 2). In an effort to calm concerns, the institution held "a faculty meeting saying, don't freak out, this is a new thing, we're gonna be letting you know every time we visit your class... we made them well aware that 'you will get feedback now every single time you teach a course you're gonna get feedback.' And once they got used to that idea, they really, really bought into the program and they actually really love it" (For profit 2). Although there was initial resistance to this new process of evaluations that focused on teaching behaviors, it became a positive process for instructors. They began to really appreciate the regular feedback they received on their teaching.

Although there are not as many best practices that focus on the teaching perspective, those that have been described here are vital to instructors feeling valued in the process and feeling motivated to better serve their students. It is important that instructors feel trusted by those performing the evaluation. Too much evaluation can negatively impact how trusted instructors feel. Administrators can find ways to foster feelings of trust, including highlighting the great things instructors are doing and not only those things in which they need to improve. It is also important that evaluations focus on those things that instructors can control, including their teaching behaviors. It may mean that institutions need to make adjustments to their rubrics in order to focus more on these teaching behaviors.

ARTICLE 3

Online Teaching Competencies in Post-Secondary Observational Rubrics:

What Are Institutions Evaluating?

Jonathan E. Thomas

Charles R. Graham

Brigham Young University

Abstract

Evaluation of online instructors is a developing field of research. Institutions have made great progress in recent years to refine their efforts to evaluate online teaching in order to improve online teaching. Some institutions, however, still struggle to evaluate online instructors who did not design the course they are teaching. The purpose of this study is to identify some of the most important observable teaching behaviors identified in evaluation rubrics and how these compare to online teaching competencies outlined by Bigatel, Ragan, Kennan, May, and Redmond, (2012). Findings indicate that there is a growing consensus of what are the most important, observable teaching behaviors and that this can help improve existing models of online teaching competencies.

Keywords: online courses, faculty evaluation, evaluation methods, evaluation research, evaluation rubrics, online teaching competencies

Online Teaching Competencies in Post-Secondary Observational Rubrics:

What Are Institutions Evaluating?

Enrollment in online courses at colleges and universities continues to increase, even while all enrollment in post-secondary institutions is decreasing (Allen & Seaman, 2017). Since their inception, alternatives to face-to-face courses have faced criticism and charges of low quality, much of which was probably warranted (Allen & Seaman, 2003). However, a recent meta-analysis has demonstrated that much of the research over the last 20 years is inconclusive regarding whether or not there is a difference in quality, in terms of satisfaction and academic achievement (Smith & Macdonald, 2015).

Others have documented that online instructors typically receive lower student ratings than their face-to-face counterparts (Young & Duncan, 2014; Loveland, 2007). There are a variety of factors that may contribute to this decrease, based on modality. Some have suggested that this may be a result of using a student evaluation instrument designed for face-to-face courses to evaluate online teaching (Loveland, 2007). While this may be possible, others have countered this idea demonstrating that the same instruments can be used regardless of the modality (Dziuban & Moskal, 2011).

Lower scores may be the result of inadequate training and support of online instructors. Teaching in an online environment requires different competencies than teaching face-to-face (Creasman, 2012). Consequently, instructors who receive high ratings in face-to-face courses may not necessarily receive equally high ratings in online courses simply because there may be skills that they have not yet developed that are unique to the online environment. Additional training and support can help instructors to develop these additional competencies (Wilson & Stacey, 2004).

By establishing and implementing a full and well-rounded evaluation process, institutions can help support instructors in their efforts to teach effectively and inform training to improve in those efforts (ASCCC, 2013; Dana, Havens, Hochanadel, & Phillips, 2010; DeCosta, Bergquist, & Holbeck, 2015; Mandernach, Donnelly, Dailey, & Schulte, 2005; Palloff & Pratt, 2008).

Unfortunately, evaluation of teaching in online courses has not kept pace with the rapid development of online programs. Berk (2013) noted that “evaluation of these online courses and the faculty who teach them lags far behind in terms of available measures, quality of measures, and delivery systems” (p. 141). It has also been observed that evaluation of online teaching often focuses only on course design (Piña & Bohn, 2014; Thomas & Graham, 2017). This is problematic because many institutions utilize a master course model where the instructor of a course is not responsible for course design but only the facilitation of the course (Cheski & Muller, 2010; Hill, 2012). More research is needed to inform and improve evaluation practices of online instructors that are only responsible for course facilitation.

Review of Literature

Some of the institutions that are performing evaluations may be making assumptions that are not wholly supported by data. Research that focuses on online teacher evaluation is based on one of the following assumptions: (a) Traditional courses and online courses are similar enough that they can be evaluated with the same evaluation instruments and system, (b) online courses are different enough from traditional courses that they require different instruments and methods for evaluation, or (c) some combination of the two.

Similarities and Differences of Evaluating Online and Traditional Courses

In developing evaluation systems for online courses, some feel that online courses could be evaluated in the same way as traditional courses. For example, Cordeiro and Muraoka (2015) described the process of adapting an undergraduate business program to an online format. As part of this process, they explained that they saw no need to alter the evaluation system already in place for traditional classes. They used the same student evaluation instrument as courses concluded, and also maintained their practice of doing a “classroom visitation” in which they observed the instructor’s behavior in the course for a designated period of time (2015, p. 6). They mentioned that they made some modifications to this peer review process, but few other changes were made to address online learning as being different from traditional courses.

There has been a great deal of research that focuses on student evaluation instruments. Many researchers have concluded that it is appropriate to use the same instrument in face-to-face and online courses (Benton & Cashin, 2012; Benton, Webster, Gross, & Pallett, 2010). Some have called this practice into question because instructors who teach online typically receive lower student evaluation ratings (Loveland, 2007; Terry, 2007; Stanišić’ Stojic’ et al., 2014). A study by Loveland (2007) suggested that the lower ratings may indicate that a face-to-face instrument does not accurately evaluate online teaching. Studies by Dziuban and Moskal (2011) as well as Moskal, Dziuban, and Hartman (2013) reasserted that the aspects of teaching that students can evaluate are similar enough in all modalities, including blended learning, that the same instrument can be used.

When designing an evaluation process, Berk (2013) identified seven different options an institution may adopt to evaluate online instructors, including using traditional face-to-face evaluations. He also acknowledged that online and face-to-face courses may be more similar

than different. However, Berk (2013) also included the option of recreating entirely new instruments to evaluate online instructors. He recognized that there are aspects of instructor behaviors that may be different in an online course compared to a traditional course. He made insightful recommendations while ultimately leaving the decision to be made by each respective institution. In particular, Berk suggested that a new instrument ought to be developed for peer or self-evaluation that is specific to online instructors. Many feel that colleagues are better equipped to evaluate teaching effectiveness than students (Darling, 2012), and this may even be more true in online courses where additional competencies are necessary to be an effective instructor.

