International Journal of Engineering Education Vol. 37, No. 3, pp. 608–623, 2021
Printed in Great Britain

0949-149X/91 $3.00+0.00

Exploring Engineering Faculty Beliefs and Practices on Student Evaluation and Pedagogy*

AHMED OSMAN[1], ERIC CUELLAR[1], SHANE A. BROWN[2] and BENJAMIN D. LUTZ[1]
[1] Department of Mechanical Engineering, California Polytechnic State University, San Luis Obispo, California, 93407, USA.
[2] School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, 97331, USA. E-mail: blutz@calpoly.edu

Evaluation of teaching is an integral component of engineering education that is designed to improve student learning and faculty practices. But the data and methods used to conduct evaluation can affect the way faculty make pedagogical choices. The purpose of this research is to explore the ways in which institutional evaluation approaches influence the perceptions and actions of engineering faculty and in particular, the way faculty use these data to make changes or improvements to their teaching. We ask the following question: How do engineering faculty beliefs and practices about pedagogy influence their perceptions of institutional evaluation methods? We conducted interviews with 29 engineering faculty at a high-research activity university in the Pacific Northwest. Using thematic analysis, we examined beliefs and practices related to pedagogy and evaluation. Findings are presented in terms of four major themes: (1) pedagogy beliefs, (2) pedagogy practices, (3) evaluation beliefs, and (4) evaluation practices. Faculty beliefs about teaching inform a range of responses related to their concrete practices in the classroom. At the same time, however, faculty were unsure as to whether institutional evaluation practices offered meaningful insight regarding their effectiveness as educators. Findings suggest that while faculty recognize the importance and value of different evaluation methods and recognize the role of evaluation in improving their teaching, they also identify a range of challenges associated with existing measures common to most universities. Given the range of purposes that evaluation data is designed to serve, and the fact that the same data might be used to offer both formative and summative feedback, the authors recommend aligning evaluation practices and instruments with evidence-based instructional strategies which are responsive to the contexts in which they are deployed.

Keywords: qualitative research; student evaluation of teaching; engineering pedagogy; interviews

1. Introduction

Student Evaluation of Teaching (SET) is a vital component in engineering education that improves both faculty practices and student learning. These evaluation data are vital because they support a range of decisions regarding personnel and educational decision making. However, evaluations are notoriously difficult to interpret and are further complicated by differing notions of "effectiveness." Nonetheless, beliefs about the data and methods used to conduct evaluation affect how faculty make pedagogical choices and improvements. It is therefore important that evaluation data not only offer accurate reflections of faculty effectiveness but also provide meaningful information regarding how they might improve that effectiveness. While evaluations are intended to improve teaching and learning, there is limited research in engineering surrounding how faculty make sense of and respond to evaluation respond data. The purpose of this research is to explore faculty perceptions surrounding evaluations. In our study, we examine how evaluation approaches influence perceptions and actions of faculty as well as the ways faculty use evaluation data to guide their instructional improvement. To address this issue, we pose the following question: *How do engineering faculty beliefs and practices about pedagogy influence their perceptions of institutional evaluation methods?*

We interviewed 29 engineering faculty about evaluation and teaching at a research-based university in the Pacific Northwest. Using qualitative analysis techniques, we examined beliefs and practices related to both pedagogy and evaluation. By implementing a thematic analysis approach [1], we organized data according to dominant themes and explored the relationships among beliefs and practices surrounding pedagogy and evaluations. Our findings highlight the ways SET data both influence faculty pedagogical beliefs and inform choices about how to improve their teaching and learning. Moreover, our results stress the need for greater understanding of how SET data might support pedagogical developments on the one hand and inform institutional decision making on the other. The present research has implications for how faculty might make use of and interact with evaluation data; the role of students in providing meaningful assessments of effectiveness; and the role of administration in using such data to inform personnel decisions.

The following sections will offer a brief literature review that addresse relevant issues in SET mea-

* Accepted 22 November 2020.

surement and faculty perceptions of current evaluation practices. Next, we describe our methods, including data collection and analysis. Based on our thematic analysis, we present our findings in terms of four salient themes along with supporting sub-codes. Finally, we discuss the ways our findings both align with and extend current literature on faculty evaluations and make recommendations for various stakeholders in the SET process.

2. Literature Review

Current evaluation practices typically offer some combination of student and peer perspectives, and faculty have different perceptions regarding the validity and appropriateness of the different ways the data are used. Moreover, student evaluations in particular are subject to bias from a range of different axes of privilege and highlight some of the challenges associated with current measures and practices. The following sections will discuss these themes in more detail.

2.1 Student Evaluation of Teaching (SET)

Student evaluations of teaching (SET) are the most widely utilized method to measure faculty teaching effectiveness. Researchers at Oregon State University conducted interviews with engineering faculty at a range of university types and classifications to explore existing practices [2]. Their findings show that faculty interact with a range of data, but most evaluations take place via student evaluations. While SET varies according to organizational contexts and needs, they share many of the same characteristics across universities. Most student evaluations are administered at the end of the academic term before final examinations [3]. Students typically respond to some combination of quantitative, Likert-type items and open-ended, qualitative questions regarding the instructor's effectiveness. Evaluations are typically conducted through an online format and ask questions such as, "Give an overall rating of the instructor and the course," as well as more open-ended questions like, "What were the best aspects of this course?" [4]. This data is provided to instructors so they can respond to feedback and identify areas for improvement [5].

