

Teaching Evaluation Practices in Engineering Programs: Current Approaches and Usefulness*

KEISHA A. VILLANUEVA, SHANE A. BROWN, NICOLE P. PITTERSON and DAVID S. HURWITZ
School of Civil and Construction Engineering, Oregon State University, Corvallis, OR 97331, USA.
E-mail: villanuk@oregonstate.edu, shane.brown@oregonstate.edu, pittersn@oregonstate.edu, david.hurwitz@oregonstate.edu

ANN SITOMER

School of Chemical, Biological and Environmental Engineering, Oregon State University, Corvallis, OR 97331, USA.
E-mail: ann.sitomer@oregonstate.edu

The evaluation of teaching is a critical aspect in higher education. There is substantial knowledge based on best teaching evaluation practices that can be used, yet there is a lack of research regarding teaching evaluation practices that are used specifically in engineering programs. This research characterized teaching evaluation practices within engineering programs across the country, to understand and assess the current state of practice. Three research questions were explored: (1) What teaching evaluation practices are used in engineering programs? (2) Which practices are used for formative and/or summative purposes? (3) What practices do engineering faculty report as useful? An exploratory sequential mixed-method design, utilizing interview and survey methods was used. In the qualitative phase, data were collected using semi-structured interviews followed by the quantitative phase, which included the development of a survey to more fully understand the evaluation practices of the interviewed participants. Thirty-four educators, including course instructors, department heads, and program coordinators in engineering programs participated. The result of this study showed that end-of-course student evaluation of teaching (SET) is the most common approach. In addition to SET, other approaches to teaching evaluation include classroom observation by peers or non-peers, evaluation of classroom materials, and student mid-course evaluations. It was also identified that the formative practices used mainly to gather student feedback or to improve faculty teaching are useful to the participants. Although there is substantial interest in improving teaching evaluation practices, generally current practices are still much different from identified best practices such as evaluating active learning approaches in the classroom, constructive alignment of outcomes, activities and assessments as well as the frequency and quality of feedback to students. The teaching evaluation system in engineering programs can be improved when educators become aware of and choose to adopt approaches that have been demonstrated to improve teaching and student learning.

Keywords: teaching evaluation; formative evaluations; summative evaluations; engineering; higher education

1. Introduction

A National Research Council (NRC) report proposed that “rigorous evaluation to improve teaching and learning must become integral to Science, Technology, Engineering, and Mathematics (STEM) departmental culture if the broad teaching missions of colleges and universities are to be achieved” [1, p. 5]. This report indicates that the use of evaluation in higher education to examine and improve teaching quality is increasing. However, evaluation requires an understanding of what to evaluate, how to evaluate, what data to collect and analyze, and how to implement teaching improvements based upon what is learned [2]. Evaluation is one tool that can be used to improve STEM postsecondary teaching. Moreover, evaluation is an important part of instruction and of learning because through evaluations, we can determine the effectiveness of an educator by measuring what students have learned [3].

Substantial research has been conducted on teaching evaluation in general. Student evaluation of teaching was found to be the most common

method for teaching evaluation in higher education [4]. Although it was suggested as a necessary source of evidence of teaching effectiveness, student evaluation of teaching was critiqued for being insufficient when used as the only data source because it often relies exclusively on the qualitative judgments of students rather than how the teaching supported student learning [5]. This critique has fueled the call for multiple sources of evidence in assessing teaching effectiveness [1]. Employing additional methods to evaluate teaching effectiveness can help ensure that all dimensions of teaching (e.g., content knowledge, pedagogy knowledge, pedagogical content knowledge), and the course (e.g., lectures, materials, resources, etc.) are explored [6]. Additionally, it is essential to have a clear purpose for the evaluation process used and to ensure that the process is a valid and reliable indicator of educator effectiveness [2]. However, the same method of teaching evaluation may not be valid for all subjects, student groups, levels and disciplines [7].

While there are many teaching evaluation methods available for implementation, little research has been conducted that investigates teaching evalua-

tion practices, particularly in engineering programs [1, 8]. The purpose of our research is to explore teaching evaluation practices in one discipline—engineering—at three different types of institutions. These are associate’s colleges, master’s colleges and universities, and doctoral universities. This study seeks to understand the perceptions of educators and administrators regarding the usefulness of evaluation practices. Investigating teaching evaluation in engineering is important since engineering expertise is measured by the knowledge, skills and attitudes gained from educational experience. Fink, Ambrose and Wheeler posit that, “the ultimate goal of changing engineering education is to change faculty practices to improve the quality of student learning” [8, p. 191]. Consequently, studies that explore the nature of teaching in engineering as well as how teaching is assessed as being effective or not is of much benefit to the discipline. The importance of preparing high-quality engineers, an important aspect of ABET accreditation, is a pertinent reason for conducting teaching evaluation research in engineering programs. Therefore, this study seeks to share teaching evaluation practices used in engineering departments with the broader community with the hope that engineering stakeholders will benefit from its findings.

2. Background

The purpose of teaching evaluation in higher education has changed significantly over the years [9]. Teaching evaluation in higher education has evolved from primarily relying on a department chair’s evaluation to a more systematic approach, including a variety of methods such as student end-of-course evaluations, peer reviews, peer visits, self-evaluations, document reviews, and evidence of achievement. In higher education, there have been multiple calls for evidence-based instructional practices (EBIPs), such as active learning, that align with evidence of student learning. “There is extensive evidence that active learning works better than a completely passive lecture. Despite this evidence, adoption of these evidence-based teaching practices remains low” [10, p. 1]. The NRC recommended that, “teaching effectiveness should be judged by the quality and extent of student learning, and that there are various teaching styles and methods that are likely to be effective” [1, p. 118]. Still, most colleges and institutions tend to focus more on evaluations that are easy to use and measure because a more robust evaluation system requires a larger investment [11].

“The terms assessment and evaluation are commonly used interchangeably although they have two distinct meanings. Both terms are often confused

because the same data can be used for both assessment and evaluation” [11, p. 305]. For example, final exams can be used to assess student learning in courses. Furthermore, it can be utilized in the evaluation of the effectiveness of a course. In this paper, the term evaluation is defined as a collection of evidence regarding teaching practices, whereas assessment is defined as the activity of measuring student learning. The distinction between formative and summative evaluation is described in Section 2.2 later in this paper. We also use the term educator to describe a person who provides instruction in a classroom.

2.1 What practices are used?

“Educational technology and research have produced tools and strategies to evaluate and enhance teaching effectiveness as well as insights into the underlying processes and mechanisms” [12, p. 221]. “The literature on the evaluation of college teaching varied from developing effective faculty evaluations to very specific research investigating a variety of methods for evaluating teaching” [13, p. 109]. The most frequently referenced methods include student end-of-course evaluations, peer observation, portfolios, evaluation of course materials, and self-evaluation by educators. However, there is no general agreement on how to evaluate teaching in higher education. There is no distinct formula to effectively evaluate teaching because there are considerable variations that exist across institutions and disciplines [1]. The authors contend that the utilization of teaching evaluation practices is affected by context and culture. The NRC reports there are differences in the cultures of research and teaching in higher education, as well as the criteria used in the evaluation of teaching by disciplines were examined. For example, teaching effectiveness is less valued when compared to research productivity by many colleges and universities. In community colleges, the evaluation of teaching and learning has received greater attention [1]. Therefore, it is proposed that effective teaching evaluation stems from a combination of multiple sources of evidence. For example, students (current, graduating seniors, and alumni), graduate teaching assistants, departmental and other faculty colleagues could provide evidence of effective teaching through evaluation and assessment. This group of individuals could provide critical information related to the educator’s role in students’ learning, approaches to teaching, mentoring, the currency of the materials the educator presents, level of student engagement or participation and leadership in improving undergraduate education [1, p. 3–4].