Focusing Evaluations on Instructional Behaviors Instead of Course Design

Piña and Bohn (2014) noted that there is heavy emphasis on course design among post-secondary online instructor evaluation rubrics. Focusing on course design is an important aspect of evaluating online instructors. However, many institutions are developing online programs that utilize a master course model (Cheski & Muller, 2010; Hill, 2012). In this model, the instructors that teach the course are not those who designed it. It is important that in these cases, instructors are evaluated on their instructional behaviors and not course design. Careful development of evaluations is necessary to ensure that an instructor is evaluated on his or her instructional behaviors rather than course design.

There are several examples of focusing evaluations on teaching behaviors rather than course design. Of note is the work of Mandernach et al. (2005), Schulte (2009), and Schulte, Dennis, Eskey, Taylor, and Zeng (2012) and Weschke and Canipe (2010).

Online Instructor Competencies

A variety of competencies of online teaching that inform best practices of online instructors exist (Smith, 2005; Darabi, Sikorski, & Harvey 2006; Varvel, 2007). It is also clear by doing a review of these lists of competencies that there are still some inconsistencies among the competencies identified in these models that focus on online instructors. This should come as no surprise considering that the contexts within which online learning exists vary greatly. Baran, Correia, and Thompson (2011) explained that “the prioritization of the roles and competencies of online teachers varies in the literature depending upon the context where online teaching takes place” (p. 427). In these varied contexts, different competencies are necessary.

After a thorough analysis of online instructor competencies, we found that most of these competencies focus primarily on course design (Thomas & Graham, 2017). Bigatel, Ragan, Kennan, May, and Redmond (2012), however, developed a list of online teaching competencies that only focus on teaching behaviors. They explained that “any tasks related to instructional design were excluded” (p. 64). They developed a survey based on a literature review and extensive discussions with online learning professionals (instructors, designers, evaluators, researchers) that included 64 statements of online teaching behaviors they called tasks, performed by effective online instructors. This survey was completed by 197 respondents. They were asked to rank each task on a 7-point Likert scale according to which tasks they felt were most important to online teaching during course delivery. Using exploratory factor analysis, they grouped the tasks into seven competencies. These competencies include: (1) active learning, (2) administration/leadership, (3) active teaching, (4) multimedia technology, (5) classroom decorum, (6) technological competence, and (7) policy enforcement.

The Bigatel et al., (2012) model is the only model we found in our literature review that specifically outlines the teaching behaviors of instructors during course delivery that does not focus on course design. For this reason, we have determined to use this model as a basis for our research. We recognize that there may be limitations in this model and anticipate that by checking the validity of this model as part of our study, we can either confirm it as an accurate model, or recommend ways to improve it.

Research Question

The purpose of this study is to explore current evaluation practices of post-secondary institutions through addressing the following research question:

- To what extent do online teaching evaluation rubrics from institutions with a master-course model address the online teaching competencies identified by Bigatel et al., (2012)?
 - a. How well are each of the online teaching competencies and the associated tasks represented in the rubrics?
 - b. Are there any items found in the rubrics that do not address tasks within the Bigatel et. al. model?

Methodology

In order to accomplish this purpose and answer this research question, we performed a directed content analysis. This is an appropriate method when researchers seek to support or extend existing theory (Hsieh & Shannon, 2005). We will describe in further detail our sample as well as our mode of data collection and analysis.

Sample

We identified all post-secondary institutions that grant four-year undergraduate degrees, reside in the United States, and have more than 10,000 enrolled distance education students using a report entitled, *Online Report Card: Tracking Online Education in the United States* (Allen & Seaman, 2017). Through this process, we identified 48 institutions. We contacted administrators of these programs to first find out if they utilized a master course model, and second, if they would be willing to share their observation instrument and provide additional context through interviews as needed. We continued contacting administrators until we identified at least two institutions in each of the following institutional categories: for profit institutions, private institutions, and public institutions. We felt that by including members of each of these institutional categories in our sample, we could obtain a more complete representation of current evaluation practices. Using purposive sampling in this way has also been used in other published research and accordingly has been deemed as acceptable practice (Graham, Woodfield, & Harrison, 2012). Our final sample included two for-profit institutions, three private institutions, and two public institutions.

Data Collection, Coding, and Analysis

The artifacts collected for this study are rubrics used as part of an administrative or peer evaluation. Table 1 outlines the types of institutions in our sample and the type of rubric we obtained from them:

Table 1

Peer and Administrative Evaluation Rubrics Collected by Type of Institution

Institution	Peer Evaluation Rubric	Administrative Evaluation Rubric
For profit 1	X	
For profit 2	X	
Private 1	X	X*
Private 2		X
Private 3	X	
Public 1	X	X
Public 2	X	

*Same as peer rubric.

Each of the seven competencies outlined by Bigatel et al., (2012) are further described by a list of tasks of successful online teaching. These tasks provide rich descriptions that further define the competency. These tasks provided the a priori categories for coding and are listed in Table 2.

Table 2

Competencies for Online Teaching Success

Competency	Tasks
1. Active Learning	<ul style="list-style-type: none"> • 1.1 The instructor encourages students to interact with each other by assigning team tasks and projects, where appropriate. • 1.2 The instructor includes group/team assignments where appropriate. • 1.3 The instructor encourages students to share their knowledge and expertise with the learning community. • 1.4 The instructor encourages students to participate in discussion forums, where appropriate. • 1.5 The instructor provides opportunities for hands-on practice so that students can apply learned knowledge to the real-world. • 1.6 The instructor provides additional resources that encourage students to go deeper into the content of the course. • 1.7 The instructor encourages student-generated content as appropriate. • 1.8 The instructor facilitates learning activities that help students construct explanations/solutions. • 1.9 The instructor uses peer assessment in his/her assessment of student work, where appropriate. • 1.10 The instructor shows respect to students in his/her communications with them.
2. Administration/Leadership	<ul style="list-style-type: none"> • 2.1 The instructor makes grading visible for student tracking purposes. • 2.2 The instructor clearly communicates expected student behaviors. • 2.3 The instructor is proficient in the chosen course management system (CMS). • 2.4 The instructor adheres to the university's policies regarding the Federal Educational Rights & Privacy Act (FERPA). • 2.5 The instructor integrates the use of technology that is meaningful and relevant to students.
3. Active Teaching/Responsiveness	<ul style="list-style-type: none"> • 3.1 The instructor provides prompt, helpful feedback on assignments and exams that enhances learning. • 3.2 The instructor provides clear, detailed feedback on assignments and exams that enhances the learning experience. • 3.3 The instructor shows caring and concern that students are learning the course content.