Some instructors use mid-course evaluations to make adjustments within the academic term. When student evaluations are given in the middle of the term, participation is more likely to be voluntary, informal, and used to provide formative feedback about the course and facility teaching effectiveness [6]. Instructors either mimic the typical approach of SET or have their own process for collecting data. For example, instructors might use structured evaluations, or more open-dialogue in the classroom to gather feedback about the course and their teaching (e.g., [6]). These evaluations are typically conducted because instructors want more student data to utilize which helps them make real-time changes based on the feedback collected [7].

2.2 Faculty Perceptions Regarding Evaluation Data

Nevertheless, despite their ubiquity in academia, student evaluations have remained a highly debated topic for over a century [8–10]. As a result, a number of studies have examined faculty perceptions on the ways evaluations are implemented and used. Ostensibly, evaluations are conducted by students to measure the instructor's effectiveness. But the term "effectiveness" can be problematic because its definition is rarely agreed upon between students and faculty. This disconnect raises concerns over what is actually being measured. For example, Layne (2012) found that faculty definitions of effectiveness were centered on content delivery and expertise, while student definitions tended to emphasize the instructor's ability to be relatable and show emotions [11]. Such findings demonstrate how perceptions of effectiveness may differ between those conducting and responding to SET. The inherent ambiguity of the term "effectiveness" poses challenges for interpreting and making changes based on evaluation data.

Further complicating this process is the fact that results are also used by university administration to make decisions informing promotion and tenure. In one study, Nasser and Fresko (2002) explored instructors' beliefs on SET validity and the administration's access to and use of these scores. Their research found more than one-third of instructors are opposed to administrative use of SET and more than one-half are opposed to students receiving these evaluation results [12]. Further, in light of concerns related to validity and meaningfulness of SET data, they found that faculty expressed reservations over administrative use of SET to make decisions about promotion and tenure. Such findings are important because they highlight potential disconnect between receiving meaningful feedback from students on the one hand and collecting data that impact faculty careers on the other.

Because effectiveness is not well-defined, researchers have called for the need to better understand the context in which these evaluations arise. For example, Osborne (1998) proposed combining multiple sources of data such as peer assessment alongside SET, noting that such processes are more flexible to diverse teaching methods and present a more holistic and coherent picture of faculty [13]. Still, faculty continue to express concern regarding

SET practices. For example, Lutz, Barlow, Brown, & Sanchez (2018) asked faculty to identify what is missing from current measures and offer suggestions to address existing limitations. Their findings suggest that instructors perceive a need for richer student perceptions and data, as well as third party input to help facilitate more longitudinal data and measure students' learning over time [14].

2.3 Challenges with Measurement

Despite SET being used for critical operations like faculty improvement, promotion, and tenure, research has demonstrated the presence of a range of biases that can impact scores. In response to these issues, scholars have explored a range of factors that might inequitably influence SET data. Such research is vital because while some factors might be under faculty control, others (e.g., race and gender) are beyond such control but nonetheless influence SET.

Some researchers have examined the effects of the educational setting on student evaluation results. For example, Hill and Epps (2010) explored the role of physical environmental factors (e.g., time of day, lighting, location of the classroom) on student experiences, and evaluations of instructor effectiveness. They found that students gave higher scores to classes taught in upgraded classrooms with tiered seating, better lighting, and noise control [15]. Studies have also explored the role of student grades and evaluation results [16]. Millea and Grime (2002) found that current grades and evaluation scores were positively correlated, whereas negative views towards future grades had a negative impact [17]. In contrast, however, a meta-analysis conducted by Uttl, White, and Gonzalez (2017) aggregated SET scores and final examination grades and found no correlation between the two [18]. Nonetheless, the relationship between grades and perceptions of effectiveness remains an important link to examine because, at least in theory, the grade a student earns in a course should operate somewhat independently of the effectiveness of a given instructor.

While factors such as grades and learning environment can be concretely understood and measured to some degree, personalities and preferences are both less controllable and potentially more impactful. For example, Clayson and Sheffet (2006) found that students' first impressions of faculty were significantly related to their end of course evaluations [19]. Similarly, Shelvin, Banyard, Davies, and Griffiths (2000) mapped the influence of specific personality types to evaluation scores and focused on the role of instructor charisma. Their findings suggest greater levels of charisma are associated with higher scores on measures of a lecturer's ability to convey material [10]. Such research suggests that student evaluations heavily rely on their perception of the instructor, even if those perceptions are not directly linked to concrete teaching approaches.

In addition to personality traits, researchers have investigated the role of race and gender in the evaluation process [20]. For instance, Hendrix (1998) interviewed 28 students across six sections of the same course at a school with predominantly white student enrollment and found that they tend to evaluate faculty of color using more rigorous criteria than their white counterparts [21]. Further, MacNell, Driscoll, & Hunt (2015) examined bias in terms of gender. Specifically, they administered identical online courses to participants but used different names to mark the genders for instructors. Their results showed the men received higher scores than the women despite the instructor and course delivery remaining identical. These findings are especially disconcerting in engineering contexts in particular, where the overrepresentation of white males exerts palpable influence on engineering culture and thus student expectations of engineering professors [22, 23]. Recent work has reinforced these findings and offer a more fine-grained analysis of potential inequities within SET processes. For example, Mitchell and Martin (2018) used mixed-methods approaches to explore differences between genders for both quantitative scores and language used in qualitative comments; they found that male instructors received higher scores, even when the questions were unrelated to the instructor (e.g., the use of technology or course materials) [24]. Women instructors also received more comments about their personality and their ability to keep students entertained. Their findings suggest that women and other minoritized groups in academia are at an inherent disadvantage when it comes to SET and underscore the need to better understand how such factors influence student beliefs about teaching effectiveness.

