2018-06-15

Current Practices of Online Instructor Evaluation in Higher Education

Jon E. Thomas
Brigham Young University, thomasje@ldschurch.org

Charles R. Graham
Brigham Young University

Anthony A. Piña
Sullivan University

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; Stanišic´ Stojic´, Dobrijevic´, Stanišic´, & Stanic´, 2014) as well as professional development (ASCCC, 2013; Dana, Havens, Hochanadel, & Phillips, 2010; DeCosta, Bergquist, & Holbeck, 2015; Mandernach, Donnelli, 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.

Compora, 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., Donnelli, E., Dailey, A., & Schulte, M. (2005). A faculty evaluation model for online instructors: Mentoring and evaluation in the online classroom. *Online Journal of Distance Learning Administration*, *8*(3), 1–28. Retrieved from 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/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

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

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