- 3.4 The instructor helps keep the course participants on task.
 - 3.5 The instructor uses appropriate strategies to manage the online workload.
4. Multimedia Technology
- 4.1 The instructor uses a variety of multimedia technologies to achieve course objectives.
 - 4.2 The instructor uses multimedia technologies that are appropriate for the learning activities.
5. Classroom Decorum
- 5.1 The instructor helps students resolve conflicts that arise in collaborative teamwork.
 - 5.2 The instructor resolves conflicts when they arise in teamwork/group assignments.
 - 5.3 The instructor can effectively manage the course communications by providing a good model of expected behavior for all course communication.
 - 5.4 The instructor identifies areas of potential conflict within the course.
6. Technological Competence
- 6.1 The instructor is proficient with the technologies used in the online classroom.
 - 6.2 The instructor is confident with the technology used in the course.
7. Policy Enforcement
- 7.1 The instructor monitors students' adherence to policies on plagiarism.
 - 7.2 The instructor monitors students' adherence to Academic Integrity policies and procedures.

* Bigatel, P. M., Ragan, L. C., Kennan, S., May, J., & Redmond, B. F. (2012). The identification of competencies for online teaching success. *Journal of Asynchronous Learning Network*, 16(1), 59–78.

Two trained researchers independently coded each of the rubric items using the tasks listed in Table 2 as a priori codes. A rubric item was sometimes coded to more than one task. In the case that a rubric item did not address one of these tasks, we coded it as “Other.” To determine percent agreement among researchers, we used the following formula based on the formula developed by Holsti (1965): Percent Agreement = $2m/(n1 + n2)$ (m = number of coding

decisions agreed upon by the two coders; $n1$ = number of coding decisions made by rater 1; $n2$ = number of coding decisions made by rater 2). The initial percent agreement after the first phase of coding was 76%. The researchers discussed all coding decisions that did not agree. After discussions, researchers arrived at 100% agreement on all coding decisions. Examples of coded rubric items are listed in Table 3.

Table 3

Examples of Coded Rubric Items

Competency	Task	Example of Coded Items
1. Active Learning	1.3	Solicited input and challenged students via comments, questions and forums [pull down selection rating with box for open ended comments]
	1.4	Promoted and participated in Canvas Discussion Board Forums. Where discussion topics permitted commented on a minimum of 1/3 rd of the initial discussion threads [pull down selection rating with box for open ended comments]
	1.10	Addressed student by name. Verified using 3 items in each Gradebook category: 1 perfect score, 1 zero score, and 1 other score [Select 1 for yes, 0 for no, mark only one oval] Course Communication is open, ongoing and nonjudgmental.
2. Administration/Leadership	2.2	The course grading policy is stated clearly. Specific and descriptive criteria are provided for the evaluation of learner' work and are tied to the course grading policy.
3. Active Teaching/Responsiveness	3.1, 3.2	<u>Exceeds Expectations:</u> Students receive frequent, substantive feedback that invites them to apply learning or shows other personal connections and insight in addition to basic instruction. <u>Meets Expectation:</u> Each student receives weekly feedback on assignments, as often as the course permits. Feedback left for students offers specific instruction about what has been done well, and how to improve. Feedback is relevant to the student and assignment objectives

		<u>Falls Below Expectations:</u> Feedback left for students is non-existent, minimal, generic, and/or irrelevant to student or lesson objectives. [radio button next to each option, open field for additional comments]
	3.3	Monitors student understanding of course materials through careful review of discussions and assignments.
4. Multimedia Technology	4.1	Utilizes a variety of instructional materials including charts, graphs, power points, videos, etc. [open field for comments]
5. Classroom Decorum	5.3	Quality discussion “model” provided to the students. Modeling of good discussion participation practices by the instructor.
6. Technological Competence	6.1	Possessed and utilized the requisite technology during the conduct of their courses [pull down selection rating with box for open ended comments]
7. Policy Enforcement	7.1	Any single assessment that comprises 15% or more of the total grade uses appropriate security measures such as plagiarism detection and/or proctoring services.

Findings and Discussion

In this section we will address our findings in response to our single, over-arching research question: To what extent do online teaching evaluation rubrics from institutions with a master-course model address the online teaching competencies identified by Bigatel et al., (2012)? We have divided this question into two sub-questions: (1) How strongly are each of the online teaching competencies and the associated tasks represented in the rubrics? (2) Are there any items found in the rubrics that do not fit or address the tasks in the Bigatel et. al. model?

After providing a general response to our over-arching question, we will address the focused research questions specifically.

After coding and analyzing the data we found that the rubrics in this sample, collectively, address all of the online teaching competencies listed in Bigatel et al. (2012) to varying degrees. Table 4 displays our findings by rubric. All coding decisions for each rubric are listed by competency and includes the percentage of coding decisions that address each competency.

Table 4

The Number and Percentage of Coding Decisions in Each Rubric that Address Each

Competency

Competency	Rubrics							
	FP1 # (%)	FP2 # (%)	Prv1 # (%)	Prv2 # (%)	Prv3 # (%)	Pub1A # (%)	Pub1P # (%)	Pub2 # (%)
1. Active Learning	6 (23)	3 (11)	9 (19)	5 (11)	15 (32)	8 (14)	26 (33)	29 (33)
2. Administration/ Leadership	2 (9)	2 (7)	2 (4)	5 (11)	3 (6)	14 (24)	10 (13)	11 (13)
3. Active Teaching/ Responsiveness	6 (23)	10 (36)	14 (30)	7 (15)	7 (15)	3 (5)	13 (16)	12 (14)
4. Multimedia Technology	0 (0)	0 (0)	0 (0)	2 (4)	3 (6)	3 (5)	2 (3)	0 (0)
5. Classroom Decorum	3 (12)	6 (21)	4 (9)	4 (9)	4 (9)	2 (3)	5 (6)	9 (10)
6. Technological Competence	0 (0)	0 (0)	0 (0)	7 (15)	2 (4)	0 (0)	0 (0)	0 (0)
7. Policy Enforcement	1 (4)	0 (0)	2 (4)	2 (4)	0 (0)	5 (8)	1 (1)	0 (0)
8. Other	8 (31)	7 (25)	16 (34)	14 (30)	13 (28)	24 (41)	23 (29)	27 (31)
Total	26	28	47	46	47	59	80	88

Note: FP=For Profit, Prv=Private, Pub1A=Public Administrative, Pub1P=Public Peer.

As illustrated in Figure 1, 69% of the rubric items addressed these 7 competencies. Of note is that 31% of the rubric items were coded as “Other.” These rubric items did not address the tasks associated with these competencies of online teaching identified by Bigatel et al.

(2012). We categorized these items into five task statements and one miscellaneous category.

These categories can be found in Table 5.