3. Methods

The purpose of this research is to explore the different ways engineering faculty interact with the different forms of evaluation data at their university and how that data influences their pedagogical beliefs and practices. To address this purpose, we interviewed 29 engineering faculty members regarding their beliefs and practices surrounding teaching and evaluation at their university. We used thematic analysis [1] to develop a list of dominant themes that describe faculty perceptions of the various modes and uses for evaluation data.

3.1 Recruitment and Sample

To recruit participants, we contacted administrators and school heads, described the study, and asked them to distribute a recruitment email to their respective faculty members. The recruitment email described the goal of the research and directed them to a screening survey link which probed for factors such as faculty position (i.e., instructional or research faculty), rank, number of classes taught, and the number and grade level of students in their classes. Table 1 offers an overview of the demographic information provided by respondents to the screening survey and Table 2 provides more information about the number and nature of courses taught by respondents. While we developed the screening survey to achieve maximum variation, we ultimately conducted interviews with all who responded to the survey.

Tables 1 and 2 offer data for 29 participants, though not all the totals within categories will sum to 29. For most items, faculty were able to select more than one option. Some respondents noted acting in dual roles such as in both instructional and administrative capacities. Faculty also selected more than one option when responding to questions related to class size, typical grade level of their students, and ethnicity. Moreover, one faculty member declined to participate in the screening survey and instead emailed the researcher directly to schedule an interview; that participant is not represented in Tables 1 and 2.

3.2 Interview Protocol

The interview protocol followed a semi-structured approach to both allow for consistency across participants as well as provide space for emergent conversations or aspects of evaluation that participants found personally relevant. First, we asked participants to describe the ways in which their teaching was currently evaluated, focusing primarily on the methods used and the logistics of how the evaluations were conducted. Second, we explored faculty use of the evaluation data. We probed for faculty beliefs about what the evaluations tell them about their teaching, what they do with their results and how they decide to make changes to their teaching based on the evaluation data they receive. Third, the interview protocol shifted to investigate faculty beliefs and perceptions regarding "effective teaching." In particular, we asked them to define the term (i.e., effective teaching) and to describe what it looked like for them. Fourth, we asked about how they determine success for students and what students must do to achieve that success. Finally, we asked faculty to elaborate on the degree to which they believed the current evaluation measures offered accurate depictions of their effectiveness as educators. The interviews were audio-recorded and transcribed by a professional transcription service for further analysis. All research protocols were approved by the local institutional review board (IRB #8247).

3.3 Analysis

To examine emergent themes in engineering faculty perceptions of SET data and practices, we followed recommendations for thematic analysis procedures recommended by Braun and Clarke (2006). Specifically, we followed their step-by-step guide which lays out five phases that facilitate the development

Table 1. Demographic overview of respondents

Faculty Position		**Ethnicity**		**Gender**	
Instructional	3	Non-Hispanic White	21	Male	21
Research	25	Hispanic White	1	Female	7
Administrative	3	Latino or Hispanic	2		
		Asian/Indian Subcontinent	4		

Rank		**Experience**	
Assistant Professor	12	0–5 years	15
Associate Professor	8	6–10 years	5
Professor	7	11–15 years	3
		>15 years	5

Table 2. Overview of courses taught by respondents

Number of courses taught		**Typical Grade Level**		**Approximate Course Size**	
1	1	First-year	2	1–20 students	18
2	3	Sophomore	5	21–40 students	11
3	12	Junior	23	41–80 students	17
4	11	Senior	20	81–150 students	10
> 4	1	Graduate	25	> 151 students	4

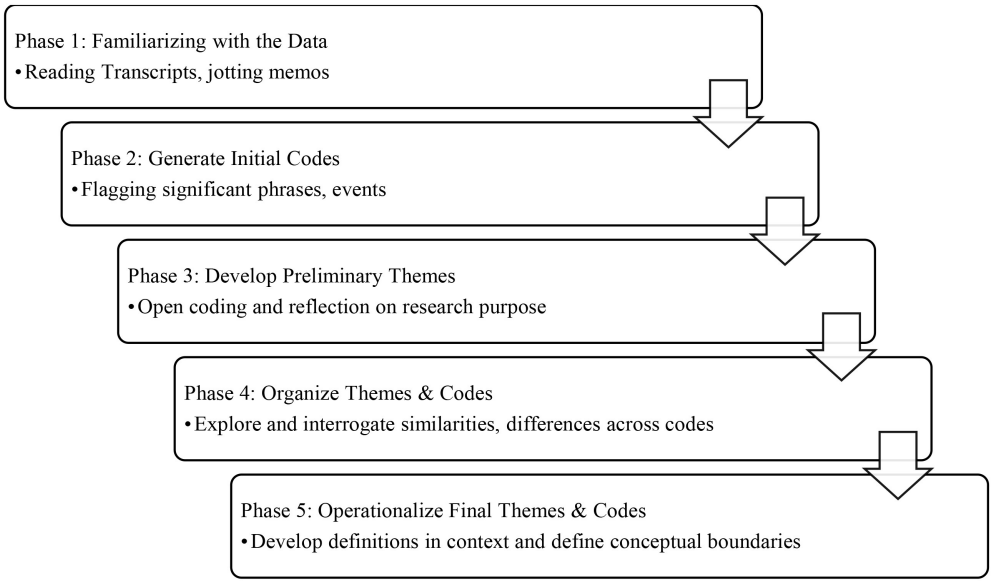

Fig. 1. Overview of Thematic Analysis recommendations by Braun & Clarke (2006).

of rich descriptive codes. We used these steps to move from qualitative interview transcripts to distilled themes and accompanying descriptive codes that support those themes. To do so, we focused on the beliefs and practices as they concerned teaching on the one hand and evaluation on the other. All analysis was conducted using Dedoose, a cloud-based qualitative analysis platform. Fig. 1 outlines the 5 major phases of thematic analysis we followed throughout our work and the following section describes each phase in more detail. Important to note here is that while the process is presented as proceeding in a linear fashion, many of the phases of this work were iterative and recursive in nature.