Student evaluation of teaching is the most common method used for evaluating teaching and

courses [12, 14]. This method is used in higher education to summarize the students' overall perceptions at the end of a term or semester. "Still, student's learning processes remain underexposed using this quantitative data analysis because the data collected focus on a narrow range of teaching behaviors, unreflective of the cognitive, and affective structures that they construct a partial image of teaching and learning" [7, p. 658]. In addition, "there also appears to be little evidence that SET forms and procedures measure or contribute to teaching quality" [15]. In response to these problems, Broder and Taylor conducted a survey of teaching evaluation in agricultural economics and related departments. They had three objectives, which included developing a theoretical model of the teaching evaluation process, examining departmental policies and assessments of teaching evaluation, and identifying factors associated with effective teaching evaluations. The data used for this study were taken from a survey conducted in 1988 where department heads in the United States ($n = 51$) and Canada ($n = 8$) were asked to complete a questionnaire on teaching evaluation in their department and to submit a sample of copies of teaching evaluation forms and statistical summaries. From this study, "empirical models were estimated which characterize faculty reliance on SETs and faculty satisfaction with SET forms. The faculty was found to have specific preferences for SET form and content, and improvements in SETs are more likely to result from increasing the quality rather than the quantity of SET questions" [16, p. 153].

The most common method, and in some cases the only method, used for the evaluation of teaching was student end-of-course teaching evaluation [12, 14]. Several studies suggest that student end-of-course evaluations are reliable, and have a high level of validity because students are in the best positioned to judge particular aspects of teaching, and the classroom [17]. "The major strength of student evaluation is that their reliability and validity have received more empirical support than any other method of teaching assessment" [18, p. 424]. However, many studies argue that evaluations by students are invalid and unreliable when these student end-of-course evaluations are the only method used to assess and evaluate teaching. Regardless of being reported to be useful, evaluation by students was found to be neither the sole nor the best way to evaluate a course [1].

2.2 Formative and summative teaching evaluation practices

Generally, there are two types of evaluation practices: formative and summative. Formative evalua-

tion incorporates evidence from other educators or students during and after the course to improve teaching and student learning. This type of information can be used to make in-course changes. Examples of formative evaluations include conversing with students informally during the course to determine what is or is not working, formal weekly meetings with a group of class representatives or dialoguing with Teaching Assistants (TAs) [19]. Summative evaluations are done after the course and are used for a variety of purposes. These evaluations conclude the educator's overall performance or status and are used to make decisions about annual performance review, promotion and tenure. For example, summative follow-up evaluations by alumni can provide feedback as to what course material has proven to be particularly useful in the industry [19].

A variety of formative and summative practices are used in the evaluation of teaching. While "SET are typically used summatively and exclusively" [20, p. 14], other measures of teaching effectiveness such as mid-course evaluation, exit evaluation, and alumni evaluation have the potential to be used formatively. Student end-of-course evaluation is the most influential measure of educator performance, used for both formative and summative decisions [21]. Exit and alumni ratings provide information about the quality of teaching, courses, curriculum, admissions, and other topics on a programmatic level [21]. In addition, graduating students and alumni could provide evidence about the educator's role in their learning [1, p. 60]. Additional sources of evidence for teaching evaluation could provide information gathered from peers and non-peers (e.g., administrators). For example, peer review is an alternative practice [21] and composed of two activities: (a) peer observation of in-class teaching performance and (b) peer review of the written documents used by an educator in a course [21]. Peer ratings of teaching performance and materials were found to be the most complementary source of evidence to student end-of-course evaluations. These practices cover the aspects of teaching that students are not in a position to evaluate [21]. Lastly, the educator could provide self-assessment of his or her teaching strengths and areas for improvements through self-evaluation by submitting a teaching portfolio.

2.3 Call for improvements to teaching evaluation practices

Critiques of current evaluation practices have fueled the call for multiple sources of evidence in assessing teaching effectiveness during the last decade [1]. Wankat et al. suggest that students are not qualified to evaluate the educator's teaching, and they often

miss the richness of ideas, which can be obtained using other evaluation procedures [19]. Student evaluation of teaching may confuse the evaluation of the educator with an evaluation of the course [1]. There are also concerns that factors other than teaching quality, including the size of the class, course grade distributions, and whether it was being taken as an elective or a requirement can influence students' responses on such evaluations and, for this reason, student end-of-course evaluations may not be directly comparable to courses or educators [1].

In terms of evaluation, one single data source, whether from students, peers or mentors, or one single collection method such as interviews, questionnaires, or discussion, may provide insight on only one aspect of teaching. For instance, a study by Iqbal showed that a peer review, which usually consists of observation and review of materials, along with student evaluations might be useful for informing the evaluation of teaching because students are best at providing feedback on the quality of student-educator interactions. On the other hand, colleagues are better positioned to comment on matters such as content expertise, instructional design, and methods of assessment [16]. Several different sources and methods should be utilized to provide a more comprehensive overview and to ensure a range of teaching processes are explored. Each of these sources can provide unique information, but when used alone each has limitations [1]. A combination of techniques can make up for the deficiencies of student end-of-course evaluations. Evaluations of teaching by students need to be supplemented by other methods, for example, peer observation or review of course artifacts, such as syllabi, course activities, and assessments [23]. Colleagues in the discipline are able to determine whether course activities are appropriately challenging and accurate [17]. By drawing upon three or more different sources of evidence, the strengths of each source can compensate for the weaknesses of the other sources, thereby, converging on a decision about teaching effectiveness that is more accurate than one based on any single source [21, 23]. For instance, Berk proposed a unified conceptualization of teaching effectiveness where evidence must be collected from a variety of sources to define the construct and to make decisions about its attainment. Each source can supply unique information, but can be weak, usually in a way different from the other sources [21].

Several comprehensive models of faculty evaluation have been proposed. They include multiple sources of evidence with a greater weight attached to student and peer input and less weight attached to self-evaluation, alumni, and administrators. All of

these models can be used to arrive at formative and summative decisions [21]. "For any formative or summative evaluation, it was found that assessment based on a single teaching activity (i.e., classroom presentation) or dependent on information from a single source (i.e., student evaluation forms) is less reliable, useful, and valid than an evaluation of an educator's strengths and weaknesses based on multiple sources" [1, p. 51]. Comprehensive evaluations of teaching are more accurate, particularly when based on the views of current and former students, colleagues, and the educator or department being reviewed. The process of evaluating teaching has been found to work best when all faculty members in a given department play a strong role in developing policies and procedures. "This is the case because evaluation criteria must be clear, well known, and understood, scheduled regularly, and acceptable to all who will be involved with rendering or receiving evaluation" [1, p. 51–52].

"Research suggests that some combination of formative and summative evidence about student learning can be helpful in evaluating and improving an educator's teaching" [1, p. 2–3]. It is known that there are practices that can contribute to effective teaching and learning criteria for assessing teaching performance [1]. Despite knowing that such teaching evaluation practices already exist, it is still unclear what quality of the measures provides evidence of teaching effectiveness and student learning. Surprisingly, little is known about what teaching evaluation practices are used, especially in engineering programs, and what practices are actually working for engineering educators in their department or institution.

2.4 What practices work for engineering educators?

There are a variety of teaching evaluation practices available for implementation. However, there are far fewer that are frequently used. The adoption of evidence-based practices is a complicated process that involves social, cultural, and individual factors. The perception of value is an important aspect of the adoption process; what educators believe is most effective is influential on what is adopted by the faculty. However, research has not identified which practices educators perceive as effective. We do not argue that what educators believe is effective actually is, but rather that effectiveness, as noted above, is an elusive goal, and perceived effectiveness may lead to increased adoption of effective and holistic practices that would be of benefit to the community.

Research in the constructivist paradigm suggests knowledge is constructed and adopted within a community when it is seen as "useful, practical and adaptive" [24]. It can be argued that when each individual member has the opportunity to

share in the development, implementation, or enactment of a particular practice embodied by the community more value is placed on said practice [25]. With respect to perspectives on teaching evaluation practices, participants' knowledge of these practices can be deemed as constructed within the social context of their institutional mission as it relates to how important effective teaching practices and evaluation are. Consequently, since human practices are constructed in and out of their interaction with each other within an essential social context, it is necessary to gauge the perspective of these individuals [25]. Building on the idea of constructivism, the authors posit that studying participants' perspectives on teaching evaluation practices utilized within their engineering programs is an important and valid area of research.

The objective of this research study is to characterize best practices for assessing teaching evaluation within engineering departments across the country. To meet this goal, three research questions must be addressed.

1. What teaching evaluation practices are used in engineering programs?
2. Which practices are used for formative and/or summative purposes?
3. What practices do engineering faculty report as useful?