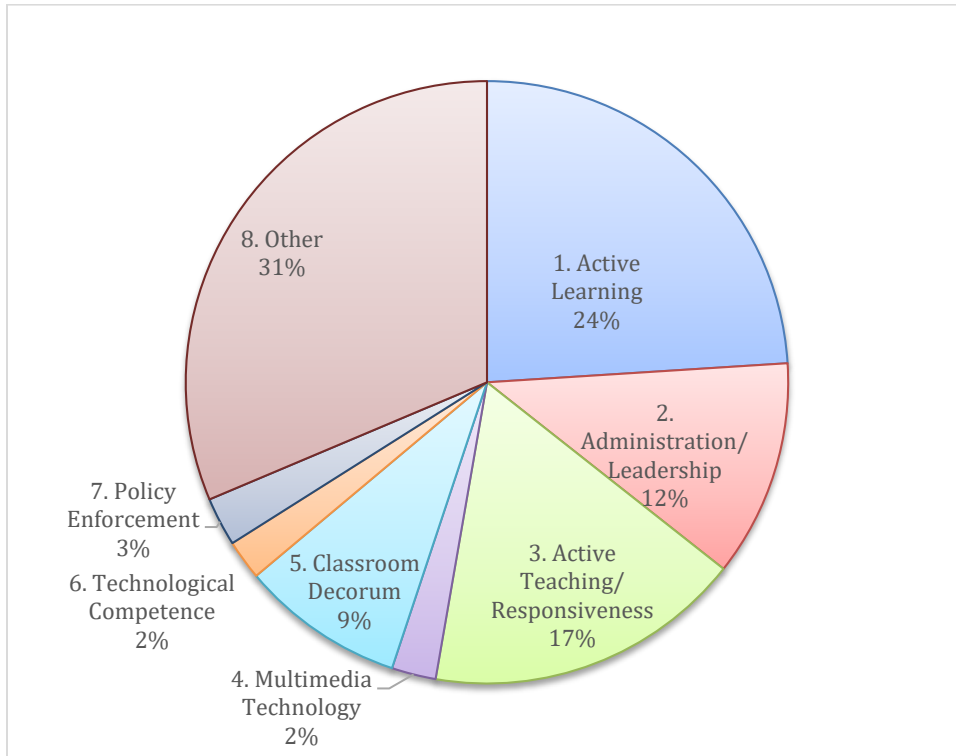


Figure 1. Percentage of total coding decisions (N=421) in eight online teaching observation rubrics coded to competencies outlined by Bigatel et al. (2012).

Table 5

Other Category Coding Groups and Rubric Item Examples

Coding Groups	Example of a Rubric Item
8.1 The faculty employs effective course design	The course learning objectives, or course/program competencies, describe outcomes that are measurable.
8.2 The faculty helps to build a learning community	The Faculty engages in the dialogue and conversation with and between students.
8.3 The faculty adapts course content to diverse needs of students as necessary	Accommodations are proactively offered for students with disabilities.
8.4 The faculty clearly explains course navigation and assignment purposes	Learners are introduced to the purpose and structure of the course.
8.5 The faculty demonstrates expertise, interest, and enthusiasm in subject matter and connects students to it in meaningful ways	Demonstrates interest and enthusiasm about content area.
8.6 Miscellaneous	Weekly announcement meets minimum length (5+ sentences).

We have also created two tables that organized the data by tasks. We have listed in Table 6 the coding results organized by task within each competency. This table displays the percentage of the total number of coding decisions within all the rubrics as well as within each competency that address each task. The final column includes the rank of each task based on the percentage of codes. This table also includes the breakdown of all of the categories organized in “Other.” During the analysis, we noticed that there were significant differences in how different types of institutions addressed the competencies. So, we created Table 7 to organize the data by rubric based on the type of institution.

We used this data to help address how strongly the online teaching competencies of online teaching listed in Bigatel et al., (2012) are represented in the rubrics of our sample. We will then address any rubric items that are not represented in the Bigatel model.

How Strongly are Each of the Online Teaching Competencies and the Associated Tasks Represented in the Rubrics?

Four of the seven competencies are represented in each of the rubrics. These include Active Learning, Active Teaching/ Responsiveness, Administration/ Leadership, and Classroom Decorum. In most of these cases, the high presence of these competencies is due to one or two tasks that are emphasized. There is often not broad coverage among the different tasks associated with the competencies.

Three competencies are not emphasized in all of the rubrics. These include Multimedia Technology, Technological Competence, and Policy Enforcement. We will describe how these competencies were represented in the rubrics as well as some of the more prominent tasks associated with each competency.

Active learning. The most prominent competency found in the rubrics was Active Learning. It represented 24% of all coding decisions (see Figure 1). Active Learning focuses on tasks that make learning more student-centered. Tasks 1.4 and 1.3 were the most emphasized tasks in this competency. Task 1.4 (encourages discussion participation) overtly addresses how an instructor encourages students to participate on a discussion board. Task 1.3 is that “the instructor encourages students to share their knowledge and expertise with the learning community.” In an online course, opportunities to share knowledge and expertise are usually provided on a discussion board.

Task 1.10 (respectful communication) was heavily emphasized by for profit institutions by almost 10% of coding decisions in the rubrics (see Table 7). It is that an instructor is respectful in all interactions with her or his students. Some rubric items addressed this task in a very general way such as an instructor using a “proper” tone or making efforts to establish a

“positive” environment. Other rubric items addressed this task by being more specific such as looking to see if instructors addressed students by name, avoided “derogatory references or sarcasm,” and used “nonjudgmental communication.”

Active teaching/responsiveness. Active Teaching/Responsiveness was the next most emphasized competency among all rubrics. It represented 17% of all coded decisions. This competency focuses on an instructor’s efforts to connect students to course content. The task for 3.3 (concern about learning content) ranked 5 among all tasks. Rubric items that addressed this task focused on seeking to connect students to the learning objectives of the course. Items may have asked if the instructor connected students to learning outcomes in the weekly announcements or perhaps in the feedback the instructor provided on an assignment. Very few of these items clearly addressed the idea of an instructor showing care and concern with regard to a student learning course content. It seemed more implied than overtly demonstrated. Tasks 3.1 and 3.2 were ranked 9 and 10 respectively. They are similar in that they both address the feedback an instructor gives. 3.1 addresses the idea of prompt feedback that is timely and regular whereas 3.2 addresses feedback that is clear and substantive. All rubrics addressed both of these items although some combined the two tasks into a single item.

Administrative/leadership. The third most emphasized competency was Administrative/Leadership. This was largely because public institutions heavily emphasized this competency in their rubric items. Of all coding decisions for this competency, 72% were from public rubrics (see Figure 2). In fact, the administrative rubric of Public 1 emphasized this competency more than any other (see Table 4). It is worth noting that the same institution did not emphasize this competency as heavily on its peer rubric. This suggests that different observable items should be emphasized depending on the type of rubric that is used.

