International Journal of Teaching and Learning in Higher Education 2022, Volume 33, Number 3, 285-297
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Faculty Use of End-of-Course Evaluations

Noelia Pacheco Diaz, John P. Walker, Louis M. Rocconi, Jennifer A. Morrow, Gary J. Skolits, Jessica D. Osborne, and T. Richard Parlier
University of Tennessee, Knoxville

End-of-course evaluations are a central part of the accountability system at American universities. They are used formatively to evaluate the way courses are delivered and the effectiveness of teaching practices. Therefore, institutions may use these instruments to make course changes or tenure and promotion decisions. The purpose of this research study was to gain a better understanding of how faculty use the evaluation results. Faculty from a southeastern U.S., research intensive university received a survey containing 28 Likert-scale items and two open-ended questions. Quantitative data were analyzed with descriptive statistics, while a thematic analysis was used to analyze the qualitative data. Most participants in our study (89%) reported that they use the feedback provided by the end-of-course evaluations to make changes in their courses. In the qualitative section, faculty believe that end-of-course evaluations provide meaningful input regarding student learning; however, they stated that they would like to see more open-ended questions within end-of-course evaluations. Findings from our study suggest that faculty value the information they receive from students, but end-of-course evaluations need to be better targeted to the needs of the course and faculty using it.

End-of-course evaluation surveys have been an integral part of the higher education system in the U.S. since the 1920s (Dommeyer et al., 2004). They were conceived as a way for students to provide an assessment of faculty course delivery and teaching effectiveness. One of the first incarnations of end-of-course evaluations was the "Purdue Rating Scale for Instructors" which was developed as a means to measure teaching quality in a more scientific way (Calkins & Micari, 2010). Most end-of-course evaluation instruments are composed of a combination of open-ended and closed-ended (e.g., rating scales) items focused on the effectiveness of teaching and the content of the course (Gravestock & Gregor-Greenleaf, 2008). Students are normally guaranteed anonymity and responses are collected at the end of the course. Today, end-of-course evaluations are a ubiquitous part of the higher education landscape.

End-of-course evaluations are so pervasive in higher education because they provide institutions with a seemingly objective measure of teaching quality. Moreover, they are usually the main component in an accountability process in which postsecondary institutions gauge the quality of the education they provide (Bradley et al., 2016). Given the content covered by these evaluations, it is reasonable to assume that end-of-course evaluations would provide a valid account of how instructors perform, at least from the perspective of students. However, there is a considerable amount of research that calls into question the validity of end-of-course evaluations (Hornstein, 2017; Jaquett et al., 2016; Uttl et al., 2017). Most of these criticisms are related to low response rates, biases in students' responses, and the ways in which these evaluations are used. Despite these concerns, end-of-course evaluations are still widely used by most universities in the United States. Therefore, understanding faculty's perceptions of end-of-course evaluations and how they use end-of-course evaluations are vitally important. However, the research on faculty's use of end-of-course evaluations is scant. Nasser and Fresko (2002) surveyed faculty in Israel about their views on end-of-course evaluations and found that while faculty viewed the instrument as only moderately useful, they reported making changes to their course based on students' responses to the instrument. Our study furthers the research in this area by examining how useful faculty view end-of-course evaluations and whether and how faculty use end-of-course evaluations to make pedagogical and curricular changes to their courses.

Review of the Literature

Current Use of End-of-Course Evaluations

End-of-course evaluations are commonly used measures of teaching effectiveness. Most end-of-course evaluations are intended to be used as a formative approach to advise faculty on ways to improve teaching and learning in the classroom. A consensus on what constitutes effective teaching, how it can be measured, and how teaching can be enhanced from student feedback is not widely accepted among faculty and faculty developers (Abrami et al., 2007; Germain & Scandura, 2005). Some researchers recommend other approaches to assess teaching effectiveness such as entrance-exit surveys or peer evaluations of teaching. For example, Ng and colleagues (2016) suggested that entry-exit surveys better account for students' perceived attainment of the intended learning outcomes and highlight the effectiveness of course assessment strategies. However, these measures also have their limitations. Students may give high ratings at the beginning of the semester, which leaves little room to show growth and makes the measurement ineffective. Because end-of-course evaluation surveys are intended

to measure teaching effectiveness, faculty are encouraged to adjust their instruction and other aspects of their course based on these results (Hobson & Talbot, 2001). End-of-course evaluations are also used as a summative measure and used in different high stakes evaluations of faculty (e.g., tenure, promotion, course offerings), although many institutions will use end-of-course evaluations as only one part of the decision-making process in high-stakes decisions (Dommeyer et al., 2004).

Previous Research on End-of-Course Evaluation Surveys

The majority of research on end-of-course evaluations has focused on identifying problems with the validity and reliability of these instruments. Some researchers have pointed to biases as the cause of these issues. Many studies have found that selection bias, grade expectation, class size, gender biases, and faculty likeability affect the results of the end-of-course evaluations, which would jeopardize the validity of the assessment (Boring et al., 2016; Dodeen, 2013; Goos & Salomons, 2017). For instance, Goos and Salomons (2017) examined the selection bias present in end-of-course evaluations at a large European university. They examined results from over 28,000 students and 3,000 courses and found a positive selection bias indicating that students who responded to the end-of-course evaluations were more likely to have higher grades in the course and have a positive opinion on the course in general. They also observed that response rates tended to be higher during the first semester than during the second semester of the academic year. Other researchers have found a positive relationship between students' expected grades or current grade point average and favorable responses to end-of-course evaluations (Dodeen, 2013; Stroebe, 2016).