3. Method

3.1 Study design

This research study is based on a constructivist paradigm, which "recognizes the importance of the subjective human creation of meaning, but does not reject outright some notion of objectivity" using a descriptive case study with multiple embedded units [26, p. 549]. This case study was used to describe the data collected as they occurred in various institutions. The data were gathered using an exploratory sequential mixed-method design to answer the research questions, which involved collecting both qualitative and quantitative data [27]. The qualitative data were collected through semi-structured interviews; the quantitative data were gathered from the survey that was developed after the preliminary analysis of the data collected from the interview. The survey was developed to further understand the practices of the research participants and to gather additional information that was not captured from the interview. Collectively, the interview and survey data provided a holistic overview of evaluation practices used by the participants and the value of such practices, as reported by the participants. This study was reviewed and approved by the Institutional Research Board (IRB).

3.2 Participant characteristics

Individuals with knowledge regarding teaching evaluation in their department and/or institution were invited via email to participate in this study. Participants included course instructors, department heads, and program coordinators in engineering programs at doctoral universities, master's colleges and universities, and associate's colleges based on the Carnegie classification of institutions. Invited participants contact information was found from their college or university website and through personal contacts. Snowball sampling was used to identify other colleagues who were recruited and whose contacts were provided by the existing participants. Upon expressing interest in participating in the research study, the individuals were contacted again to set up a telephone interview and were provided with the IRB approved consent form. The participants' location varied, but a large proportion of the participants were from the western region of the United States ($n = 23$). The rest of the participants were from Midwest ($n = 4$), Southwest ($n = 2$), Northeast ($n = 1$) or Southeast regions ($n = 4$). Table 1 shows the number of participants and their role in their department and/or institution. Fig. 1 provides the breakdown of participant by institution types. Each participant mentioned having more than one role in their department, though Table 1 shows the participants' primary role.

Table 1. Number of participants and their role in their department and/or institution

Participant's department role	Number of respondents
Department Chair	15 (44%)
Associate Department Chair	1 (3%)
Full Professor	1 (3%)
Associate Professor	6 (18%)
Assistant Professor	3 (9%)
Engineering Instructor ^a	6 (18%)
Program Coordinator	2 (6%)

^a These include educators from associate's colleges and non-tenured-track educators from universities.

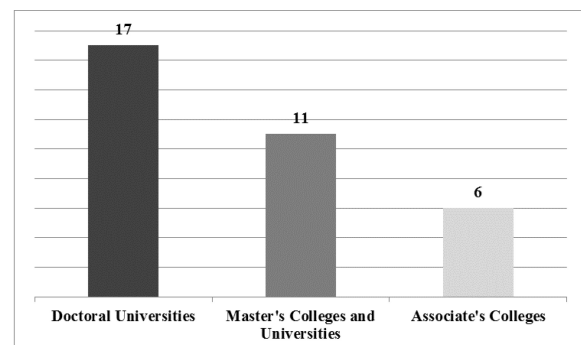


Fig. 1. Participant's breakdown using Carnegie classification of institutions.

3.3 Data collection

Data were collected using semi-structured interviews and a survey of teaching evaluation practices.

3.3.1 Semi-structured interviews

In the qualitative phase of this study, data were collected through semi-structured interviews conducted over the phone. An interview protocol was created to question the participants about the teaching evaluation practices in their department. The participants were initially asked about their background followed by the main question,

“What practices have you used or are currently using to evaluate teaching in your department?” Overall, they were asked to report what, and how teaching evaluation practices were conducted in their department. However, participants also reported their *own* practices in conducting teaching evaluations. Additional questions were asked based on the participant’s responses in order to elicit more details about particular practices the participants mentioned. Fig. 2 shows examples of the questions asked in the interview. The duration of each interview ranged between 15–60 minutes in length.

<p>Participant’s Background</p> <ul style="list-style-type: none"> • What is the name of your university? • What is the name of your department? • Could you please tell me your role in your department? <p>What practices have you used or are currently using to evaluate teaching in your department? (e.g., student end-of-course evaluation, classroom observation, and evaluation of classroom materials)</p> <ul style="list-style-type: none"> • What is the purpose of this teaching evaluation practice? • Could you please describe the process for this method? <p>Based on the participant’s response, clarification of the teaching evaluation practices may be needed. Examples 1 and 2 provide some follow-up questions that may be asked to the participants if they indicated using student evaluation of teaching and peer review.</p> <p>Example 1: Student Evaluation of Teaching (SET) Questions</p> <ul style="list-style-type: none"> • What is the purpose of the SET in your department? • How it is administered, online or paper-based and why? • Who determines what questions are on the SET? • Are the questions on SET standard? • Please provide some examples of the questions asked on SET? • Can students write comments on the SET? • Can instructors add their own questions on SET? • When do students usually take the survey? • Can students still take the survey after grades have been posted? • Are the evaluations published for students and for public to see? Explain <p>Example 2: Peer Review Questions</p> <ul style="list-style-type: none"> • What is the purpose of the peer review in your department? • What does a single peer review consist of? <ul style="list-style-type: none"> ◦ Observation of single or more lectures ◦ Review of syllabus, course notes, assignments, exams, etc. • Who is the evaluator (peer from the same and/or different department)? <ul style="list-style-type: none"> ◦ Does the evaluator get any form of training? ◦ Does he use a checklist or standard form during an evaluation? • What documentation is provided to the faculty member? <p>Does your institution have a Center for Teaching and Learning (CTL) or equivalent?</p> <ul style="list-style-type: none"> • What services does your CTL provide related to teaching evaluation and assessment? <p>Participants’ Perspectives</p> <ul style="list-style-type: none"> • What do you think about the teaching evaluation practices in your department? Are they effective or ineffective and please explain why? • What changes would you like to see regarding the teaching evaluation practices in your department? <p>Could you provide us any materials that are used for teaching evaluation in your department?</p>
--

Fig. 2. Interview questions.

3.3.2 Qualtrics survey

The interview protocol was adapted throughout the study to add questions based on what was learned during previous interviews. For this reason, a supplemental survey was developed to gather more information and to clarify some of the findings that emerged during the interviews with the participants. The survey questions contained similar questions to those from the interview, but it further explored the practices of the interview participants. The survey was distributed to interview participants via email. The time necessary for participants to complete the survey ranged between five and 30 minutes; they were given one month to complete the survey.

There were four different sections and 13 primary questions on the survey. The first section of the survey was intended for the collection of background information, which included institution and department name, average number of students who complete the engineering program annually, and degrees that the institution offers. The second section of the survey consisted of questions related to teaching evaluation practices gathered from either the educator's peers or administrators (i.e., classroom observation and evaluation of classroom materials). The third part asked for teaching evaluation practices gathered from the students (i.e., student end-of-course evaluation, student mid-course evaluation, exit evaluation, and alumni evaluation). Finally, the survey asked whether the participant's institution has a Center for Teaching and Learning (CTL) or equivalent. Finally, participants were asked whether any other teaching evaluation practices not mentioned in the survey were used. Upon indicating the use of a particular

method, sub-questions asked participants to elaborate on that method. However, if the participant indicated that their department did not use a certain method, the survey skipped the sub-questions as they would not apply to these participants. Table 2 shows the 13 primary questions asked on the survey protocol.

3.4 Data analysis

The audio recordings were transcribed either via a professional transcription company or by graduate researchers. Transcriptions were uploaded into Dedoose, a web application for managing and analyzing qualitative and mixed methods research data [28]. The participants' transcriptions were coded and arranged into four categories: method, purpose, procedure, and participant's perspective. For example, the number of participants who used student end-of-course evaluation were grouped together and counted to identify the total number of this approach. The same analysis was completed with other teaching evaluation practices using the transcription data. Then, these data were combined with the data collected from the survey to summarize and describe the overall data set that represents the study sample.

While the data were initially separated into three categories based on institution types, the unequal distribution of participants in each type meant it was not possible to provide meaningful comparisons of teaching evaluation practices across institutions types. Half ($n = 17$) of the total participants came from doctoral universities, 11 of the participants were from master's colleges and universities, and the rest came from associate's colleges ($n = 6$). Instead, the data were grouped based on their

Table 2. Primary survey questions

1. Context (Background)	What is the name of your college or university? What is the name of your department or school? On average, how many undergraduate students complete your engineering program annually? On average, how many graduate students complete your engineering program annually? What degrees are offered?
2. Teaching Evaluation Practices gathered from either Peers or Non-Peers	Does your school or department use classroom observation? Does your school or department use classroom materials evaluation?
3. Teaching Evaluation Practices gathered from Students	Does your school or department use student end-of-course evaluation? Does your school or department use student mid-course evaluation? Does your school or department use exit evaluation? Does your school or department use alumni evaluation?
4. Other Teaching Evaluation Practices (Programs such as Center for Teaching and Learning)	Does your college or university have a Center for Teaching and Learning or equivalent that provides support and services for faculty members related to teaching evaluation and/or teaching improvement? Does your school or department use any other teaching evaluation practices that are not mentioned?