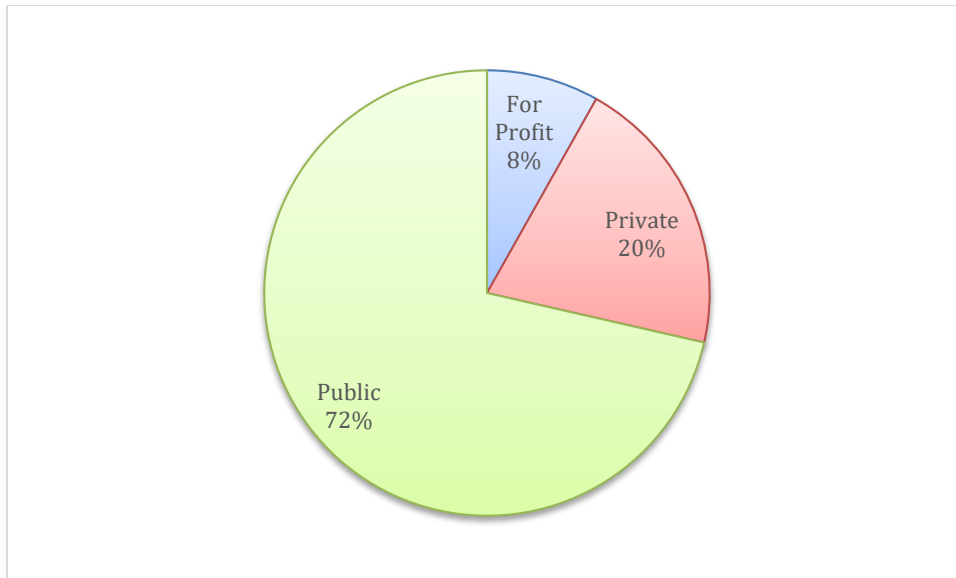


Figure 2. Percentage of coding decisions (N=49) for Administration/Leadership organized by type of institution.

Task 2.2 was ranked second among all tasks. Of all rubric items, 9% addressed this task. These items addressed a variety of expectations instructors were to share with students. These included clearly explaining instructions for assignments, clear guidelines on how they would be graded, time to completion, and other items that students would need to understand in order to be successful in the course. The other tasks associated with this competency were rarely addressed in the rubrics.

Classroom decorum. Classroom Decorum was heavily addressed by for profit rubrics representing 17% of all coding decisions, but it was also emphasized by the other rubrics as well. Similar to the Administrative/Leadership competency, this is largely because one task was heavily addressed. Task 5.3 (managing and modeling communication behavior) was the fourth task that was most emphasized in all of the rubrics. It represented almost 7% of all coding decisions. There was considerable overlap between this task and tasks 1.3, 1.4, and 8.2. This was largely because a considerable amount of course communications in an online course occur

on a discussion board. Additionally, the added element of “providing a good model” of how to communicate led to some overlap with task 1.10.

Multimedia technology, technological competence, and policy enforcement. The remaining competencies all had a weak presence within the rubrics. Multimedia Technology, Technological Competence, and Policy Enforcement were addressed by 2%, 2%, and 3% of all coding decisions. Multimedia Technology was present in half of the rubrics where Policy Enforcement was represented in 6 of the rubrics. Technological Competence was only addressed in 2 of the 8 rubrics. It was, however, heavily emphasized in the Private 2 rubric as the most emphasized competency, together with Active Teaching/Responsiveness at 15% of the rubric. These items were largely coded to task 6.1 and addressed the variety of ways an instructor can utilize technology to provide a better experience for students. These include an instructor knowing how to set-up third-party applications, provide audio or video feedback (not just text), and generally demonstrate ability to use technological tools to facilitate the course. Similar items were not included in other rubrics.

Are There any Rubric Items that do not Address the Tasks in the Bigatel et al. (2012) Model?

We identified 5 additional tasks represented in rubric items that are not listed in the Bigatel et al. (2012) model (see Table 5). We created one additional category we titled Miscellaneous into which we grouped rubric items that were context specific or were vague and required additional clarification. Together, these 6 groupings make up 31% of the coding decisions (see Figure 1). Three of the 5 tasks listed in “Other” were ranked in the top 10 most represented tasks in all rubrics. These were tasks 8.1 (7), 8.2 (1), and 8.3 (3). We will discuss more specifically each of the 5 tasks below.

Course design. This task was strongly emphasized by public institution rubrics. It addressed 11% of all coding decisions on public institution rubrics. This is no surprise. The public institutions in this sample utilize a master course model, but not as extensively as the for profit and private institutions. Many of the public, online instructors also design the course they are teaching. This would necessitate including items on an observation rubric that also address course design. No rubric items on for profit and private rubrics addressed this task.

Building a learning community. Task 8.2 was the most emphasized task on all of the rubrics. Of all the coding decisions, almost 10% addressed this task. It is widely represented across all rubrics. It is evident in rubric items that address instructor presence in the course and efforts to communicate regularly with students. It is also demonstrated as an instructor shows care and concern for students and develops relationship with and among them. The emphasis on this task demonstrates a clear agreement among evaluation rubrics that this is an important task vital to effective online teaching.

Adapting to students' diverse needs. This task addresses an instructor's ability to adapt course content to facilitate the success of students with a wide variety of needs. These needs could be due to student diversity of culture, a student's physical or mental disability, or students who are simply struggling. This task was addressed by almost 8% of coding decisions and was ranked number 3. Some examples of rubric items that addressed this task are that an instructor demonstrates sensitivity and flexibility to diverse needs of the class, provides a policy for accommodations that is clear and made available for students, and offers supplemental materials and support for students in need of remediation. This task was far more emphasized by public rubrics (11%) than for profit rubrics (0%) and private rubrics (6%).

Course navigation. Of the tasks that were not represented in the Bigatel et al. (2012) model, task 8.4 had the weakest presence and was ranked number 21 out of 36 tasks. It is focused exclusively on helping orient students to the learning objectives and structure of the course. The rubric items coded to this task addressed how an instructor explains where students should start when first arriving in the course and how to navigate the course. It also addressed whether an instructor is informing students about learning objectives for each activity in the course as well as the course as a whole.

Instructor connects students to subject matter. Task 8.5 was in the upper half of most emphasized tasks at a rank of 14 and emphasized in a little over 3% of the coding decisions. This task addresses an instructor's passion for the subject as well as his or her ability to kindle and grow students' interest in it as well. Rubric items that address this task focus on an instructor's ability to present content in an engaging way. Additionally, they addressed how an instructor meaningfully connects students to the content through personal connections with student experiences and relevancy. Private rubrics emphasize task 8.5 more than for profit and public rubrics.

Conclusion and Recommendations

It is apparent after comparing observational rubrics at institutions that some online teaching competencies are more emphasized in evaluations than others. This may simply be because some tasks are more observable than others. It also may be that the tasks address beliefs and attitudes that a potential online instructor must possess to be successful rather than behaviors they regular perform while teaching a course. Observational rubrics place greater emphasis on Active Learning and Active Teaching/ Responsiveness than on the other competencies. To address Active Learning, rubrics focus on the behaviors of an instructor to help students take an

active part in their own learning and sharing what they are learning with others. Active Teaching/ Responsiveness is the next most emphasized competency in the rubrics. It is the only competency where most of the tasks are heavily weighted and similarly emphasized in course observations.