During phase 1, the authors familiarized themselves with the data; reviewing the transcripts in their entirety, jotting memos, and taking notes related to the ideas of pedagogy and evaluation. Led by the final author, the first two authors explored the data and developed initial ideas about the content of the interviews and the nature of the discussions within them. During phase 2, we generated initial codes. Given that our focus was on teaching and evaluation, passages were that referred to either of these ideas were "flagged" and organized for subsequent analysis.

These data were then sorted for phase 3, during which we generated initial themes. Guided by our interest in how *beliefs and practices* might align both within and across discussions of teaching and evaluation, we conducted open coding [25] that characterized the nature of the passage as it pertained to our broader research goals. For example, some initial themes we identified were those related to beliefs about what "effective teaching" or "effective learning" looked like to participants. We also identified initial themes related to faculty perceptions of what is being evaluated within these processes, both implicitly and explicitly. Identifying these initial themes was instrumental to informing a more coherent organizational scheme developed in phase 4.

Phase 4 consisted of reviewing these themes and organizing them according to similarities and differences. We thus grouped these codes to focus on two primary goals: internal consistency and conceptual boundaries. To address the former, we organized passages according to similarities in discussion and ensured that each passage within a given code addressed the same core concept. For the latter issue, we compared passages across codes to examine the possibility of conceptual overlap and make sure that each code addressed a distinct idea within the data set. This process facilitated the organization of codes into broader ideas that would form the supporting basis for our final themes.

Phase 5 entailed defining and naming our final themes. Because we were interested in beliefs and practices and the relationship between them, we organized our codebook according to four overarching themes, resulting in a 2x2 matrix that described faculty beliefs and practices about evaluation and pedagogy. The results of the thematic analysis are provided in Table 3, which offers an overview of both the final themes and the supporting codes.

3.4 Credibility and Trustworthiness

To establish credibility and trustworthiness of our findings, we implemented two primary approaches following recommendations from Rossman & Rallis (2003) and Miles, Huberman, & Saldaña

Table 3. Overview of final thematic codebook

	Pedagogy	Evaluation
Beliefs	• Effective teaching • Effective learning • Motives to improve	• Impact on career • Validity and reliability of data • Implicit evaluation criteria
Practices	• Content delivery • Assessing student learning	• Institutional feedback • Improvised feedback

(2013) as well as Patton, (2005). In particular, we focused on *peer debriefing* and *triangulation* through multiple analysts. During peer debriefing, we provided the codebook and results to researchers in engineering education who also have studied SET processes and incorporated their feedback. We asked the peers to identify any concerns or questions they had with our codebook and we worked to incorporate those concerns into subsequent iterations of analysis. A major result of this process was the suggestion to focus on a simpler 2x2 approach to presenting the codes such that readers could more clearly understand how codes were both different from one another but also relate to each other in their ability to represent the central phenomenon (i.e., engineering faculty perceptions of SET).

In addition, we used triangulation through independent researchers. To accomplish this triangulation, the final codebook was developed by the first two authors and provided to the last author for analysis. The final author then independently coded a subset of the data and examined the agreement between the way the codes had been applied across researchers. In the case of discrepancies between coders, disagreements were argued iteratively to consensus and the codebook was modified to account for noted differences. The result was a final list of four themes and supporting codes to depict faculty beliefs and practices regarding pedagogy and evaluation.

3.5 Limitations

There are several limitations that should guide the reader's interpretation of our findings presented below. First, as can be seen in Table 1, our data come from a relatively skewed sample of engineering faculty. That is, most participants were white and male, and most were either tenured or tenure-track professors (as opposed to instructional faculty). It is therefore important to recognize that these faculty perceptions might not represent the full breadth of perceptions expressed by a more diverse group of faculty or those who engage in more teaching than research. Nonetheless, it is important to understand these faculty perceptions because at least in the present moment, they are relatively representative of the demographics of many engineering departments at high-research activity institutions [28]. Thus, while future research should more deeply explore the beliefs of a broader range of SET stakeholders – especially minoritized groups in engineering – this work illuminates concerns and challenges that can potentially transfer to a broad swath of engineering faculty nationwide.

Another important limitation concerns the lack of actual evaluation data from participants. This is a limitation for two key reasons. First, while we did not ask directly about how individual participants performed on the SET, they did often note receiving relatively high scores. Therefore, faculty who participated in this research might have been willing to do so because they generally received higher SET scores than their peers. Second, and related, perceptions of SET might be based on faulty receiving particular kinds of scores. That is to say, faculty who receive relatively high scores might have a personal incentive to accept the validity of their evaluation, while those who receive less favorable evaluations might be more likely to discount the usefulness of the data or insights offered by them. Additional work should examine the relationship between actual student and faculty evaluation data and faculty perceptions of these processes in ways that can better triangulate findings.