Table 3. Categories in Dedoose and description

Categories in Dedoose	Description of what is included in each category
Method	Variety of approaches that are currently used to evaluate teaching in engineering departments and by individual participants such as student end-of-course evaluation, student mid-course evaluation, classroom observation (peers and non-peers), alumni evaluation, exit evaluation, review of classroom materials and classroom assessment techniques
Purpose	Formative evaluation, summative evaluation, or both. For instance, promotion, tenure, teaching improvement, accountability, feedback, annual review, disciplinary action, newly hired assessment
Procedure	Process for how evaluation methods were enacted in the department. For example, voluntary or mandatory, formal or informal
Participant's Perspective	Perspective regarding what teaching evaluation practices were found to be useful by our participants

purposes and usefulness for further analysis. Table 3 provides a short description of the four main categories.

4. Results

4.1 What teaching evaluation practices are used in engineering programs?

The participants commonly referenced the following approaches: classroom observation by peers and/or non-peers, evaluation of classroom materials, student end-of-course evaluation, student mid-course evaluation, exit evaluation, and alumni evaluation. Results gathered from exit and alumni evaluations were sometimes used to provide formative feedback to educators, but more frequently for students to provide feedback about their experience in the engineering program. Some participants mentioned additional methods such as teaching portfolios, self-assessment, and informal in-class assessment techniques. However, the use of these methods to evaluate teaching was not explicitly discussed. Shown below are more detailed accounts of the practices used at participating institutions.

4.1.1 Classroom observation

Classroom observation by peers and/or non-peers:

A. Classroom observation by peers.

With this method, the evaluator is usually a senior colleague or tenured faculty member. About 60% ($n = 20$) of peer evaluators were from the same department. This process is typically voluntary, where the educator asks their peers to come visit a single lecture. The purpose varied by departments; it could be used for promotion purposes or to gather information for formative feedback as part of a program review, for example. However, this method is usually used to obtain constructive feedback to help improve the individual educator's teaching. This method was also used as part of a mentorship program to assist newly hired educators.

B. Classroom observation by non-peers.

A non-peer is considered a dean, program director, tenure committee member, department head or an external evaluator, such as someone who works with the institution's Center for Teaching and Learning (CTL), who is knowledgeable about pedagogical strategies or the content being taught. Non-peer classroom observation was typically evaluated by the following: department chairs ($n = 14$), CTL personnel ($n = 8$), deans ($n = 7$), committee members ($n = 3$), and other external teaching evaluator ($n = 1$). This process could be voluntary or mandatory depending on the department. The information gathered may be used to provide constructive feedback to the educator. However, it was usually conducted for promotion purposes, annual review, or in response to student complaints or low scores on student evaluation of teaching.

Overall, the process for conducting a classroom observation was similar when conducted by both peers and non-peers. Twenty-nine out of 34 participating institutions used classroom observation by peers or non-peers to evaluate an educator's teaching. The process usually consists of three parts: pre-observation, actual classroom visit, and post-observation. During the pre-observation, the educator, and the observer meet to clarify expectations for the evaluation and the best way to conduct the observation. The observer may also request classroom materials for reviews, such as syllabus, exams, projects, assignments or quizzes. Then, the observer visits an educator's lecture. The observer evaluates the educator's teaching performance by focusing on the educator's content knowledge, delivery, teaching methods and learning activities. The actual observation is followed by a post-observation where the observer and educator meet again to discuss what went well during the observation and items that may need attention to help improve the educator's teaching performance.

Fourteen participants revealed that educators only receive a single classroom observation per

course. The number of observations or whether an educator gets evaluated using this method depends on their rank, position, and employment status (e.g., full-time or part-time faculty). For instance, at doctoral universities, master's colleges and universities, tenure-track and adjunct instructors were sometimes the *only* ones who got evaluated or they received more evaluations than tenured faculty members. The tenured faculty members either do not get evaluated or if they did, they were not evaluated as regularly as pre-tenure and adjunct educators. Some associate's college participants said that they receive very few observations once they achieved continuous employment.

Eight participants indicated that their department used a standard form or checklist during the classroom observation. Twenty-one participants indicated that their department did not have a standard form or checklist for evaluators to use during the observation. Some evaluators were provided a template as a guide for writing their observations while most evaluators wrote their observations without a template. About 85% ($n = 29$) of the total participants stated that the evaluator or observer did not receive any sort of training or if they did, it was minimal. If training occurred, it usually consisted of a workshop or ongoing training provided by the CTL.

4.1.2 Evaluation of classroom materials

Participants revealed that classroom materials were typically requested and evaluated as part of the classroom observation process. Fifty percent ($n = 17$) of the total participants indicated that their department collects and evaluates classroom materials to assess faculty teaching. Different types of documents, such as course materials (syllabus, and assignments) and/or result of student works (exams, projects, and quizzes), were evaluated. The evaluation of course materials usually occurred when the purpose was to provide formative feedback to the educator. Certain types of documents were additionally requested when the information gathered from this type of evaluation was intended to be used for promotion purposes, annual review, and/or ABET accreditation. In engineering disciplines, ABET accreditation provides evidence that the program meets the essential standards needed to ensure graduates are ready to transition to work as an engineer. It was found that a peer from the same department was the most common choice to evaluate these materials. However, non-peers such as deans, department chairs, an ABET reviewer, or a tenure committee member might evaluate these materials. Many of the individuals evaluating course materials were not trained and did not

have a standard form or checklist to review these materials.

4.1.3 Student end-of-course evaluation

All study participants indicated that they used student end-of-course evaluations, yet they did not all conduct this practice in the same way. There were two methods of student evaluations of teaching: online and the traditional paper format. About 68% ($n = 23$) of the participants used an online student end-of-course evaluation to which students were provided a link, while 26% ($n = 9$) of the participants used the traditional paper-based format to survey their students. The remaining participants ($n = 2$) indicated using both methods. The traditional paper-based format survey was given during an educator's class. Depending on the department, the student end-of-course evaluation results can also be seen and accessed by the dean, program chair, review committee, or tenure committee. Many of the participants mentioned that their department had switched from paper-based evaluations to online evaluations. However, our data show that as institutions considered moving to online evaluations, educators' concerns about response rates emerged.

Some participants (6%, $n = 2$) indicated there were different forms of the student end-of-course evaluations from which they could choose that depend on the class level or size of the class that they are teaching. The majority of the participants (94%, $n = 32$) said that there was only one form that they could use. The questions on the student end-of-course evaluations were standardized either by the institution, by the college or school of engineering, or by the department only (e.g., civil, mechanical, chemical, electrical). Nearly 62% ($n = 21$) of the participants stated that their institution used a set of standardized questions on the student end-of-course evaluations. Four participants discussed that their department developed their own questions for the student end-of-course evaluations, and three participants indicated that the entire college or school of engineering has their own questionnaire for student end-of-course evaluations. Half of the participants ($n = 17$) said that they were able to add supplemental questions on the SET to help them improve their teaching and to measure student learning. However, some participants indicated that although adding questions was possible, not many educators took advantage of this opportunity.

4.1.4 Student mid-course evaluation

Approximately 74% ($n = 25$) of the participants used student mid-course evaluations to assess their own teaching. This evaluation was completed in various

ways depending on the educator's preference. Overall, educators undertook student mid-course evaluations voluntarily, that is, the information gathered from this type of evaluation was only seen and used by the educator. There were different ways to administer this type of evaluation. The educator conducting his or her own student mid-course evaluation was the most common way. For example, some of the participants specified using different classroom techniques such as minute paper, muddiest point paper, or having their students write to gather feedback about their teaching and to assess student learning. Only a few of the participants specified asking another, such as a person from a Center for Teaching and Learning, to come and perform mid-course evaluations, providing feedback on how the educator may improve his or her teaching. The intention was normally to facilitate assessment and improvement of teaching for newly hired educators or an educator teaching a class for the first time, or to help an educator trying a new teaching technique. Questions that educators asked their students when administering a student mid-course evaluation included questions such as:

- “What is working in the classroom that helps enhance the student learning?”
- “What specific improvements would help the student learn better?”
- “What could the educator do differently to help this particular class?”