The other two competencies that were emphasized in the rubrics included very few tasks that were emphasized in all the rubrics. The other three tasks did not have broad coverage or were completely absent. Classroom Decorum is emphasized widely in the rubrics. To address this competency, rubric items can focus on how an instructor models appropriate behavior in course communications. This can be done primarily on discussion boards, but also in other types of communication, including email or instant messaging within the course delivery system. Administration/Leadership is emphasized in rubric items that address how clearly and regularly an instructor explains student expectations. Instructors should regularly communicate to students regarding expectations on assignments as well as in the overall course. This can include grading guidelines.

As a result of our findings, we make three recommendations: (1) Evaluations of online instructors should include course observations that focus on instructional behaviors, (2) Observational rubrics should include items that focus on building relationship and community with and among students, and (3) Efforts should be made to establish a more comprehensive model of online teaching competencies that can be addressed in observational rubrics.

Focus on instructional behaviors. Contrary to conclusions of previous research (Thomas, & Graham, 2017), this study provides evidence that post-secondary institutions have established effective evaluation processes. The rubrics in this study demonstrate substantial overlap among institutions in the competencies that are evaluated when performing course

observations. These observations focus largely on instructional behaviors rather than simply course design as identified in previous online evaluation research (Piña & Bohn, 2014).

Institutions should develop evaluation processes that perform regular evaluation of instructional behaviors that occur during course delivery. This will allow administrators to not only help encourage the most effective teaching practices, but also help to provide more focused and timely training and professional development.

Emphasize building relationships and community. The most emphasized instructional behaviors in the observational rubrics are interactions where an instructor seeks to establish and nurture relationships with and among students. These relationships between instructors and their students as well as among students can help motivate students to learn (Sher, 2009). Developing a community in an online learning context can also contribute to a student's success in an online course (Shea, Sau Li, & Pickett, 2006; Garrison, Anderson, & Archer, 2000). Course observation rubrics should emphasize regular interactions between the instructor and students. These interactions often occur on discussion boards, within feedback for assignments, or on an announcement page. These interactions should demonstrate thoughtful and substantive interaction that demonstrates an instructor's awareness of the student as an individual (demonstrated by personalized interaction) as well as concern that they are learning course content (demonstrated by efforts to connect students to course content and learning objectives).

A need for a more comprehensive model of online teaching competencies. There were several tasks that were not present in the Bigatel et al. (2012) model but strongly emphasized in the rubrics. These include how an instructor is building a learning community within the course, adapting to students' diverse needs, and connecting students to course content in relevant and personalized ways. These have all been confirmed as critical to effective online

teaching (Garrison, et al., 2000; Graham, Cagiltay, Lim, Craner, & Duffy, 2001; Fish & Wickersham, 2009). This suggests that this model may be incomplete. Bigatel et al. (2012) acknowledged that there were behaviors not included in the model that are critical behaviors to effective online teaching. “For reasons not totally evident,” they explained that these behaviors “simply did not load onto the top seven competency categories” when they performed an exploratory factor analysis (p. 73).

The heavy emphasis of observational rubrics on these behaviors indicates that these are important behaviors to effective online teaching. These rubrics have been refined over time with reviews of literature and a wide variety of experience. They are not perfect but do provide a strong argument for some of the most important instructional tasks. The Bigatel et al. (2012) model establishes a good basis for online teaching competencies to build on and refine. We recommend additional research to help establish a consensus of observable online teaching practices. To this end, administrators can do more to share information and experience with one another. This can include sharing instruments, practices, and experiences to further refine and establish shared best practices. By establishing this consensus, institutions will be better equipped to perform effective evaluations that help to improve online instruction.

References

- Academic Senate for California Community Colleges. (2013). *Sound Principles for Faculty Evaluation*. Sacramento, California: Author. Retrieved from https://asccc.org/sites/default/files/publications/Principles-Faculty-Evaluation2013_0.pdf.
- Allen, I. E., & Seaman, J. (2003). Sizing the Opportunity: The Quality and Extent of Online Education in the United States, 2002 and 2003. *Sloan Consortium (NJ1)*.
- Allen, I. E., & Seaman, J. (2017). *Digital learning compass: Distance education enrollment report 2017*. Babson Park, MA: Babson Survey Research Group, e-Literate, and WCET.
- 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. <http://doi.org/10.1080/01587919.2011.610293>
- 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.
- Benton, S. L., Webster, R., Gross, A. B., & Pallett, W. H. (2010). IDEA technical report no. 15: An analysis of IDEA student ratings of instruction in traditional versus online courses, 2002–2008 data. *The IDEA Center, (IDEA Paper #15)*, 1–38.
- 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.
- Bigatel, P. M., Ragan, L. C., Kennan, S., May, J., & Redmond, B. F. (2012). The identification of competencies for online teaching success. *Journal of Asynchronous Learning Network, 16*(1), 59–78.

- 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.
- 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.
- Creasman, P. A. (2012). Considerations in Online Course Design. *The IDEA Center*, (IDEA Paper #52), 1–12.
- 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(1), 105–122. <http://doi.org/10.1080/01587910600654809>
- Darling, D. D. (2012). *Administrative evaluation of online faculty in community colleges*. Fargo: 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.
- 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>
- Fish, W. W., & Wickersham, L. E. (2009). Best practices for online instructors: Reminders. *Quarterly Review of Distance Education*, 10(3), 279–284.

- Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3), 87–105. [http://doi.org/10.1016/S1096-7516\(00\)00016-6](http://doi.org/10.1016/S1096-7516(00)00016-6)
- 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/
- Graham, C. R., Woodfield, W., & Harrison, J. B. (2012). A framework for institutional adoption and implementation of blended learning in higher education. *The Internet and Higher Education*, 18(1), 4–14. <http://doi.org/10.1016/j.iheduc.2012.09.003>
- Hill, P. (2012). Online Educational Delivery Models: A Descriptive View. *EDUCAUSE Review*, 47(6), 84–86. Retrieved from <http://ezproxy.lib.utexas.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ996668&site=ehost-live%5Cnhttps://net.educause.edu/ir/library/pdf/ERM1263.pdf>
- Holsti, O. (1969). *Content analysis for the social sciences and humanities*. Don Mills, ON: Addison-Wesley.
- Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <http://doi.org/10.1177/1049732305276687>
- 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., Donnelly, 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>
- Moskal, P., Dziuban, C., & Hartman, J. (2013). Blended learning: A dangerous idea? *Internet and Higher Education*, 18(1), 15–23. <http://doi.org/10.1016/j.iheduc.2012.12.001>
- Palloff, R. M., & Pratt K. (2008) Effective course, faculty, and program evaluation. Paper presented at the Annual Conference on Distance Teaching & Learning, Madison, WI, 2008. University of Wisconsin.
- 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.
- 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.
- Shea, P., Sau Li, C., & Pickett, A. (2006). A study of teaching presence and student sense of learning community in fully online and web-enhanced college courses. *Internet and Higher Education*, 9(3), 175–190. <http://doi.org/10.1016/j.iheduc.2006.06.005>
- Sher, A. (2009). Assessing the relationship of student-instructor and student-student interaction to student learning and satisfaction in Web-based Online Learning Environment. *Journal of*