4. Results

The purpose of this research is to better understand the ways SET approaches within a high-research activity university shape faculty beliefs and practices surrounding their own pedagogy. We present our results in terms of four emergent themes that focus on both beliefs and practices as they concern evaluation and pedagogy. Table 4 offers operational definitions for each of the themes identified, and the following sections describe each theme in more detail and offer examples through participant quotes.

4.1 Pedagogy Beliefs

We defined Pedagogy Beliefs in terms of three major codes. First, effective teaching concerns faculty beliefs about what it means to be a good teacher. On the other hand, effective learning describes faculty expectations surrounding student learning. Finally, motives to improve captures the

Table 4. Themes and operational definitions

Theme	Operational Definition
Pedagogy Beliefs	Beliefs about how to engage in meaningful teaching practices. Perspectives and opinions on teaching methods and student learning, as well as individual teaching motivations and approaches to improving teaching.
Pedagogy Practices	Descriptions of the methods and modes related to the ways faculty interact with and evaluate students in their classes. The ways in which faculty deliver their material and measure students' understanding.
Evaluation Beliefs	Faculty beliefs about the evaluation process, including how they feel they are evaluated and what they extract from institutional and improvised feedback, as well as the impact it has for promotion and tenure.
Evaluation Practices	Descriptions of the means by which evaluations are conducted by the professor or the university. Implementation of various institutional evaluation processes, as well as the way these practices might are utilized by the instructors and/or administration.

ways in which the previous two codes interact and offers descriptions of the reasons faculty modify their beliefs about teaching and learning.

4.1.1 Effective Teaching

We defined *effective teaching* as faculty beliefs on what "effective" teaching means and the role they play in their student's learning. In some cases, instructors believe that their role in being an effective teacher is to be organized and motivate students. As one respondent noted,

> "Effectiveness is embodied in what you're bringing in terms of either an organization of the material, or an approach to the material that motivates the student" (01).

To be effective in the classroom, faculty believe they must communicate effectively and clearly and that doing so is motivating for students.

Some faculty view their role as one of a coach and see themselves as resources to help students get to where they need to be.

> "I can really have an impact on them. I can coach them. I can say oh you know it's just like this or I see when you solve problems you always do it this way but if you did it this way you could keep track of things a little better. For those people I have a great impact. In this room and those tables I often have students sitting in here in my office hours coming in when they have a question. That's when I think I'm being very effective as a teacher" (05).

Rather than simply working example problems, this faculty member notes how careful observation of student preferences and problem-solving approaches can facilitate effective learning. Faculty beliefs such as those expressed here offer insight into the ideas that shape their classroom as well as their expectations of students.

4.1.2 Effective Learning

Effective learning captures the teacher's perception of their relationships with and the expectations they have for students. Some faculty noted the importance of simply showing up and engaging with the class. One participant noted how attending class and office hours are major factors in being a successful student.

> "Attend class. I want them to participate and make effort to do their homeworks, make effort to come meet with me during office hours. I tell them in advance that, 'Hey, you guys got to put in effort here in this class.' If someone is not being responsive, then they're not engaging in the learning process. They're disconnected, and so I tell them" (06).

Faculty also stressed the importance of being mentally present and "engaged". One participant noted a distinction between simply being present and being meaningfully, cognitively engaged. When asked about what makes an effective learner, they noted the following:

> "Some would say, 'show up'. . . Many of my students do well. I think they do have to show . . . you can pass my class being in your head, but you succeed in my class when you can show up with your whole self" (15).

The interesting point to note here is the distinction between passing the class and succeeding in the class. By showing up with their "whole self," a reference that seems to signal full engagement, students are more likely to succeed. Better understanding faculty expectations of students can help clarify and bring alignment to the classroom experience for both students and instructors.

4.1.3 Motives to Improve

At the intersection of effective teaching and effective learning sits faculty beliefs and motivations for improving their pedagogical approaches. We identified *motives to improve* to be instances in which faculty discuss their drive to make changes to their teaching styles and the reasons for doing so. One instructor discussed that negative comments help them become better instructors for the future.

> "I made a major change in the [redacted] course, and it improved the score from a five to a six. So it was a big jump. And I was like, 'Oh, okay! They responded.' I thought they would like it better for the class, and then this kind of reaffirmed it. And then as far as the little written comments, which you see, not very many. But you just kind of read through them, and you savor the

> nice ones. And then the ones that are a little more harsh, you're like, 'Okay. Point taken. I'm not gonna think too much about it for a few months.'" (20).

In the example above, the motive to improve and the evidence used to make those improvements center on their SET results. The "big jump" in their score was reflected in the quantitative results ("the score went from a five to a six"), while the written comments appear to be a source of triangulation for the noted improvements. However, not all faculty share this belief. Some faculty derive motivation from the desire to not raise any red flags or introduce complications into other evaluations. As one participant noted,

> "I do care so bad [about my evaluations], the number is so terrible that on average, that it becomes a red flag for [the] P&T process. Then I'm like, 'Oh, shoot! I need to do something creative here'" (06).

Some instructors use SET as a tool to ensure they are not underperforming. When evaluation numbers start falling, red flags may be raised by administration, which is a common motive to improve. This motive to improve stands in contrast to those that seem to be fueled by a sense of "care".

> "Some of those things, I think I did and some I don't. I tend to be more controlling and so forth. I feel like my teaching, oh yeah I am a better teacher than I used to be, but it's not fueled by feedback I got through evaluation processes.
>
> Ben: Okay. So then what is it fueled by?
>
> Respondent: I would say is fueled because I care so deeply about it. I'm a workaholic. That's why I'm retiring, I can't control it. I'm here constantly" (04).