Wolbring and Treischl (2016) also examined the selection bias present in end-of-course evaluations. Using data from a single, German university, Wolbring and Treischl examined the relationship between students' responses to course evaluations at the beginning of the term and end of the term. They noted that while end-of-course evaluations are not entirely free of selection bias, variations in students' ratings within each course were small. Therefore, the researchers concluded that end-of-course evaluations seemed to be a good representation of students' ratings for the courses. However, variations in ratings were larger when comparing different courses; thus, the researchers concluded that end-of-course evaluations should not be used to compare different courses.

Gender bias is also a significant concern in end-of-course evaluations. Boring et al. (2016) examined data from a French university and found that female instructors were evaluated more harshly compared to their male counterparts. An experiment conducted by MacNell and colleagues (2015) found that students rated faculty with male names consistently higher than faculty with female names. Bianchini and colleagues (2013) found that female faculty were generally rated worse than male faculty, but this was dependent on the type of course being rated. In a recent study, Peterson and colleagues (2019) conducted an experiment in which students were randomly assigned to two different formats of end-of-course evaluations within each course; these courses corresponded to four faculty members (two males and two females). One form had language warning students about possible gender biases, and the other one was the regular end-of-course evaluation survey. They found that the instrument with language referring to gender bias had a small positive effect on the students' ratings for female faculty. They concluded that gender bias could be alleviated by using language that makes students aware of gender biases.

Some research has pointed to how a faculty member's personality might dictate how students evaluate faculty (Calkins & Micari, 2010; Hatfield & Coyle, 2013). For instance, students with this bias, whether negative or positive, will account for their opinion of the faculty's character (i.e., likeability, sense of humor, personality) rather than their teaching style (Clayson, 2009). Other researchers (e.g., Ory, 2001; Theall & Franklin, 2001) have contended the idea that faculty's personality has any influence on students' end-of-course evaluation ratings. Ory (2001) stated that faculty's so-called "personality attributes" could be a reflection of good teaching practices (e.g., showing enthusiasm) and may improve the quality of teaching. A final bias discussed in the literature concerns grade inflation as an unintended consequence of end-of-course evaluation. Grade inflation occurs when faculty provide advantageous grading schemes to increase the chance of favorable evaluations. Hoefer and colleagues (2012) found that students that provided more favorable reviews of instructors tended to have higher grades, particularly among male students.

Faculty Views and Use of End-of-Course Evaluations

The research on the value of the end-of-course evaluations is mixed. While faculty tend to report they use end-of-course evaluations "quite a bit" (BrckaLorenz et al., 2014), they have differing opinions on how helpful end-of-course evaluations can be for classroom improvement. Baker (1992) contended that end-of-course evaluations are a waste of time and contain an inordinate amount of bias. Other faculty recommend using midterm evaluations along with end-of-course evaluations to make course modifications that

will benefit currently enrolled students (Senior, 2000). For example, Diamond (2004) conducted midterm course evaluations in a private university with 14,000 students and 775 full-time faculty and found that by using midterm course evaluations, faculty were able to make immediate course changes (e.g., teaching methodology, grading, discussing expectations with students) that improved the instructional quality of the courses.

Small sample sizes, whether a result of small class size or low response rate, impact how and whether faculty use end-of-course evaluations. Faculty with small class sizes and few responses are less likely to use end-of-course evaluation results compared to faculty with larger class sizes and high response rates on end-of-course evaluations (Kane & Staiger, 2002). In a qualitative study, Iqbal and colleagues (2016) interviewed faculty about their motivations to use end-of-course evaluations and observed that faculty felt more compelled to use this feedback when the response rate was high.

Purpose and Research Questions

While previous research on end-of-course evaluations has focused on validity, bias, and alternative measures, the literature on how faculty actually *use* the feedback provided by end-of-course evaluations remains limited. The present study aimed to further research this area by investigating how faculty use end-of-course evaluations to make pedagogical and curricular changes to their courses. Given the critical role end-of-course evaluations play in faculty promotion, curriculum and instruction revision, and improvement, gaining a better understanding of how faculty use end-of-course evaluation data can help inform better practices both locally and nationally. The research questions guiding this study are as follows:

1. To what degree do faculty find end-of-course evaluations useful?
2. How are faculty using information from end-of-course evaluations?

Methods

Participants

The sampling frame used in this study consisted of 317 faculty members from a public research university in the southeastern U.S. who had a working relationship with the school's student success center. Participants in this study consisted of both non-tenure track and tenure track faculty members who currently, or within the last academic year, taught one or more courses, including online courses. Faculty received a recruitment e-mail with a link that prompted them to complete an online survey via Qualtrics. In the initial email invitation, participants were notified of any possible risks and were given assurance their survey responses and personal information would be kept confidential. Participants who had not started the survey received a reminder email two weeks after the initial invitation. The response rate for the survey was 17.4% and the final sample consisted of 55 participants. Participants were primarily non-tenured faculty (78%), White/Caucasian (73%), and female (51%). Participants who worked ten years or more at the university made up the plurality of the sample at 44%, while those who had worked 6-9 years or 3-5 years consisted of 29% and 24% of the sample, respectively. Table 1 displays a summary of the demographic characteristics of the sample.

Table 1
Descriptive Statistics of the Sample

Variable	N	%
Faculty Appointment Track		
Tenured track	9	16.4
Non-tenured track	43	78.2
Prefer not to say	3	5.5
Gender		
Male	19	34.5
Female	28	50.9
Prefer not to say	8	14.5
Race/Ethnicity		
Caucasian/White	40	72.7
Non-White	4	7.2
Prefer not to say	11	20.0
Years Worked at Institution		
1-2 years	0	0.0
3-5 years	13	23.6
6-9 years	16	29.1
10 or more years	24	43.6
Prefer not to say	2	3.6

Instrument

The survey consisted of 28 Likert-scale items that asked faculty about how well end-of-course evaluations informed them of the student experience, the usefulness of end-of-course evaluations for course planning and teaching, and their usage of end-of-course evaluations to make changes to course content and teaching practices. The survey also included two open-response items that asked about faculty usage of end-of-course evaluation and other methods of assessing the quality of faculty's courses. Additionally, four demographic items were asked related to appointment track (i.e., non-tenure track and tenure track), gender identity, current length of

appointment, and race/ethnicity. See Appendix A for the survey in its entirety.