- Interactive Online Learning*, 8(2), 102–120. Retrieved from <http://www.ncolr.org/jiol/issues/pdf/8.2.1.pdf>
- Smith, M. M., & Macdonald, D. (2015). Assessing Quality and Effectiveness in Fully Online Distance Education. *Journal of Perspectives in Applied Academic Practice*, 3(1), 24–38.
- Smith, T. C. (2005). Fifty-One Competencies for Online Instruction. *The Journal of Educators Online*, 2(2), 1–18. Retrieved from http://web.kennisnet2.nl/attachments/session=cloud_mmbase+1690923/Ted_Smith_Final.pdf
- Stanišić Stojic', S. M., Dobrijevic', G., Stanišić', 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.
- 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>
- Thomas, J. E., & Graham, C. R. (2017). Common practices for evaluating post-secondary online instructors. *Online Journal of Distance Learning Administration*, 20(4). Retrieved from https://www.westga.edu/~distance/ojdla/winter204/thomas_graham204.html
- Varvel, V. E. (2007). Master Online Teacher Competencies. *Online Journal of Distance Learning Administration*, 10, 1–47. Retrieved from <http://www.westga.edu/~distance/ojdla/spring101/varvel101.htm>
- 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.

- Wilson, G., & Stacey, E. (2004). Online interaction impacts on learning: Teaching the teachers to teach online. *Australasian Journal of Educational Technology*, 20(1), 33–48. Retrieved from <http://www.ascilite.org.au/ajet/ajet20/wilson.html>[2/07/2009]
- Young, S., & Duncan, H. E. (2014). Online and Face-to-Face Teaching: How Do Student Ratings Differ? *Journal of Online Learning and Teaching*, 10(1), 70–79.

APPENDIX A

Table 6

All Coding Decisions Organized by Tasks, Including Percentage of Total Coding Decisions, Percentage of Competency, and Rank among All Tasks

Competency	Tasks	# (% of Total)	# (% of Competency)	Rank
1. Active Learning	1.1	7 (1.66)	7 (7)	18
	1.2	5 (1.19)	5 (5)	22
	1.3	18 (4.28)	18 (18)	8
	1.4	22 (5.23)	22 (22)	6
	1.5	4 (0.95)	4 (4)	24
	1.6	10 (2.38)	10 (10)	16
	1.7	5 (1.19)	5 (5)	23
	1.8	16 (3.80)	16 (16)	12
	1.9	2 (0.48)	2 (2)	30
	1.10	12 (2.85)	12 (12)	15
2. Administration/Leadership	2.1	4 (0.95)	4 (8)	25
	2.2	38 (9.03)	38 (78)	2
	2.3	2 (0.48)	2 (4)	31
	2.4	1 (0.24)	1 (2)	35
	2.5	4 (0.95)	4 (8)	26
3. Active Teaching/Responsiveness	3.1	17 (4.04)	17 (24)	9
	3.2	17 (4.04)	17 (24)	10
	3.3	24 (5.70)	24 (33)	5
	3.4	14 (3.33)	14 (19)	13
	3.5	0 (0.00)	0 (0)	36
4. Multimedia Technology	4.1	6 (1.43)	6 (60)	20
	4.2	4 (0.95)	4 (40)	27
5. Classroom Decorum	5.1	2 (0.48)	2 (5)	32
	5.2	2 (0.48)	2 (5)	33
	5.3	29 (6.89)	29 (78)	4
	5.4	4 (0.95)	4 (11)	28

6. Technological Competence	6.1	7 (1.66)	7 (78)	19
	6.2	2 (0.48)	2 (22)	34
7. Policy Enforcement	7.1	3 (0.71)	3 (27)	29
	7.2	8 (1.90)	8 (73)	17
8. Other	8.1	5.23	17	7
	8.2	9.98	32	1
	8.3	7.60	24	3
	8.4	1.43	5	21
	8.5	3.09	10	14
	8.6	4.04	13	11

APPENDIX B

Table 7

Percentage of All Coding Decisions by Task Divided into Three Categories of Post-Secondary Institution Rubrics; For Profit, Private, and Public

Competency	Tasks	For Profit (N=54) # (% of FP)	Private (N=140) # (% of Priv)	Public (N=227) # (% of Pub)
1. Active Learning	1.1	0 (0)	0 (0)	7 (3)
	1.2	0 (0)	0 (0)	5 (2)
	1.3	0 (0)	7 (5)	11 (5)
	1.4	0 (0)	7 (5)	15 (7)
	1.5	0 (0)	1 (1)	3 (1)
	1.6	2 (4)	7 (5)	1 (.4)
	1.7	0 (0)	0 (0)	5 (2)
	1.8	2 (4)	4 (3)	10 (4)
	1.9	0 (0)	0 (0)	2 (1)
	1.10	5 (9)	3 (2)	4 (2)
2. Administration/ Leadership	2.1	0 (0)	2 (1)	2 (1)
	2.2	3 (6)	6 (4)	29 (13)
	2.3	1 (2)	1 (1)	0 (0)
	2.4	0 (0)	0 (0)	1 (.4)
	2.5	0 (0)	1 (1)	3 (1)
3. Active Teaching/ Responsiveness	3.1	5 (9)	5 (4)	7 (3)
	3.2	4 (7)	6 (4)	7 (3)
	3.3	4 (7)	11 (8)	9 (4)
	3.4	3 (6)	6 (4)	5 (2)
	3.5	0 (0)	0 (0)	0 (0)
4. Multimedia Technology	4.1	0 (0)	3 (2)	3 (1)
	4.2	0 (0)	2 (1)	2 (1)
5. Classroom Decorum	5.1	1 (2)	0 (0)	1 (.4)
	5.2	1 (2)	0 (0)	1 (.4)
	5.3	7 (13)	10 (7)	12 (5)
	5.4	0 (0)	2 (1)	2 (1)

6. Technological Competence	6.1	0 (0)	7 (5)	0 (0)
	6.2	0 (0)	2 (1)	0 (0)
7. Policy Enforcement	7.1	0 (0)	1 (1)	2 (1)
	7.2	1 (2)	3 (2)	2 (1)
8. Other	8.1	0 (0)	0 (0)	22 (10)
	8.2	7 (13)	16 (11)	19 (8)
	8.3	0 (0)	8 (6)	24 (11)
	8.4	0 (0)	2 (1)	4 (2)
	8.5	1 (2)	8 (6)	4 (2)
	8.6	7 (13)	9 (6)	1 (.4)

DISSERTATION CONCLUSION

The purpose of this study was to explore the current state of online instructor evaluation in post-secondary institutions. Inherent in this, was a desire to not only identify some of the problems that exist, but also the challenges and what institutions are doing to overcome them. Enrollment in online courses at post-secondary institutions continues to swell requiring growing numbers of online instructors (Allen & Seaman, 2017). As a result, institutions are grappling with issues of hiring, training, and developing online instructors. Ensuring instructional effectiveness is a key purpose to regular online teacher evaluations (ASCCC, 2013; Dana, Havens, Hochanadel, & Phillips, 2010).