4.1.5 Exit evaluation

Exit evaluations were given mostly to graduating seniors before they completed their engineering program. This typically applied to students who graduated from institutions such as doctoral universities and master's colleges and universities. For associate's colleges, this evaluation was given to students who were transferring to a four-year institution. The main purpose of the exit evaluation was to assess what the students have learned and to guide longer-term assessment of the program. It was also used to make adjustments in the engineering program in terms of course offerings or who teaches the courses. This process was not used to evaluate a specific course, although students could comment on specific courses. Exit evaluations assess the overall quality of the courses taught in the department. Generally, this was used to help the department identify strengths and areas for improvement, and eventually to implement a change to improve the program. However, some educators indicated that they used the results from exit evaluations as formative feedback to improve their teaching as well as student learning. This type of evaluation involved questions such as if the

students were satisfied with the program, changes the students would like to see within the program, and if they found themselves ready to enter the workforce. Below are sample questions from exit evaluations:

- “What did you do in your class?”
- “What did you learn?”
- “What do you think about the overall quality of education in your program?”

Students were either given a form to fill out or a link to the survey. However, some participants mentioned administering or accompanying the survey with an oral interview where the students were given an opportunity to discuss and explore their answers in more detail. Approximately 71% ($n = 24$) of the participants practiced this method in their department. Nearly 15% ($n = 5$) of these participants indicated using this type of evaluation as part of the ABET accreditation program. About 6% ($n = 2$) of the participants indicated that the senior exit survey was part of the National Survey of Student Engagement (NSSE), which gathers information nationally about students' participation in programs and activities at the institution.

4.1.6 Alumni evaluation

Approximately 32% ($n = 11$) of the total participants administered an alumni evaluation. The alumni evaluation was very similar to the exit evaluation; however, it was usually given to previous students who have successfully completed the engineering program. About 21% ($n = 7$) of the participants indicated using this as a formative assessment, and others ($n = 5$) as part of the ABET accreditation process. Just like the exit evaluation, the alumni evaluation was typically not used to evaluate an individual educator. Rather, it was focused on the engineering program as a whole. Yet, some educators used results from alumni evaluations to help improve their teaching when the students provided comments regarding specific instructors or courses.

4.1.7 Practices usage summary

Table 4 summarizes the teaching evaluation practices used. The most common method used to evaluate teaching within engineering departments was a student end-of-course evaluation. It was followed by classroom observation with 85% utilization. The student mid-course evaluation was used by 74% of participants, although this study revealed that departments did not require this method.

Table 5 shows the methods usage by institution type using the Carnegie classification of institutions. Doctoral universities, master's colleges and universities, and associate's colleges used similar evalua-

Table 4. Teaching evaluation practices usage summary

Teaching evaluation practices	Usage by number of total participants
Student end-of-course evaluation	34 (100%)
Classroom observation either by peers or non-peers	29 (85%)
Student mid-course evaluation	25 (74%)
Student exit evaluation	24 (71%)
Evaluation of classroom materials	17 (50%)
Alumni evaluation	11 (32%)

tion approaches. For example, student end-of-course evaluation was used 100% of the time at all three institution types investigated. Similarly, classroom observation either by peer or non-peer and student mid-course evaluation had comparable usage across all three institutions. Master's colleges and universities institutions reported the highest usage of classroom material evaluation with 73%. Doctoral universities tend to utilize student exit evaluation and alumni evaluation more frequently than other institution types.

4.2 Which practices are used for formative and/or summative purposes?

The exit and alumni evaluations were not usually used to evaluate teaching but to assess the program as a whole. They were both used for formative feedback or for ABET accreditation purposes. Table 6 shows the most common methods used to evaluate an educator's teaching, its purpose, and how it is used. In this study, mandatory evaluations correspond to teaching evaluation methods that were required of an educator, while voluntary evaluations were methods that educators used at their own discretion.

Information gathered from classroom observation was commonly used for promotion purposes ($n = 11$), annual review ($n = 9$), and teaching improvement ($n = 17$), regardless of who the observer was: peer or non-peer. Fourteen participants indicated that both peers and non-peers could be the evaluator. Other participants indicated that only their peers ($n = 7$) or non-peers ($n = 8$) evaluated them. The review of classroom materials that may accompany classroom observations were reported by around 26% ($n = 9$) of the total participants to be used for promotion purposes, 24% ($n = 8$) for annual review, 18% ($n = 6$) for teaching improvement, and 21% ($n = 7$) for ABET accreditation purposes. The student end-of-course evaluation was used for two very different purposes within departments: formative and summative. Nearly 15% ($n = 5$) of the participants described using these evaluations both solely and formatively to improve teaching. The same number of participants used the student end-of-course evaluations solely for making summative decisions for educators' promotion, tenure, or annual review. Many participants ($n = 24$) used the student end-of-course evaluation information for making both formative and summative decisions. All participants ($n = 34$) used the results gathered from the student mid-course evaluation as formative feedback. Fig. 3 provides the number of teaching evaluation practices usage by purposes, formative or summative.

4.3 What practices do engineering faculty report as useful?

Participants' perspectives on the effectiveness of teaching evaluation methods were explored with the interview question, "What do you think about

Table 5. Teaching evaluation practices usage summary by institution type

Teaching evaluation methods by institutions	Usage by participants' institution		
	Doctoral Universities	Master's Colleges and Universities	Associate's Colleges
Student end-of-course evaluation	17 (100%)	11 (100%)	6 (100%)
Classroom observation by peers or non-peers	15 (88%)	9 (82%)	5 (83%)
Student mid-course evaluation	12 (71%)	9 (82%)	5 (83%)
Student exit evaluation	15 (88%)	7 (64%)	2 (33%)
Evaluation of classroom materials	7 (41%)	8 (73%)	2 (33%)
Alumni evaluation	7 (41%)	3 (27%)	1 (17%)
Total number of participants	17	11	6

Table 6. Methods, purposes, and how teaching evaluation is used within engineering departments

Teaching evaluation methods	Purposes	How it is used
Classroom observation by peers and non-peers	Formative and Summative	Mandatory and Voluntary
Evaluation of classroom materials	Formative and Summative	Mandatory and Voluntary
Student end-of-course evaluation	Formative and Summative	Mandatory
Student mid-course evaluation	Formative	Voluntary

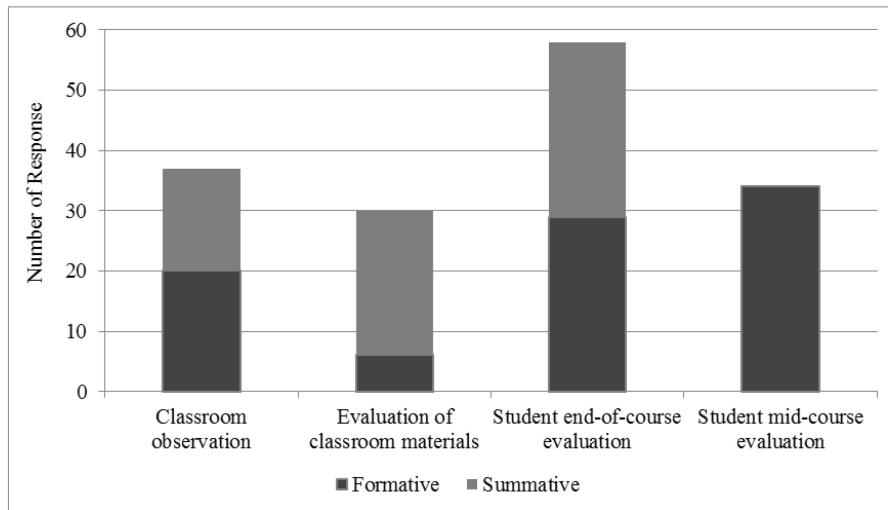


Fig. 3. Formative and summative teaching evaluation practices.

the teaching evaluation practices in your department?" The participants were also asked about what practices they found to be effective or ineffective. Table 7 shows the participants' perspectives about the collective teaching evaluation practices in their department, which were broken down into three parts: somewhat effective, ineffective, or neither.