Analyses

We analyzed both descriptive statistics and qualitative responses to answer our two research questions. Responses to five survey items were used to answer the first research question (i.e., to what degree faculty find end-of-course evaluations useful). These items included (a) how useful faculty feel end-of-course evaluations are for planning their course; (b) how often faculty make changes to their course as a result of the end-of-course evaluations; (c) how often faculty share end-of-course evaluation results with current students; (d) how often faculty use end-of-course evaluations to make changes to their teaching and pedagogy; and (e) how useful faculty feel end-of-course evaluations are for improving their teaching and pedagogy. The alpha reliability coefficient for these items was 0.78. We also conducted a thematic analysis on the open response questions to identify, analyze, and report patterns within data (Braun & Clarke, 2006).

To answer our second research question (i.e., how are faculty using information from the end-of-course evaluations), we examined responses from a set of items that asked faculty how well feedback from end-of-course evaluations informs various aspects related to the student experience, course planning, and teaching and pedagogy. The alpha reliability coefficient for these items was 0.96. We again used both descriptive statistics and thematic responses to gauge how faculty use the information in end-of-course evaluations. Respondents were asked to provide specific ways in which they used end-of-course evaluations.

Limitations

As with all research, ours is not without its limitations. Foremost, our response rate and sample size are low. Due to financial limitations, we were not able to offer an incentive to participate, which likely attributed to our low response rate. Even though researchers should strive to achieve maximal response rates, research has shown low response rates may not be detrimental to a study (Lambert & Miller, 2013), and even with a minimum of 50 responses, respondents may be able to provide reliable estimates (Fosnacht et al., 2017). Our findings are also limited by the fact that participants are from only one institution and weighted towards non-tenure track faculty (78% of respondents). While we advise readers to consider these factors when interpreting our findings, we still believe our study provides valuable information to add to the literature on faculty perceptions of end-of-course-evaluations.

Results

RQ 1: To What Degree do Faculty Find End-Of-Course Evaluations Useful?

We asked faculty members, "In general, how useful are end-of-course evaluations for planning your course?" The plurality of participants (42%) generally found end-of-course evaluations to be only moderately useful, while 16% found end-of-course evaluations to be "very" or "extremely" useful. Most participants (89%) reported that they used the feedback from the end-of-course evaluations to make changes to their course at least once a year. However, most faculty have kept their end-of-course evaluation results private from their current students. These faculty members (56%) stated that they *never* share the results of end-of-course evaluations with their current students. Approximately 60% of participants also found end-of-course evaluations to be at least moderately influential when it came to improving their teaching or pedagogy (i.e., 47% moderately influential, 9% very influential, 4% extremely influential). Nearly all participants (94%) shared that they used the feedback from the end-of-course evaluations to make changes to their teaching practices at least once a year. Table 2 provides additional descriptive statistics from these responses.

We constructed thematic codes from open-ended responses that addressed faculty members' thoughts on the usefulness of end-of-course evaluations. Two themes emerged: ways to improve end-of-course evaluations and other means of collecting information for course improvement. The first theme focused on ways to improve the end-of-course evaluation for better or more strategic use. In particular, faculty members noted that the use of end-of-course evaluations was limited because of the bias that persisted with them. As one faculty member expressed, "they provide a loose approximation of student perceptions of the course, although the correlation between the positivity or negativity of these assessments and student grades (or perceived physical attractiveness) limits their usefulness." Participants reported that end-of-course evaluations do not objectively measure the quality of the course or instruction but rather the mood of students. Participants attributed some of this bias to the timing of the end-of-course evaluations. For example, a faculty member stated:

> Student course evaluations should be given half-way or two-thirds into the course rather than at the very end...you would alleviate those students venting because they earned a failing grade, i.e. those simply angry/frustrated at not doing the work and looking to blame the instructor...I believe this

Table 2
Use of End-of-Course Evaluations by Faculty (N = 55)

Item	Mean	SD	% 4 or 5	% 1
In general, how useful are end-of-course evaluations for planning your course?[a]	2.62	1.03	16.4	16.4
How often during the school year do you use end-of-course feedback to make changes to your course?[b]	2.60	0.89	12.7	10.9
How often do you use end-of-course evaluations to make changes to your teaching practices?[b]	2.70	0.88	10.9	5.6
How often do you share the results of end-of-course evaluations from previous courses with your current students?[c]	1.75	1.09	10.9	56.4
Overall, how useful are end-of-course feedback with improving your teaching/pedagogy?[d]	2.65	0.93	12.7	10.9

[a]1 = not at all useful, 2 = slightly useful, 3 = moderately useful, 4 = very useful, 5 = extremely useful
[b]1 = never, 2 = once a year, 3 = twice a year, 4 = three times a year, 5 = four or more times a year
[c]1 = never, 2 = sometimes, 3 = about half the time, 4 = most of the time, 5 = always
[d]1 = not influential at all, 2 = slightly influential, 3 = moderately influential, 4 = very influential, 5 = extremely influential

would be a more fair, accurate assessment of the instructor in particular.

Some faculty would follow these criticisms with suggestions to remove bias-prone questions, specifically closed-ended responses. These respondents believe closed-ended responses are unusable to make informed decisions for improving overall course quality.