In Article 1, “Common Practices for Evaluating Post-Secondary Online Instructors,” we sought to identify some of the current problems, identified in the literature, with online teacher evaluations. The research on this topic was rather sparse and underdeveloped considering the growth of online programs. The earliest research focused on how comparable online teaching is to face-to-face teaching and what the similarities and differences are (Darabi, Sikorski, & Harvey, 2006; Tallent-Runnels, Cooper, Lan, Thomas, & Busby, 2005). This also led to the question of whether it was appropriate to use the same evaluation instruments and processes (Berk, 2013). This discussion largely focused on student evaluation instruments. Some felt strongly that online instructors were receiving lower ratings on student evaluations than face-to-face instructors simply because using the same instrument for both modalities unnecessary penalized online teachers (Loveland, 2007). It is clear, however, from additional research that good teaching is good teaching regardless of the modality and student evaluation instruments can be independent of modality and provide accurate measures of effective teaching.

Although the literature has provided extensive discussion of student evaluations of online teaching, another alarming trend apparent in the research is the neglect of other types of evaluation of online teaching. This includes peer evaluations, administrative evaluations, and self-evaluations. We also found that post-secondary institutions are focusing evaluations extensively on course design rather than teaching behaviors during course delivery. More institutions are adapting a master course model (Cheski & Muller, 2010; Hill, 2012) where the instructors that are teaching a course are not responsible for the course design. These problems of inadequate evaluation measures and inappropriate emphasis of existing evaluations propelled us into the other two articles of this dissertation.

Article 2, “Current Practices of Online Instructor Evaluation in Higher Education,” focused on how institutions provide adequate measures of online instructor evaluations in addition to student evaluations of online instructors. After collecting data from 10 different administrators of online programs we found that one of our conclusions in Article 1 was inaccurate. We concluded in Article 1 that developing evaluation systems at post-secondary institutions is a neglected area of research and may reflect that it is also a neglected area by post-secondary institutions. We were concerned that institutions were not providing adequate attention to establishing evaluations that help to ensure online teaching effectiveness. A growing trend among post-secondary institutions is their efforts to employ more than one type of evaluation of online instructors. The institutions in our study largely used student, peer, and administrative evaluations to evaluate instructors using a variety of different perspectives that each provide important pieces to inform instructors and administrators regarding a faculty’s teaching effectiveness. No single one of these types of evaluations provides all of the information necessary to judge teaching effectiveness.

Additionally, from this study, we found that many institutions perform course visits in order to evaluate instructional behaviors during course delivery. In order to ensure that these observations follow a similar structure, observers utilize a rubric that guides the evaluator as to what to look for in his or her observation. These rubrics are used in both peer and administrative evaluations. The use of these instruments helps the evaluation to be more objective. We felt that these rubrics could help to confirm or expand existing models of online teaching competencies. This led us to the final article of this dissertation.

In Article 3, “Online Teaching Competencies in Post-Secondary Observational Rubrics: What Are Institutions Evaluating?” we collected 8 observational rubrics used by post-secondary institutions to evaluate online instructors. We performed a directed content analysis in order to compare what institutions were evaluating with an existing model of online teaching competencies (Bigatel, et al., 2012). These rubrics focused on instructional behaviors rather than course design. The most emphasized behaviors on the observational rubrics were how an instructor establishes and maintains relationships with and among students and fosters a learning community in the course. We were surprised to find that these instructional behaviors are not included in the Bigatel, et al. (2012) model. This suggested that the model may not address the most important behaviors of effective online teaching. Additionally, surprising was the amount of agreement among institutions as to what online teaching competencies should be addressed in their rubrics. These findings suggest that more can be done to establish a more complete model of online teaching competencies that can inform evaluation of online instructors. This can help to refine rubrics and establish more effective practices by focusing on the most important instructional behaviors during course delivery.

In conclusion, this dissertation has provided a more clear representation of the current state of online teaching evaluation. It is more promising than we had first anticipated. This research has also highlighted additional problems that can be improved through increased sharing of instruments, practices, and experiences. Where one institution faces challenges with evaluating online instructors, another may have already discovered a potential solution. This can help to develop evaluation processes that are scalable as well as personable that benefit the individual instructor as well as the institution. Improving evaluation practices will help to improve the educational experience for students which is the aim of all online learning programs.

DISSERTATION REFERENCES

- Academic Senate for California Community Colleges. (2013). *Sound Principles for Faculty Evaluation*. Sacramento, California: Author. Retrieved from https://asccc.org/sites/default/files/publications/Principles-Faculty-Evaluation2013_0.pdf.
- Allen, I. E., & Seaman, J. (2017). *Digital learning compass: Distance education enrollment report 2017*. Babson Park, MA: Babson Survey Research Group, e-Literate, and WCET.
- 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.
- Bigatel, P. M., Ragan, L. C., Kennan, S., May, J., & Redmond, B. F. (2012). The identification of competencies for online teaching success. *Journal of Asynchronous Learning Network*, 16(1), 59–78.
- 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(1), 105–122. <http://doi.org/10.1080/01587910600654809>
- 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*. St. John’s, NL, Canada: Memorial University of Newfoundland, Distance Education, and Learning Technologies. Retrieved from http://www.mun.ca/educ/faculty/mwatch/laura_treslan_SPETHE_Paper.pdf

- 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>
- 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.
- Hill, P. (2012). Online educational delivery models: A descriptive view. *EDUCAUSE Review, 47*(6), 84–86. Retrieved from <http://ezproxy.lib.utexas.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ996668&site=ehost-live%5Cnhttps://net.educause.edu/ir/library/pdf/ERM1263.pdf>
- 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., Donnelly, 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 <http://www.westga.edu/~distance/ojdla/fall83/mandernach83.htm>.

- Palloff, R. M., & Pratt K. (2008) Effective course, faculty, and program evaluation. Paper presented at the Annual Conference on Distance Teaching & Learning, Madison, WI, 2008. University of Wisconsin.
- 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.
- 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://doi.org/10.1207/s15389286ajde1901>
- Rothman, T., Romeo, L., Brennan, M., & Mitchell, D. (2011). Criteria for assessing student satisfaction with online courses, *International Journal for e-Learning Security*, 1(1-2), 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>
- Schnitzer, M., & Crosby, L. S. (2003). Recruitment and development of online adjunct instructors. *Online Journal of Distance Learning Administration*, 6(2), 1–7. Retrieved from http://www.westga.edu/~distance/ojdla/summer62/crosby_schnitzer62.html
- 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>
- Stanišić Stojić, S. M., Dobrijević, G., Stanišić, N., & Stanić, 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.
- 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.