Faculty offered a range of reasons for why they might want to improve based on their SET data. And understanding why and how faculty make choices to improve their teaching is important for developing high-quality faculty development and meaningful interventions.

4.2 Pedagogy Practices

Faculty described their teaching practices in terms of two overarching dimensions. *Content Delivery* describes how faculty make choices about the content they cover and the ways that they cover it. *Gauging Student Learning* thus describes the tools and assessments that faculty use to ensure that the content they covered is being retained by students in their classes.

4.2.1 Content Delivery

We defined *content delivery* to be different stylistic approaches to teaching a class and sharing knowledge and information. While all faculty agreed that they should teach in an effective manner, there exists a range as to how to best achieve that effectiveness. For example, to make the information relevant, some instructors stressed how they bring in real world applications.

> "I tried to be sort of like a real mix on content delivery and then application of that material, in close proximity to one another. So I thought that was a ... It was certainly striving to be more active learning, and more applications of what they're learning, as opposed to just me lecturing at them and then giving them a homework set" (01).

Some faculty emphasized the role of language and discussed the importance of word choice for student learning. Delivering content in clear phrases enables students to easily grasp concepts and theories.

> "I think about lectures a lot and I choose the words extremely carefully and I know that a good portion of that is always lost. I choose the phrase to exactly encompass the concept, the theory, the assumptions, the application and all they hear is this and that" (05).

Such practices demonstrate the importance of content delivery because they show the impact that different methods of delivery can have on student learning. By anchoring examples to practical student experiences and carefully selecting the language used in class, faculty in this study were able to maximize the effectiveness of the content they delivered.

4.2.2 Gauging Student Learning

Just as instructors have differing methods of delivering their content, their practices for gauging student learning and understanding vary as well. *Gauging Student Learning* is defined as the ways in which teachers measure student comprehension and performance in class. Although professors had some differences in how they assess student understanding, many used in-class assignments as assessments of learning.

> "I would just give the problems and just sit back, but then once I started walking around, I started noticing no one's doing anything and when I went and stood next to someone he was like, 'I don't even know where to start.' So those are indications of what's challenging" (00).

By having examinations and giving in-class assignments, professors can gauge the general learning and evaluate gaps that students might have (e.g., "where to start"). Further, teachers also make assumptions during class based on student reactions and mannerisms. For example, one participant noted, "it's body language feedback. So, you can pick up a lot of immediate feedback based on body language" (07). By knowing how students react in class and how they are performing, faculty can gauge student understanding.

In other instances, faculty administered pre-

examinations to gauge the general level of understanding for the entire class. For example, the following quote illustrates how some faculty conduct assessments of prior knowledge to obtain a baseline of student understanding.

> "I do pre-assessment for students at the beginning of the class. I assess them on the previous [subdiscipline] classes, because mine is the third in the series, they need to build on the other two. And then I assess basic freshman chemistry. It turns out that the freshman chemistry is what gets them" (14).

By giving benchmarks prior to class, faculty can assess where all of their students stand with respect to the subject and can more effectively address gaps in knowledge.

4.3 Evaluation Beliefs

Faculty in this study described a range of beliefs concerning SET usefulness, the impact on their careers, and the subtler, implicit criteria that might bias the scores they receive. *Validity and reliability* concerns beliefs about the general usefulness of the data and what faculty glean from it. These results are elaborated on by faculty beliefs about *implicit evaluation criteria* that might be inadvertently applied as students complete their evaluations. Finally, in light of these beliefs, faculty described the ways in which SET *impact their career* and influence personnel decisions at their university.

4.3.1 Validity & Reliability of SET Results

Validity and Reliability is defined in terms of faculty perceptions of the accuracy of SET. While faculty recognize that SET is measuring something, precisely what it is measuring is often unclear. For example, some instructors believe that SET does not have much if any validity in assessing their effectiveness, and instead provide some coarse measurement of overall satisfaction.

> "I think they [SET] provide an accurate description of how many students like me as a person. . . If the students like me as a person, they are giving me good grades, but I think the effectiveness comes more from the outcomes of how well they actually do in the content itself or how much participation did they have in the class or how much did they pay attention to the class? I think they are correlated, because if they like me, they do pay attention to a lot of things and they actually do make an effort" (06).

Here, the participant describes SET in terms whether or not students "like [them] as a person." At the same time, others value the written feedback section of SET results as it gives direct feedback to help faculty improve.

> "As I have kind of alluded to before, I think the comments are really nice, so we really should keep the comments. It's a nice way, students are not afraid of being judged by the professor since it's done afterwards. So the professor doesn't see this before the final grade's getting posted. [. . .] So I really believe the comment section needs to be there. It's really, really helpful, it's nice" (32).

This qualitative data enables faculty to see specific areas for improvement rooted in authentic student experiences. These instructors firmly believe that, "[students] give genuine feedback. So, I take those, and I try to improve my teaching" (17).

Some faculty view student comments as valid assessments of their effectiveness, while others believe it is more of a measure of whether (or to what extent) students like the professor. Such beliefs suggest that faculty are potentially split regarding the validity and/or reliability of existing SET measures, but most seem to confer higher value to written student comments.

4.3.2 Implicit Evaluation Criteria

Implicit Evaluation Criteria are defined as factors that influence SET scores beyond measures of effectiveness and that conflate broader social, cultural, political, logistical, etc. expectations of engineering faculty. For faculty in our study, SET can be affected by unconscious biases that can both positively and negatively influence results.