To address the bias and provide more useful information for teaching and course development, faculty members suggested the inclusion of more open-ended questions in the end-of-course evaluations. Faculty expressed that open- ended responses helped to clarify issues with the course or teaching style and provided a medium for students to provide suggestions for improvement. As one faculty member voiced,

Additional open-ended responses such as: "If you could make one change in the course, what would it be? I feel like the question about what they liked about the course already addresses this, but this is more pointed to a change student would make. I would look through these responses for responses that show up a lot."

Open-ended responses were overwhelmingly the more popular item type among respondents. Most respondents who favored this type of item were adamant they provided better feedback and were actionable. Closed-ended items offered little to no substantive feedback for improvement, as they are generalized to accommodate a spectrum of disciplines and course types.

One respondent considered the quantitative feedback to be “largely useless” to make course improvements. Another faculty member stated that closed-ended responses were limiting to make improvements because they were too generic, stating that “what's okay for one instructor isn't okay for another.” A closed-ended question that is useful for one faculty member, may not have any relevant value for another faculty member to inform them on effectiveness in teaching and learning.

In alignment with the suggestion for more open-ended items, faculty shared their desire to see more student self-reflection in the form of open-ended response options. Faculty members provided examples of open-ended questions that could be incorporated into an end-of-course evaluation. For instance, a faculty member suggested, “if you could go back and give yourself advice at the start of the class (or start of a multi-semester sequence) what would it be?” Others shared they would have questions about the frequency and length of time studying for the course, or conversing with faculty or other students about coursework. Still, others focused on questions related to self-advice for the future to improve academic performance.

A second theme that emerged focused on other means of collecting information for the course and pedagogical improvements. Some faculty did not find the end-of-course evaluations to be useful for their course or pedagogical development. Respondents who expressed this opinion mentioned that they focused on more objective, direct methods of evaluations and assessments such as test scores, essays, and projects. One faculty member stated they examined “test scores, percentage of students who complete homework correctly, percentage of students who even attempt

homework, class attendance and participation" to better understand where the course, teaching, or students faltered or succeeded. Others used alternative indirect measures to assess the quality of their classroom and teaching. One faculty member stated, "I build a strong rapport with my students and many of them come to me throughout the semester to give feedback." Many faculty members shared that they are using end-of-course evaluations to make alterations to their courses and teaching methods, even though some of them reported using alternative evaluation options (e.g., mid-semester evaluations).

RQ 2: How are Faculty Using Information from End-Of-Course Evaluations?

We surveyed faculty about the usefulness of end-of-course evaluations. We asked faculty how much does feedback from end-of-course evaluations inform their understanding of the student experience, planning of their course, and teaching and pedagogy. Table 3 outlines the faculty responses to these questions. Faculty indicated that end-of-course evaluations were most informative regarding how effectively the faculty member explained content (M = 3.53, SD = 1.16), how they interact with students (M = 3.49, SD = 1.14), and how clearly they explained difficult material (M = 3.40, SD = 1.11) when it relates to the student experience. End-of-course evaluations were least informative when it came to understanding students' level of learning in the class (M = 2.64, SD = 1.19) and the level of difficulty of exams (M = 2.64, SD = 1.07). When asked about the ability of end-of-course evaluations to inform on course planning, faculty reported that identifying issues and problems related to course design as the most informative (M = 2.91, SD = 0.92), while the effectiveness of content (M = 2.11, SD = 1.08) and information about learning outcomes (M = 2.11, SD = 1.06) were the least informative. End-of-course evaluations as they related to teaching and pedagogy were most informative to learn about the effectiveness of lectures (M = 2.51, SD = 0.99), but least informative for understanding the effectiveness of demonstrations (M = 2.27, SD = 1.07).

We constructed thematic codes from open-ended responses that addressed faculty members' experiences with end-of-course evaluations. Two themes emerged: use of end-of-course evaluations for course improvement and job assessment. The first theme focused on faculty use of end-of-course evaluations to improve courses. As stated previously, faculty members reported that the most useful part of the end-of-course evaluations was the open-ended responses. One faculty member stated, "The written responses are the most powerful. I read through all of them. If there are multiple students mentioning something, I enact change on that issue." Other faculty members noted the use of open-ended responses for improvement of courses and pedagogical considerations. A faculty member reported,

> I try to see on which issues student feedback is consistent across semesters, and then work to improve on those points (e.g., when I see that there is too much information in lecture / talks too fast from multiple individuals, I know to thin out my lectures and slow down).

Some faculty shared how they were using end-of-course evaluations, while others shared disillusionment over the type of items used and the questions included in the evaluations. Faculty alluded to the irrelevant questions or the limited information the closed-ended items provided. Although the closed-ended responses provide quantifiable evaluations of teaching performance from the student perspective, some faculty do not find the information as valuable or resourceful as the open-ended responses. One faculty member shared that "it [the end-of-course evaluation] has made me less of a believer in and more of a skeptic of online numerical student evaluations of instruction." The open-ended responses allow faculty to identify specific elements of the course or pedagogy that need to be adjusted, altered, or eliminated, especially if common complaints or suggestions are prominent. As one faculty member noted,

> I use the open-ended questions to see if some aspect comes up often in those open-ended comments. If it seems fair (e.g., isn't just that the class is too early), I might adjust my course to include more of that aspect if students said it worked well or less of that aspect if most students say it didn't work.

Although faculty are receptive to student suggestions, the suggestions need to be practical and reasonable; otherwise, faculty consider open-ended responses to be as unusable as the generic closed-ended responses.