Participants found two practices, student end-of-course evaluation and classroom observation, more useful for evaluating teaching when the purpose was to gather students' or colleagues' formative feedback to improve teaching, but less useful when used for promotion and tenure purposes. For instance, participants reported student end-of-course evaluation to be a useful tool because it gathers student feedback regarding the course and the educator's teaching; students spend more time in the classroom than an observer. However, participants felt it was beneficial to gather additional data, such as data from classroom observations. Examples of student end-of-course evaluation being effective based on two participants' response are shown below:

(a) Participant 150, "Student course evaluation is a very important instrument. It creates evidence when a faculty member is misbehaving or underperforming."

(b) Participant 1250, "I generally find it [SET] pretty helpful and effective. I think that it allows me to reflect on what I do pretty well, and so I don't find it to be too lacking or anything like that. Students are not hesitant to be critical. In terms of at least thinking about how I can improve, that's a good way to do that, I find the student evaluations to be more helpful."

Based on many participants' perspectives, they believed that student end-of-course evaluations and classroom observations were only useful to some extent and would benefit from improvements. The improvements or modifications that 44% of the participants ($n = 15$) would like to see were aligned with why participants believe that these two teaching evaluation practices were ineffective. Participants critiqued both classroom observation and the student end-of-course evaluation process. Suggestions from participants included using standardized forms or rubrics during a classroom observation that could provide specificity as well as remind the evaluator of what needs to be assessed. In addition, regular observations and using a trained observer, who knows what to evaluate and how, it would help improve classroom observation. Furthermore, the student evaluation of teaching (SET) process was weak and imperfect in some ways. These reasons include, but are not limited to, the type of questions being asked in the SET survey, timing, regularity of the survey, what it measures, consistency of the system, students as evaluators, and the purpose that it served. Table 8 provides some participant perspectives on the SET process.

Participants reported finding other voluntary and informal practices to be highly useful for improving their teaching. These included admin-

Table 7. Effectiveness of teaching evaluation practices based on participants' perspectives

Teaching evaluation practices effectiveness	Number of response
Somewhat Effective	15 (44%)
Ineffective	15 (44%)
Neither	4 (12%)
Total number of participants	34

Table 8. Reasons to improve the SET process based on sample participants' perspectives

Reasons to improve the SET process	Participants' response based on their own perspectives
Students as evaluators	<p><i>PARTICIPANT 1200:</i> "The student evaluation process I feel is highly flawed because it is so affected by the passion the students feel about their grade. I honestly, I really feel like that's huge. Students can't separate their feelings about the class and the message from the grade that they are earning. It's not very objective."</p> <p><i>PARTICIPANT 800:</i> "The SET score is informative, when I read my student comments, it is informative for me, and it is nice to know what they are thinking. Students are stressed out during the end of the term when they are given the SET and highly emotional."</p>
SET questions	<p><i>PARTICIPANT 1050:</i> "I would say on the student evaluation, I think we need to do better. All that it really boils down to is one or two questions. Was the teacher effective overall, or something like that? I think that's a bad way to evaluate somebody's teaching."</p> <p><i>PARTICIPANT 1300:</i> "I think our student evaluations are not effective. I think they really tend to be more of like I said earlier, do I like this person or do I not like this person? I have a hard time looking at my student evaluations because it feels more personal than it should be, and I think it has to do with the kind of questions that they're asked. I don't think that I generally get a lot of constructive information very often in the evaluations. I would like to see questions that are a lot more content-based."</p>
Regularity	<p><i>PARTICIPANT 1300:</i> "I also wish that there was maybe more of a formal process for students to do it more often than at the end of the term because the classes that I teach, I teach once a year. If somebody says something didn't work for him or her and then I have a whole year before I can change it. I find that when I look at these evaluations, I think if I had only known, I would have had the opportunity to fix it, versus getting it after the final."</p>
Timing	<p><i>PARTICIPANT 900:</i> "The end of the quarter evaluation, it's sort of like it's too late to do anything. Still, it gives you a pretty good sense of what happened, . . . That really helps over time. It helps you figure out what's effective."</p>
What it measures	<p><i>PARTICIPANT 750:</i> "I think it's really weak, I don't think it really tells us much about how instructors are actually helping students learn. I don't think there's much sure there are things associated with the actual practices of teaching but I think there are essentially zero assessment students learning how professors are actually getting their student to learn that's the weakness we have."</p>
Consistency	<p><i>PARTICIPANT 550:</i> "I think we definitely take it seriously. I think there are a few places where there are gaps, and there are... even with a really, really well-designed system, which we have, there still is room sometimes for interpretation which makes the system sometimes a little inconsistent. One professor might be praised for one thing and the next person who does a very similar thing might get a different type of feedback. That's a little troubling sometimes."</p>

istering a survey during the course, using classroom assessment techniques, and self-reflection or assessment. For example, surveys given during the course (mid-course evaluation) help educators know what areas they need to focus on to support struggling students. Additionally, using classroom assessment techniques help the educators improve their teaching because they receive a better response rate and more feedback from their students. Eight participants have completed self-assessments and teaching portfolios which allow them to reflect on their teaching to make necessary adjustments the next time they teach the same course. Five participants found it helpful to have short discussions during meetings, and office hours with their colleagues and students about what is working and not working in the class. The educators could talk and share their strategies that they use or should be using in the classroom with other colleagues, while students can discuss what is working and not working for them.

4.3.1 Role of rewards

Recognizing and rewarding educators' performance with teaching awards may be necessary to retain effective educators and perhaps to make teaching an attractive career choice [29]. For example, five study participants (15%) mentioned that there is an annual teaching excellence award where their peers and students nominate an educator. But, a majority of the participants said that there is no teaching reward being offered. According to participant 1300, "rewards or an incentive might help change the dynamic in the classroom because we would be incentivized to get good reviews so some people would really change their teaching approach."

5. Discussion

Study participants have expressed interest in identifying alternative ways to evaluate teaching that could be implemented their departments. They are also interested in learning about the practices that

have been adopted at other institutions for several reasons. These include identifying what teaching evaluation practices are available and useful, and to know what can be done to improve the current system.

5.1 Teaching evaluation practices are used for formative or summative purposes

The purpose of evaluating teaching was found to be either formative, summative, or both, depending on the department or institution. For instance, the evaluation method was used formatively when the results are used to identify areas of improvement for an individual educator and for professional development. It is summative when the results are utilized for promotion and tenure purposes and annual reviews. Teaching evaluation practices used for formative purposes are found to be more useful and effective by many participants because they get better response rates and the feedback they collect allows them to make adjustments and reflect on the course and their teaching to help students learn.

5.2 Importance, critiques and possible improvement of SET

“Student evaluations of teaching are now administered in almost all colleges and universities in the United States, and are becoming common in other countries” [30, p. 326]. Study participants discussed multiple methods used for teaching evaluation, yet it was identified that the student end-of-course evaluation remained the most commonly used. The purpose of SET is to collect data for personnel evaluation and to improve faculty’s teaching. The SET surveys are easy to conduct, inexpensive, and less time consuming than other approaches. However, some participants questioned the usefulness of the student end-of-course evaluation. For example, the questions asked on the survey are so general that they may be irrelevant to a particular class, and even if relevant, are worded so generally that they offer little guidance for improvement [30]. Some participants found that students often got confused whether they were evaluating the course or the instructor (Table 8). There are other factors such as course size and class difficulty that should be considered because every institution, even the departments and disciplines, are different. McKeachie suggested using a variety of forms for student evaluation of teaching that will be mostly relevant to the particular class that they are teaching [30]. Nevertheless, a large number of our participants still believe that the student end-of-course evaluation remains an important part of teaching evaluation.

5.3 Approximately 85% of the evaluators who observe the classroom are not trained

The majority of the participants used classroom observations along with student end-of-course teaching evaluations. Our study showed that most evaluators, peers, and non-peers who observe the classrooms, did not receive any type of training. An evaluator’s lack of training and/or not feeling qualified to evaluate teaching poses another challenge to reliability and validity of the method, as does the fact that they are based usually on only a small portion of faculty member’s teaching activities, such as a single classroom observation of teaching [31]. The authors suggest that it is essential that any evaluator must receive and continuously undergo training. It may also be beneficial to consider bringing in an expert evaluator who has knowledge about teaching evaluation and assessment for different disciplines and topics. Well-trained evaluators who know the process on how to conduct effective observation and how to use the specific peer observation instrument are ideally suited to evaluate an instructor’s teaching effectiveness [31].