One implicit criterion results from the time of day or the room the class is held in. For example, one participant stated that, "8AM recitation section, you're gonna get a worse score. It's objectively true" (07). This belief demonstrates how some factors are beyond the control of any teacher.

Other faculty perceive SET results to be more about popularity or the degree to which faculty characteristics align with dominant cultural or social archetypes within engineering.

> "They are students . . . it's a popularity vote. There is . . . and you know this kind of thing, surely, that there is research that's shown that they are highly tied to expectations of students and those expectations are norm . . . around the dominant paradigms. So the more you are white, male, straight, you know, etc. English speaking" (15).

This quote illuminates the lack of diversity within engineering and demonstrates the belief that faculty who do not embody the "dominant paradigms" are at a natural disadvantage. More research in such areas is needed, but other scholars have demonstrated the negative impacts on SET results for women and other minoritized groups in STEM [29, 30]. Relatedly, some faculty noted a range of characteristics that, while unlikely to be related to effective teaching in any appreciable way, nonetheless influence the scores students will give. The following quote illustrates a number of perceived

implicit criteria upon which faculty are evaluated, such as charisma or sternness.

> "I think a lot of it has to do with just like charisma and how approachable you are, and stuff like that. I could see where it could be unfair to somebody who's a little more stern but maybe expects more of the students, and kind of that old school mode of you've gotta learn this stuff, you know" (16).

These implicit evaluation criteria are critical to note because they suggest that although student evaluation is a vital component of faculty evaluation, dominant cultural beliefs and related expectations might unduly influence the perceptions of students and skew evaluations in inequitable ways.

4.3.3 Impact on Career

While most participants were aware that SET results had a relatively strong influence on their careers (at least in terms of retention and promotion), they varied in their beliefs about exactly *how* the results influence them. *Impact on Career* was defined as faculty beliefs about institutional feedback methods and their relation to promotion and tenure.

Faculty in this study expressed beliefs that institutional feedback was typically used by administration to monitor faculty performance rather than reward the top instructors and promote further growth beyond some average score. For example, one participant noted that, "as long as you're average you're fine" (05). Relatedly, another participant stated, "they may use those things, mostly to possibly punish us, but not to really reward us for it" (33). Such beliefs demonstrate that some faculty do not believe SET has a major role in their career and that as long as faculty do not fall below average, nothing is necessarily wrong. (Interesting to note here is that, at least in this current research context, "average" was often somewhere around a quantitative score of 5 out of 6, which corresponded to a rating of "very good.")

In contrast, some faculty believed that SET results played a sizable role in personnel choices.

> "As far as I am aware, I know they're heavily used in the tenure process. And being that I'm going up in the tenure process now, I know that they're used as a yardstick. I do know that everyone admits they're probably not the best yardstick, but they seem to be still used. 'Cause they're not great, but they're there, so we'll use 'em" (20).

Noteworthy here is the suggestion that "they're [SET data] not great, but they're there, so we'll use 'em." Such beliefs are interesting because they suggest that the primary reason these data are used is simply because they exist, which raises questions about what other kinds of data could be made to exist and how such data could be used. Although this excerpt discusses how SET may not be the best gauge or indicator, it is still used by administration to make decisions on choices that may impact faculty careers.

4.4 Evaluation Practices

Evaluation Practices describes the different mechanisms employed by faculty and their institutions to capture SET data.

4.4.1 Institutional Feedback

Institutional feedback captures evaluation practices used by the institution and the ways faculty use and interact with this information. In interacting with this institutional data, most participants reported that they prioritize the quantitative data first and supplement with qualitative comments. The quantitative data is important because those numbers are often used and compared both within departments and across the university.

> "I normally take a look at the number first. They have two numbers that are, I would say, the most critical. It's the contribution of the course to their overall knowledge. And then the contribution of the instructor in that same context. So, I look at those two numbers just to get a gauge as to how they perceive that I performed. I don't read the comments right away, because I think it's too close. I kind of tend to get upset sometimes when I read them right away. So, I wait. Maybe sometimes until . . . so if I teach a class in Fall, I wait until the Summer to take a look at them" (21).

> "I go from year to year like whether I am . . . Kind of look at past performance. I mean I do look at how I'm doing relative to the college average or whatever those things that have to go in my dossier. And then I read the comments, and I try to make notes of things there. We do course summaries as part of our ABET process, and I try to keep a running diary through the term on things that I think are going well or not, so I can have some information to make changes for the next year. So I usually try to at least summarize that stuff when I finish the term" (01).

The numbers are perceived as "most critical" because they offer a quantifiable metric across which different faculty can be compared. They are also the data that go into faculty dossiers for promotion and tenure. Student comments, on the other hand, are examined differently and tend to be used more qualitatively to help faculty make decisions about what they might change in future iterations of the course. These quotes highlight the important differences between summative and formative feedback and point to a potential disconnect between the two. That is, faculty in this study were keenly aware of the importance of their numerical scores, but at the same time seemed to use students' qualitative comments to inform pedagogical decisions.

4.4.2 Improvised Feedback

Improvised Feedback was used when faculty solicited feedback beyond the traditional institutional practices and most often took the form of a mid-course evaluation or some form of informal check-in. Improvised feedback was noted as beneficial because it offered more specific, localized, insight into faculty effectiveness and offered opportunities to make changes throughout an academic term. The kinds of improvised feedback varied across participants, but most were used to gauge students' level of knowledge and satisfaction. As one instructor specified, "I have an archive of all the feedback that I've gotten per class basis. I try to detect trends" (11). When using improvised feedback mechanisms, faculty tend to employ a similar instrument as the SET process and add in some of their own questions.