A second theme emerged around faculty job assessment. Some faculty members mentioned that end-of-course evaluation data was included in their portfolios for promotion and other job-related purposes (e.g., course scheduling, future course offerings). A respondent described this as "I have to submit them as a part of my teaching portfolio so that I can be evaluated annually by the department." Participants also reported using end-of-course evaluations to make changes in their course. These changes were designed to make specific improvements to teaching styles, assignments, grading procedures, assessment techniques, communication with students, and the use of technology.

Table 3
Informative Feedback from End-of-Course Evaluations

Item	Mean	SD	% a lot or great deal	% none at all
Student Experience				
Student's level of learning in the class	2.64	1.19	23.6	21.4
The required effort of the student	2.84	0.99	20.0	7.3
Level of difficulty with exams	2.64	1.07	18.2	16.4
Level of difficulty with assignments[a]	2.74	1.02	20.4	11.1
Teacher feedback to the student regarding overall progress	2.84	1.16	27.3	14.5
Teacher feedback to the student regarding assignments/tests	2.76	1.09	21.8	14.5
Student interest in course	3.16	0.99	40.0	3.6
Student interest in assignments (readings, homework, labs)	2.82	1.11	29.1	14.5
Your knowledge of course content	3.04	1.35	43.6	20.0
How effectively you explain content	3.53	1.16	56.4	7.3
How effectively you use course time	3.35	1.08	45.5	3.6
How clearly you explain difficult material	3.40	1.11	47.3	5.5
How you interact with students	3.49	1.14	52.7	5.5
Course Planning				
Identifying Issues/Problems	2.91	0.92	25.5	5.5
Content[a]	2.11	1.08	13.0	37.0
Experiential Methods	2.16	0.99	9.1	29.1
Learning Outcomes	2.11	1.06	9.1	34.5
Assessment	2.35	1.05	12.7	20.0
Teaching and Pedagogy				
Lecture	2.51	0.99	16.4	12.7
Discussion	2.31	1.13	12.7	29.1
Self-Instruction (assignments, readings, etc.)	2.40	1.11	16.4	25.5
Group Projects[a]	2.39	1.11	16.7	25.9
Demonstrations (e.g., presentations)	2.27	1.07	10.9	29.1

Note. 1 = None at all, 2 = a little, 3 = a moderate amount, 4 = a lot, 5 = a great deal; N = 55; [a]N = 54

Accounts of end-of-course evaluations for assessment of teaching in the classroom were common. Some faculty stated using them as a means to promote their course for future students, provided the results were positive. Others noted a more administrative purpose for end-of-course evaluations. These faculty members reported that their departments use results to show evidence of effective instruction in teaching portfolios for tenure and promotion. However, faculty expressed concern and skepticism regarding the real value and validity of end-of-course evaluations. Many respondents show concern over the singular use of end-of-course evaluations and the persistent bias that accompanies this form of faculty member assessment.

Discussion

End-of-course evaluations are the most common approach to assess teaching effectiveness (Hornstein, 2017). Marks (2012) found that end-of-course evaluation instruments influence how students and faculty perceive the quality of classroom pedagogy, which in turn, impacts pedagogical decisions. Research suggests that student end-of-course evaluations do contribute considerably to the enhancement of teaching and learning when supplemented with teacher consultations (Marsh & Roche, 1993; Murray, 1997); therefore, the use of end-of-course evaluations is important to understand and investigate. Participants in this study were split on the usefulness of end-of-course evaluations. While nearly all faculty participants reported that they used the feedback to make changes to their courses and teaching practices, just over half reported that these evaluations were influential in improving their teaching or pedagogy. Most faculty indicated that end-of-course evaluations provided informative feedback on the quality of lecture and self-instruction (e.g., assignments, readings, and independent study), but over a quarter of faculty shared that the end-of-course evaluations provided no valuable feedback on the quality of group projects or demonstrations (e.g., presentations). Still, other respondents noted that closed-ended items did not provide them with useful feedback.

Although quantitative measures are the most common form of end-of-course evaluations (Davis, 2009), participants in our study alluded to open-ended responses as the most useful form of feedback from the end-of-course evaluations. Specifically, participants recommended using open-ended questions to alleviate the bias found in end-of-course evaluations. They suggested more open-ended questions be added to the instrument to improve its usefulness to improve teaching and learning in the classroom. Other researchers (Steyn et al., 2019) have found that students provided more actionable recommendations for course improvement when student evaluations used open-ended questions in lieu of quantitative items. Furthermore, faculty rated evaluative data from group interviews as more accurate, believable, and usable than closed-ended or open-ended items (Wachtel, 1998). Although some faculty in our study found group interviews as a more credible form of obtaining student evaluations of teaching, the logistics and time associated with collecting this information may not be feasible. Thus, faculty would likely benefit from more open-ended questions in the end-of-course evaluations. Some faculty suggested further modifications to end-of-course evaluations. These faculty wanted to see more individual, faculty designed open-ended questions for their specific courses. They felt that faculty designed questions may yield more useable results to make improvements in the classroom compared to standardized items.

Some participants were skeptical of the feedback they received from the end-of-course evaluations. Questions surrounding the validity and biases of end-of-course evaluations were abundant, despite numerous participants reporting that they use the feedback provided by the instrument. Skepticism about end-of-course evaluations is not unusual. Senior (2000) noted that faculty cynicism over the use of end-of-course evaluations for pedagogical and instructional decisions is persistent, citing concerns over the validity and bias of such instruments. The types of students that respond to end-of-course evaluations add to the claims of bias and questions related to validity (Goos & Salomons, 2017). For example, students' motivation to complete the end-of-course evaluations weighed heavily towards those who had largely extreme views of the course or instructor. Jaquett et al. (2017) found that students with positive views are more likely to respond to end-of-course evaluations.