5.4 Alignment with evidence-based teaching practices

A glaring omission from the literature is in the identification of practices and how they align with the evidence-based teaching practices (EBIPs). This omission is surprising because ideally teaching evaluation efforts would be aligned with best practices in course development and implementation, as well as with student learning. Alignment with student learning is very challenging due to several issues; high quality measures of student learning are not consistent across courses, measures of student learning vary widely across courses and offerings of a particular course, and determining any kind of correlational or causal relation between a particular aspect of practice and its effect on learning are nearly impossible at a large scale. Doing so would require significant attention to practices, learning, improvement of practices, and student work, and the relationship between each of these. However, there is an abundance of research that supports the efficacy of educational practices with student learning, as well as other outcomes [32]. It is therefore important that the link between research and teaching is maintained as “research can inform teaching in a myriad of ways such as providing information about how to teach better and how to help students learn better” [33, p. 381]. Additionally, these practices align with what is reasonably observable with common teaching evaluation practices.

We suggest that these observable core compo-

nents of a course include active learning in the classroom; the alignment of outcomes, activities, and assessments, and the frequency and quality of feedback to students. Active learning has been shown to be effective for student learning through multiple sources [32]. Active learning is operationalized in detail in the ICAP (Interactive, Constructive, Active, and Passive) framework [34], where it is shown that I, C, A, and P learning environments are decreasingly effective to student learning. The alignment of outcomes, activities and assessments are supported as a best practice in the educational literature. Two examples are constructive alignment and backward design [35, 36]. Both are based on using intentional instructional design to align the outcomes of a course with the types of learning activities and subsequent assessment measures. The focus of these approaches is to ensure that course outcomes are reflected and enacted in every facet of the course. This alignment can be observed through the course syllabus and examples of student assignments and assessments and how they align with outcomes. There is vast literature supporting the importance of the quality and quantity of feedback to students. For example, Juwah et al. posit that student learning is enhanced when formative feedback measures are used as teachers can often adjust their lesson delivery to meet the needs of students [37]. According to Shute, “the main aim of formative feedback is to increase student knowledge, skills and understanding in some content areas or general skill” [38, p. 156].

Ideally, both formative and summative teaching evaluation practices would be aligned with these best practices. However, our findings showed that there is a lack of alignment between the current evaluation practices and EBIPs previously defined. For instance, our study showed that the majority of classroom observers and evaluators of classroom materials receive minimal or no training. The classroom observation process usually only consists of a single classroom observation with one observer, and only half of the participants in this study indicated that classroom materials were evaluated. Based on these results, active learning in the classroom is probably not considered or impossible to observe during this process. Classroom observers often focus on the educator’s strengths and weaknesses such as their teaching skills, teaching methods, and content knowledge, and less on the connection between these skills and student-learning outcomes. Additionally, only eight participants revealed using standard forms or checklists in their department during classroom observation. There is also no protocol for document analysis to help the evaluators know what and how to evaluate the evidence they gather. Without any standard procedures and

proper training, this could lead to an educator getting very different scores and feedback depending on who observed them. The evidence gathered without using standard-based measures during document analysis and without proper training for observing teaching might be based on one individual’s judgment and does not provide evidence that might help an educator improve their teaching [39].

Despite our findings, we would like to provide descriptions of how currently used teaching evaluation practices might be better aligned with EBIPs, see Table 9.

Although there is substantial interest in improving teaching evaluation practices in engineering programs, it was identified that the existing practices are still somewhat different from the identified best practices in the literature.

6. Study limitations

A key limitation of this study is the lack of equal distribution among the three institution types studied. There were 34 participant institutions as highlighted in the methods section. However, of the 34 institutions, half ($n = 17$) were doctoral universities, 11 were master’s colleges and only six were associate’s colleges. With the disproportionate numbers, it was impossible to make meaningful comparisons across the three institution types. More research is necessary to investigate the reported usefulness of the practices identified in this study across different institutions with varying commitments to teaching and student learning. Additionally, 23 of the institutions were situated in the western region of the United States. Further research efforts could include more institutions in the Midwest and Eastern regions. Given the sample distribution across institution types and geographical location the sample is not a good representation of the overall higher education population. Despite these limitations, this study has highlighted some key teaching evaluation practices and has the potential to initiate the conversation among engineering educators and other interested stakeholders about improvements that can be made to teaching evaluation.

7. Conclusions

There are a variety of teaching evaluation practices available for implementation. However, this study showed that student end-of-course evaluation is the most widely used because it is fairly easy to administer, inexpensive, and requires less time than the other teaching evaluation practices. A large number of engineering programs also conduct classroom observations as part of teaching evaluation, but

Table 9. Aligning current teaching evaluation practices with EBIPs

Teaching evaluation practices	Description of how currently used teaching evaluation practices could be aligned with EBIPs
Classroom observations	The observer and the educator being evaluated must discuss the evidence obtained in the observations. The conversation should incorporate evidence of student learning so that educators focus primarily on whether students are learning rather than focusing only on instruction [34].
Evaluation of classroom materials	Collecting classroom artifacts, such as assignments and resulting student work, can provide evidence of students' understanding of particular concepts that could lead to instructional design, and strategies to improve student-learning outcomes [34].
Student end-of-course evaluations	Designing and using the survey questions to collect specific information about the learning environment and the effectiveness of instructional practices.
Formative assessments (Also known as classroom assessment techniques)	This type of assessment consists of activities designed for helping educators assess and measure students understanding of the course content. The educator can reflect on the feedback gathered from their students. The two types of formative assessment are formal and informal techniques. Formal techniques include in-class activities, quizzes, and online learning modules. For example, if students do poorly on a quiz, then an educator might examine their instructional design or teaching strategies. In-class activities, where students work in pairs or small groups to solve classroom problems, can help students who are struggling. This approach can help educators to revise concept delivery or to reinforce important concepts in a different manner to help students. Informal techniques include written reflections, poll/surveys, and wrappers. For example, some of our participants used Poll Everywhere (an audience response system application that uses mobile technology to assess students understanding of particular concepts or misconceptions) to check what is working with students and what they find confusing with the lecture. Another participant used clicker questions to get instant feedback with questions to test students' understanding.
Summative assessments	Using these techniques could help educators improve their course, leading to improved student learning. These include collecting exams, projects, and presentations, and teaching portfolios. For instance, teaching portfolios provide examples of student works or evidence of participation, and these could be submitted for promotion purposes.

the lack of rubric use and training that might help observers focus on the connection between teaching strategies and student learning is problematic. In addition, classroom materials may be requested and collected from an educator when they undergo an observation, but as in the case of classroom observations using these documents to evaluate or improve teaching requires attention to opportunities for student learning. However, the choice of methods used ultimately depends heavily on the purpose of teaching evaluation, which may vary in terms of department or institution objectives.

Another important finding was that the methods used to evaluate teaching were either used for formative or summative evaluation. All but one of the teaching evaluation methods discussed in this study were used for both formative and summative purposes. For example, there was an almost equal split between formative and summative use for student-end-of course evaluation while the evaluation of classroom materials was mostly used for summative assessment. When the purpose is to assess the quality of student learning or for the faculty member to use data to improve teaching, then the methods were used formatively (e.g. student mid-term evaluations). In cases where the evaluation was ultimately used in the institution's rewards system such as promotion and tenure or continued employment for non-tenure line faculty, then the method was used summatively.

This work answers the call for research focused on teaching related issues in engineering education. As the discipline continues to grow and more formalized programs are established, it is necessary to link teaching strategies reported as effective to enhance student learning to methods of assessing quality teaching in engineering. In this study, we sought to explore the range of practices used to conduct teaching evaluation in engineering departments across the country. However, the number of participants were not evenly distributed across the three types of higher education institutions. This remains an area in which further investigation could be conducted. This would facilitate comparison of practices within and across institutions with varying value and focus on quality teaching.