Higher education institutions need to increase their efforts to elicit diverse student voices to participate more regularly in end-of-course evaluations. Institutions could do this by promoting practices that will increase end-of-course evaluation response rates, such as moving from online student evaluations to traditional classroom-based student evaluations (Stanny & Arruda, 2017; Stowell et al., 2012), email reminders to encourage students to complete the evaluations (Crews & Curtis, 2011), and explaining how the evaluations will be used (Anderson et al., 2005). Response rates improved when faculty members demonstrate that the end-of-course evaluations are used to adjust course structure or pedagogy (Chapman & Joines, 2017). Conversely, increased response rates of end-of-course evaluations would expand faculty use of the evaluation to improve courses and teaching practices. Many faculty noted that low response rates prevented them from trusting student evaluation results because they lack a strong representation of the student voice. Promotion of suggested practices could increase response rates and the subsequent use of end-of-course evaluations for instructional and pedagogical purposes.

An important idea emerged surrounding the use of other formative evaluations to understand the quality of teaching and learning throughout the semester. Several faculty indicated using midterm evaluations to make actionable pedagogical decisions. Midterm evaluations have been shown to elicit timely and useful student feedback of courses (Senior, 2000; Spencer & Schmelkin, 2002). Moreover, pedagogical changes made on the basis of midterm evaluation have the advantage that they can be used to implement changes within the same cohort of students who provided the feedback. Veeck et al. (2016) suggested the use of online collaborative evaluations in which students collectively provide feedback on classroom structure and pedagogy to improve the course. In these evaluations, students work in small teams to provide comments on the course. The feedback from these small teams has been shown to provide more useful feedback compared to standard individual evaluations (Veeck et al, 2016). Other faculty noted the use of embedded formative assessments, such as one-minute papers, self-reflection papers, and in-class clicker questions, to provide evidence for pedagogical and instructional adjustments. Embedded formative assessments happen during class time. This method of formative assessment allows faculty to quickly assess student progress, receive timely student feedback, and provide evidence for further pedagogical and instructional decision making (McMillan, 2018). Embedded formative assessments in conjunction with end-of-course evaluations and midterm evaluations provide a triangulation of evidence for faculty to make informed pedagogical and instructional decisions.

Participants in our study indicated that end-of-course evaluations could inform some aspects of teaching, such as how clearly the material was explained and the quality of student-faculty rapport; however, faculty reported that end-of-course evaluations do not measure how much students actually learned. Many outcomes of teacher effectiveness, such as student learning, are typically not measured directly in an end-of-course evaluation (Clayson, 2009). Although there

have been attempts made to statistically measure teacher effectiveness and student learning (Marlin & Niss, 1980), studies have shown no statistical association between end-of-course evaluations and student learning (Boring et al., 2016; Uttl et al., 2017). Concerns among faculty surface when end-of-course evaluations are the sole measure of teaching effectiveness, specifically when used to make considerations for tenure and promotion. Pounder (2007) expounded on the use of end-of-course evaluations for this purpose and noted that questions surrounding the usefulness of end-of-course evaluations to measure teacher effectiveness continues to be controversial.

Conclusions

End-of-course evaluations are a ubiquitous part of American higher education; therefore, understanding faculty members' perceptions of end-of-course evaluations and how they use the data gathered from end-of-course evaluations are critical. While the overwhelming majority of faculty members reported using end-of-course evaluations to make changes to their course at least once a year, less than one-fifth of respondents rated end-of-course evaluations as very or extremely useful for planning their course or improving their teaching and pedagogy. Faculty raised validity and bias concerns with end-of-course evaluations particularly when these are the sole measure of teaching effectiveness used to make considerations for tenure and promotion. Our findings suggest that while faculty value the information they received from students in end-of-course evaluations, end-of-course evaluations need to focus more on actionable open-ended responses and be targeted to the needs of the course and instructor.

References

Abrami, P.C., d'Apollonia, S., Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective.* Springer. https://doi.org/10.1007/1-4020-5742-3_10

Anderson, H.M., Cain, J., & Bird, E. (2005). Online student course evaluations: Review of literature and pilot study. *American Journal of Pharmaceutical Education, 69*(1), 34-43.

Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of Political Economy, 100*(3), 598-614.

Bianchini, S., Lissoni, F., & Pezzoni, M. (2013). Instructor characteristics and students' evaluation of teaching effectiveness: Evidence from an Italian engineering school. *European Journal of Engineering Education*, *38*(1), 38–57. https://doi.org/10.1080/03043797.2012.742868

Boring, A., Ottoboni, K., & Stark P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Computer Science*. http://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Bradley, K. D., Snyder, E. M., & Tombari, A. K. (2016). Higher education end-of-course evaluations: Assessing the psychometric properties utilizing exploratory factor analysis and Rasch modeling approaches. *International Journal of Assessment Tools in Education*, *3*(1), 3–22.

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

BrckaLorenz, A., Mccormick, A. C., & Peck, L. (2014). *Exploring how course evaluation outcomes are collected, shared and used*. Association for Institutional Research Annual Forum.

Calkins, S., & Micari, M. (2010). Less-than-perfect judges: Evaluating student evaluations. *Thought & Action*, 7–22.

Chapman, D. D., & Joines, J. A. (2017). Strategies for increasing response rates for online end-of-course evaluations. *International Journal of Teaching and Learning in Higher Education, 29*(1), 47-60.

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn?: A meta-analysis and review of the literature. *Journal of Marketing Education, 31*(1), 16-30. https://doi.org/10.1177/0273475308324086

Crews, T. B., & Curtis, D. F. (2011). Online course evaluations: Faculty perspectives and strategies for improved response rates. *Assessment and Evaluation in Higher Education, 36*(7), 865-878. https://doi.org/10.1080/02602938.2010.493970

Davis, B. G. (2009). *Tools for Teaching* (2nd ed.). John Wiley & Sons.

Diamond, M. R. (2004). The usefulness of structured mid-term feedback as a catalyst for change in higher education classes. *Active Learning in Higher Education*, *5*(3), 217–231. https://doi.org/10.1177/1469787404046845

Dodeen, H. (2013). Validity, reliability, and potential bias of short forms of students' evaluation of teaching: The case of UAE university. *Educational Assessment, 18*(4), 235-250.

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment and Evaluation in Higher Education*, *29*(5), 611–623. https://doi.org/10.1080/02602930410001689171

Fosnacht, K., Sarraf, S., Howe, E., & Peck, L. K. (2017). How important are high response rates for college surveys? *The Review of Higher Education, 40*(2), 245-265.

Germain, M. L., & Scandura, T. A. (2005). Grade inflation and student individual difference as systematic bias in faculty evaluations. *Journal of Institutional Psychology*, *32*(1), 58-67.

Goos, M., & Salomons, A. (2017). Measuring teaching quality in higher education: Assessing selection bias in course evaluations. *Research in Higher Education, 58*, 341-364.

Gravestock, P., & Gregor-Greenleaf, E. (2008). Student course evaluations: Research, models and trends. *Toronto: Higher Education Quality Council of Ontario*, 150. http://www.heqco.ca/SiteCollectionDocuments/Student Course Evaluations.pdf

Hatfield, C. L., & Coyle, E. A. (2013). Factors that influence student completion of course and faculty evaluations. *American Journal of Pharmaceutical Education*, *77*(2), 1–4. https://doi.org/10.5688/ajpe77227

Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations: What all faculty should know. *College Teaching, 49*(1), 26-31. https://doi.org/10.1080/87567550109595842

Hoefer, M. P., Yurkiewicz, J., & Bryne, J. C. (2012). The association between students' evaluation of teaching and grades. *Decision Sciences Journal of Innovative Education, 10*(3), 447-459. https://doi.org/10.1111/j.1540-4609.2012.00345.x

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education, 4*(1), 1-8. https://doi.org/10.1080/2331186X.2017.1304016

Iqbal, I., Lee, J. D., Pearson, M. L., & Albon, S. P. (2016). Student and faculty perceptions of student evaluations of teaching in a Canadian pharmacy school. *Currents in Pharmacy Teaching and Learning*, *8*(2), 191–199. https://doi.org/10.1016/j.cptl.2015.12.002

Jaquett, C. M., VanMaaren, V. B., & Williams, R. L. (2017). Course factors that motivate students to submit end-of-course evaluations. *Innovative Higher Education, 42*, 19-31.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*(4), 91–114. https://doi.org/10.1257/089533002320950993

Lambert, A. D., & Miller, A. L. (2014). Lower response rates on alumni surveys might not mean lower response representativeness. *Educational Research Quarterly, 37*(3), 40-53.

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*, 291-303. https://doi.org/10.1007/s10755-014-9313-4

Marks, P. (2012). Silent partners: Student course evaluations and the construction of pedagogical worlds. *Canadian Journal for Studies in Discourse and Writing/Rédactologie*, *24*(1). https://doi.org/10.31468/cjsdwr.16

Marlin, J. W., & Niss, J. F. (1980). End-of-course evaluations as indicators of student learning and instructor effectiveness. *The Journal of Economic Education, 11*(2), 16-27.

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal, 30*(1), 217-251. https://doi.org/10.3102/00028312030001217

McMillan, J. H. (2018). *Classroom assessment: Principles and practice that enhance student learning and motivation* (7th ed.). Pearson Education.

Murray, H. G. (1997). Does evaluation of teaching lead to improvement of teaching? *International Journal for Academic Development, 2*(1), 8-23. https://doi.org/10.1080/1360144970020102

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching, *Assessment & Evaluation in Higher Education*, *27*(2), 187-198. https://doi.org/10.1080/02602930220128751

Ng, A. K., Kiang, K. M., & Cheung, D. H. (2016). Assessing students' attainment in learning outcomes: A comparison of course-end evaluation and entry-exit surveys. *World Journal of Education, 6*(3), 56-65. http://doi.org/10.5430/wje.v6n3p56

Ory, J. C. (2001). Faculty thoughts and concerns about student ratings. *New Directions for Teaching and Learning*, 87, 3–15. https://doi.org/10.1002/tl.23

Peterson, D. A. M., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS ONE*, *14*(5), 1–10. https://doi.org/10.1371/journal.pone.0216241

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile?: An analytical framework for answering the question. *Quality Assurance in Education, 15*(2), 178-191.

Senior, B. A. (2000). Student teaching evaluations: options and concerns. *Journal of Construction Education, 5*(1), 20-29.

Spencer, K. J., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education, 27*, 397-409.

Stanny, C. J., & Arruda, J. E. (2017). A comparison of student evaluations of teaching with online and paper-based administration. *Scholarship of Teaching and Learning in Psychology, 3*(3), 198-207. https://doi.org/10.1037/stl0000087

Steyn, C., Davies, C., & Sambo, A. (2019). Eliciting student feedback for course development: The application of qualitative course evaluation tool among business research students. *Assessment and Evaluation in Higher Education, 44*(1), 11-14. https://doi.org/10.1080/02602938.2018.1466266

Stowell, J. R. Addison, W. E., & Smith, J. L. (2012). Comparison of online and classroom-based student evaluation of instruction. *Assessment and Evaluation in Higher Education, 37*(4), 465-473. https://doi.org/10.1080/02602938.2010.545869

Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. *Perspectives on Psychological Sciences, 11*(6), 800-816. https://doi.org/10.1177/1745691616650284

Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research,* 109*,* 45–56. https://doi.org/10.1002/ir.3

Uttl, B., White, C. A., & Gonzalez, D. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22-42. https://doi.org/10.1016/j.stueduc.2016.08.007

Veeck, A., O'Reilly, K., MacMillan, A., & Yu, H. (2016). The use of collaborative midterm student evaluations to provide actionable results. *Journal of Marketing Education, 38*(3), 157-169.