Acknowledgments—This material is based upon work supported by the National Science Foundation under Grant No. DUE-1347817. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. National Research Council, *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics*, National Academy of Sciences, Washington D.C., 2003, pp. 1–127.
2. P. Morgan, The course improvement flowchart: A description of a tool and process for the evaluation of university

- teaching, *Journal of University Teaching Learning and Practice*, **5**(2), 2008, pp. 1–14.
3. P. L. Dressel, Teaching, learning, and evaluation, *Improving College and University Teaching*, **8**(1), 2015, pp. 11–15.
 4. J. Campbell and W. Bozeman, The value of student ratings: Perceptions of students, teachers, and administrators, *Community College Journal of Research and Practice*, **32**(1), 2007, pp. 13–24.
 5. G. Mohanty, J. Gretes, C. Flowers, B. Algozzine and F. Spooner, Multi-method evaluation of instruction in engineering classes, *Journal of Personnel Evaluation in Education*, **18**(2), 2005, pp. 139–151.
 6. H. C. Hill, D. L. Ball and S. G. Schilling, Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students, *Journal for Research in Mathematics Education*, **39**(4), 2008, pp. 372–400.
 7. C. Nygaard and D. Z. Bellugi, A proposed methodology for contextualised evaluation in higher education, *Assessment and Evaluation in Higher Education*, **36**(6), 2011, pp. 657–671.
 8. L. D. Fink, S. Ambrose and D. Wheeler, Becoming a professional engineering educator: A new role for a new era, *Journal of Engineering Education*, **94**(1), 2004, pp. 185–194.
 9. J. C. Ory, Teaching evaluation: Past, present and future, *New Directions for Teaching and Learning*, **83**, 2000, pp. 13–18.
 10. S. L. Eddy, M. Converse and M. P. Wenderoth, PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes, *CBE—Life Science Education*, **14**(3), 2015, pp. 1–16.
 11. T. C. Reeves, How do you know they are learning? The importance of alignment in higher education, *International Journal of Learning Technology*, **2**(4), 2006, pp. 294–309.
 12. M. L. Goldschmid, The evaluation and improvement of teaching in higher education, *Department of Psychology and Higher Education, Swiss Federal Institute of Technology*, **7**, 1978, pp. 22–245.
 13. T. D. Johnson and K. E. Ryan, A comprehensive approach to the evaluation of college teaching, *New Directions for Teaching and Learning*, **83**, 2000, pp. 109–123.
 14. Y. Chen and L. B. Hoshower, Student evaluation of teaching effectiveness: An assessment of student perception and motivation, *Assessment and Evaluation in Higher Education*, **28**(1), 2003, pp. 71–88.
 15. L. A. Braskamp, D. C. Brandenburg, E. Kohen, J. C. Ory and P. W. Mayberry, *Evaluating Teaching Effectiveness: A Practical Guide*, Sage Publications, Beverly Hills, CA, 1984.
 16. I. Iqbal, Academics' resistance to summative peer review of teaching: questionable rewards and the importance of student evaluations, *Teaching in Higher Education*, **18**(5), 2013, pp. 1–13.
 17. R. M. Felder and R. Brent, How to evaluate teaching, *Chemical Engineering Education*, **38**(3), 2004, pp. 200–202.
 18. P. A. Koushki, A. Christoforou, A. M. Larkin and A. Al-Roomi, Engineering students' attitudes towards teaching and teachers, *International Journal of Engineering Education*, **21**(3), 2005, pp. 424–433.
 19. P. C. Wankat and F. S. Oreovicz, Teaching engineering, *IEEE Engineering Management Review*, **37**(1), 2009, pp. 1–9.
 20. C. Henderson, C. Turpen, M. Dancy and T. Chapman, Assessment of teaching effectiveness: Lack of alignment between instructors, institutions, and research recommendations, *Physical Review Special Topics—Physics Education Research*, **10**(1), 2014, pp. 1–20.
 21. R. A. Berk, Survey of 12 strategies to measure teaching effectiveness, *International Journal of Teaching and Learning in Higher Education*, **17**(1), 2005, pp. 48–62.
 22. J. E. Miller and P. Seldin, Changing practices in faculty evaluation, *Academe*, **100**(3), 2014, pp. 35–38.
 23. J. Ohanlon and L. Mortensen, Making teacher evaluation work, *Journal of Higher Education*, **51**(6), 1980, pp. 664–672.
 24. R. Ferguson, Constructivism and social constructivism, in G. M. Bodner and M. Orgill (eds.), *Theoretical Frameworks for Research in Chemistry/Science Education*, Prentice Hall, New Jersey, 2007, pp. 28–49.
 25. M. Crotty, *The Foundations of Social Research: Meaning and Perspective in the Research Process*, SAGE Publications Inc., Thousand Oaks, CA, 1998.
 26. P. Baxter and S. Jack, Qualitative case studies methodology: Study design and implementation for novice researchers, *The Qualitative Report*, **13**(4), 2008, pp. 544–559.
 27. J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 4th edn, SAGE Publications, Thousand Oaks, California, 2014, pp. 43–46.
 28. Dedoose Version 7.1.3, <http://www.Dedoose.com>, Accessed 15 April 2016.
 29. C. Knapper, Broadening our approach to teaching evaluation, *New Directions in Teaching and Learning*, **88**, 2001, pp. 3–9.
 30. W. J. McKeachie, *McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers*, 11th edn, Houghton Mifflin Co., Boston, 2002.
 31. M. Yon, C. Burnap and G. Kohut, Evidence of effective teaching: Perceptions of peer reviewers, *College Teaching*, **50**(3), 2016, pp. 104–110.
 32. M. Prince, Does active learning work? A review of the research, *Journal of Engineering Education*, **93**(3) 2004, pp. 223–232.
 33. J. Turns, R. Adams, A. Linse, J. Martin and C. J. Atman, Bridging from research to teaching in undergraduate engineering design education, *International Journal of Engineering Education*, **20**(3) 2004, pp. 379–390.
 34. M. T. H. Chi and R. Wylie, The ICAP framework: Linking cognitive engagement to active learning outcomes, *Educational Psychology*, **49**(4), 2014, pp. 219–243.
 35. J. Biggs, Enhancing teaching through constructive alignment, *Higher Education*, **32**(3) 1996, pp. 347–364.
 36. G. Wiggins and J. McTighe, What is backward design? In *Understanding by Design*, 2nd edn, Association for Supervision and Curriculum Development, Alexandria, VA, 2005 pp. 7–19.
 37. C. Juwah, D. Macfarlane-Dick, B. Matthew, D. Nicol, D. Ross and B. Smith, Enhancing student learning through effective formative feedback, *Higher Education Academy*, **68**, 2004, pp. 1–41.
 38. V. J. Shute, Focus on formative feedback, *Review of Educational Research*, **78**(1), 2008, pp. 153–186.
 39. A. Goe, L. Biggers and K. Croft, *Linking Teacher Evaluation to Professional Development: Focusing on Improving Teaching and Learning*, National Comprehensive Center for Teacher Quality, Washington DC, 2012.

Keisha A. Villanueva is a graduate student in the School of Civil and Construction Engineering at Oregon State University. She earned her B.S. in Civil Engineering from Oregon State University in Corvallis, Oregon.

Shane A. Brown is an associate professor in the School of Civil and Construction Engineering at Oregon State University. His research interests include studies of engineering practice within theories of situated cognition and studies of students and engineers understanding of engineering concepts and conceptual change. He is a recipient of the NSF CAREER Award and numerous teaching, research, and service awards. Dr. Brown earned his B.S. and Ph.D. in Civil Engineering from Oregon State University.

Nicole P. Pitterson is a postdoctoral scholar in the School of Civil and Construction Engineering at Oregon State University. Her research interests include eliciting conceptual understanding of complex circuit concepts using active learning strategies. Nicole holds a Ph.D. in Engineering Education from Purdue University and other background degrees in Manufacturing Engineering (Western Illinois University) and Electrical and Electronic Engineering (University of Technology, Jamaica).

David S. Hurwitz is an associate professor in the School of Civil and Construction Engineering at Oregon State University. He also serves as the associate director at OSU for the Pacific Northwest Transportation Consortium (PacTrans) and as the director of the Driving and Bicycling Research Laboratory. His research interests include transportation human factors and safety and engineering education with a particular focus on conceptual understanding, problem solving, and the adoption of evidence-based instructional practices. Dr. Hurwitz earned his BSCE, MSCE and Ph.D. in Civil Engineering from the University of Massachusetts Amherst.

Ann Sitomer is a postdoctoral scholar in the School of Chemical, Biological and Environmental Engineering at Oregon State University. Ann's professional background is in community college mathematics teaching; her research background is in mathematics education. Her research interests include postsecondary teaching development and organizational change. Ann holds a Ph.D. in Mathematics Education from Portland State University, a M.A. in Mathematics from Arizona State University, a B.A. in Mathematics from the University of Southern Maine, and a B.A. in Liberal Arts from St. John's College.