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education*, *23*(2), 191-212. https://doi.org/10.1080/0260293980230207

Wolbring, T., & Treischl, E. (2016). Selection bias in students' evaluation of teaching: Causes of student absenteeism and its consequences for course ratings and rankings. *Research in Higher Education*, *57*(1), 51–71. https://doi.org/10.1007/s11162-015-9378-7

NOELIA PACHECO DIAZ is an Assessment Coordinator at Student Life Assessment and Strategic Initiatives at the University of Tennessee. She manages assessment plans, survey requests, and Key Performance Indicators (KPIs). Noelia's research interests include student engagement and higher education assessment.

JOHN P. WALKER is a Psychometrician at Renaissance Learning. Prior to this role, he served as a Research and Assessment Coordinator for Teaching and Learning Innovation at the University of Tennessee, Knoxville. His research interests include psychometrics and higher education assessment.

LOUIS M. ROCCONI is an Assistant Professor in the Evaluation, Statistics, and Methodology program at the University of Tennessee, Knoxville. He is also an affiliated faculty member with the Intercollegiate Graduate Statistics and Data Science Program. His research interest include program evaluation and assessment in higher education and methodological issues in educational research.

JENNIGER ANN MORROW is an Associate Professor in the Evaluation, Statistics, and Methodology program at the University of Tennessee, Knoxville. She is also an affiliated faculty member with the Intercollegiate Graduate Statistics and Data Science Program. Her research interests include higher education assessment and evaluation as well as methodology training for emerging educational professionals.

GARY SKOLITS is an Associate Professor in the Evaluation, Statistics, and Methodology program at the University of Tennessee, Knoxville. His research interests encompass program evaluation, higher education governance as well as methodology training for emerging educational professionals.

JESSICA D. OSBORNE is the Director of First-Year Programs at the University of Tennessee, Knoxville. She recently completed her PhD in Evaluation, Statistics, and Measurement from UT, Knoxville. Her research interest includes program evaluation and assessment in student success for undergraduate and graduate student populations.

RICHARD PARLIER is an Assessment Specialist at the Georgia Department of Education in Atlanta, Georgia. He provides quality control for the data and reporting of K-12 standardized assessments. His research interest includes data analytical decision-making in higher education, K-12 standardized assessments, and psychometrics.

Appendix

End-of-Course Evaluation Survey

Instructions: Thank you for participating in this survey regarding end-of-course evaluations. The data collected from the following questions will help us determine what aspects of the evaluation feedback you find *useful*. In addition, this survey will also inform what, if any, future improvements need to be made to end-of-course evaluations. For each survey item, please select the response or responses that most reflect how *useful* a particular aspect of feedback is to you. This survey should only take five minutes or less to complete. All responses will be kept strictly confidential.

Q1. How much does feedback from end-of-course evaluation inform the following aspects of the student experience?

None at all, A little, A moderate amount, A lot, A great deal

a. Students level of learning in the class
b. The required effort of the student
c. Level of difficulty with exams
d. Level of difficulty with assignments
e. Teacher feedback to the student regarding overall progress
f. Teacher feedback to the student regarding assignments/tests
g. Student interest in course
h. Student interest in assignments (readings, homework, labs, etc.)
i. Your knowledge of course contact
j. How effectively you explain contact
k. How effectively you use course time
l. How clearly you explain difficult material
m. How you interact with students

Q2. In general, how useful are end-of-course evaluations for planning your courses?

Not at all useful, Slightly useful, Moderately useful, Very useful, Extremely useful

Q3. How often during the school year do you use end-of-course feedback to make changes to your courses?

Never, Once a year, Twice a year, Three times a year, Four or more times a year

Q4. How often do you use end-of-course evaluation to make changes to your teaching practices?

Never, Once a year, Twice a year, Three times a year, Four or more times a year

Q5. How often do you share the results of end-of-course evaluation from previous courses with your current students?

Never, Sometimes, About half the time, Most of the time, Always

Q6. Overall, how useful are end-of-course feedback with improving your teaching/pedagogy?

Not influential at all, Slightly influential, Moderately influential, Very influential, Extremely influential

Q7. How much does feedback from end-of-course evaluations inform the following aspects of your course planning?

None at all, A little, A moderate amount, A lot, A great deal

a. Identifying issues/problems
b. Contact
c. Experiential methods
d. Learning outcomes
e. Assessment

Q8. How much does feedback from end-of-course evaluations inform the following aspects of your teaching/pedagogy?

None at all, A little, A moderate amount, A lot, A great deal

a. Lecture
b. Discussion
c. Self-instruction (assignments, readings, etc.)
d. Group projects
e. Demonstration (e.g., presentations)

Q9. Please list any other ways in which you have used end-of-course evaluations. Be as specific as possible.

Open-ended response

Q10. Other than end-of-course evaluations, what other methods do you use for assessing your courses?

Open-ended response

Q11. What is your current faculty appointment track?

Tenure track, Non-tenure track

Q12. What is your gender identity?

Male, Female, Other (with write-in response), Prefer not to say

Q13. How long have you worked as a faculty member/instructor at the [this institution]?

0-2 years, 3-5 years, 6-9 years, 10 or more years

Q14. What is your Race/Ethnicity?

African-American/Black, Caucasian/White, Hispanic, Asian/Pacific Islander, Middle Eastern, Native American, Mixed Race, Other, Prefer not to say