> "I pretty much mimic the [SET] and then I ask, sometimes I would throw in very specific questions about course content. If I was curious about that. Like we covered, if we covered something new that I didn't in the previous year. I would say, 'How helpful is this?'" (07).

Here, asking questions specific to the course allow faculty to more effectively assess student understanding and development. Collecting data during the course rather than the end helps teachers implement additional feedback because they can pivot and alter their lesson plans.

Another improvised method of data collection was a loose, informal discussion with students. One instructor noted the use of a kind of continuous feedback gathered daily through questions and conversations during class.

> "I wanted students to feel like they at least had a forum to provide me feedback. And sometimes I ask . . . every day I start class, 'Anyone have any questions?', and it's usually like no one. Usually one or two people ask, but it's like another opportunity for them to feel like they can give feedback and that someone's going to listen to what they say. I just felt like that's the sort of relationship I want to have with my students, so it seemed like a pretty easy way to try to forge that sort of relationship" (01).

In this particular excerpt, faculty administer improvised feedback by simply having an open forum with their class so that questions can be answered when they arise. Improvised feedback enables faculty to have more access to student data and know how to make quicker adjustments in the classroom.

4.5 Summary of Results

Overall, faculty in this study noted a range of beliefs and practices as they pertain to pedagogy and evaluation. In many cases, the relationship between teaching and evaluation resembles that of a cycle in which evaluation data is collected to inform pedagogical beliefs and thus classroom choices, and the data collected from the next round of evaluations informs subsequent changes and beliefs. Perhaps reassuringly, participants here noted a deep sense of care for how they perform in class and encourage learning for their students. And in order to nurture this sense of care, it is important that faculty have the appropriate data to meaningfully inform their pedagogy.

5. Discussion & Implications

Faculty believe there is value in conducting evaluation of teaching and measuring effectiveness, but also perceive room for improvement in current evaluation practices. In particular, our findings point to three key implications for improving faculty evaluation approaches. First, given the confusion surrounding the term "effectiveness" and the resulting inequities, it is important that evaluation methods devise ways to align with evidence- and research-based practices. Second, while faculty recognize the importance of their own behaviors in the classroom, they also noted the role of students in engaging in effective learning. Evaluations should therefore take into account student learning behaviors and acknowledge the reciprocal nature of teaching and learning. Finally, given the confusion many faculty noted regarding the use of evaluation data, it is vital that all stakeholders involved in the evaluation process offer greater clarity and transparency in handling the data. We argue for greater transparency concerning the ways in which the myriad SET data is used across the diverse stakeholders involved in the process (i.e., students, faculty, administrators). Our key findings are summarized in Table 5 according to the different stakeholders in the evaluation process and are elaborated on in the following sections.

5.1 Research-Based Instructional Strategies and Effective Teaching

Faculty in this study expressed a range of beliefs about what it means to be effective. This was true both for their own definitions and those they believe students hold about the same concept. When faculty and students share a mutual and equitable meaning of effectiveness, evaluations can be clearer and better aligned across parties. Better alignment will minimize gaps within evaluation practices by creating a more coherent understanding of what is being measured. The importance of minimizing these evaluation gaps is reflected in our findings and problematized by the literature on biases and their role in evaluation [11, 30, 31]. As Osborne

Table 5. Summary of findings aligned with recommendations for improving the use of SET in engineering

Stakeholders	Challenges	Recommendations
Faculty	Conflict between critical and positive feedback	Different data for formative and summative purposes
	Conflating factors influence SET scores (e.g., race, gender)	Redesign SET to mitigate implicit biases and non-teaching factors (e.g., class time, room assignment)
Students	Lack of input during academic term, when it matters most	Increased opportunities for feedback during the term
	Limited data on how students engage with a course	Probe for concrete learning behaviors in addition to teaching behaviors
Administrators	Confusion about use of SET data in merit decisions	Greater transparency in data use
	Lack of consensus on what constitutes "effective teaching"	Probe for use or prevalence of Research-Based Instructional Strategies

(1998) observed, ensuring that students and faculty work together in the evaluation process may help present more cogent data because effectiveness will be evaluated based on a shared understanding and definition.

To achieve greater alignment and shared understanding of effectiveness, we recommend that SET processes probe for objective behaviors and practices that are well-known to be effective. For example, researchers have long noted the value of various student-centered or active learning strategies [32, 33]. Understood broadly as *Research-based Instructional Strategies* (RBIS), these approaches and their value have been documented by numerous researchers (e.g., [34]) and their use offers insight into how SET might be conducted in a more objective, equitable fashion. To mitigate the ambiguity surrounding effectiveness, then, universities should apply this knowledge when developing SET and create items that probe for the use of RBIS. For example, SET might include items that ask students to report the frequency with which faculty implement well-known effective teaching techniques, such as how often faculty engage student in inquiry-based learning [35] or facilitate think-pair-share activities [36]. Moreover, probing for the use of RBIS in SET can offer guidance for the instructor in terms of what effective teaching looks like and thus encourage innovative teaching techniques in ways that enhance student learning. By asking about specific teaching practices linked to the use of RBIS, SET can generate data that speaks directly to evidence-based notions of effective teaching and avoids the ambiguity and problematic biases inherent in asking about effectiveness in general. Such changes can lead to more equitable evaluation of engineering faculty.

5.2 Effective Learning and Teaching

Most existing measures of faculty evaluation emphasize the behaviors and actions of the faculty member. This emphasis makes sense because understanding how faculty create spaces for effective learning is important for understanding how they might replicate or improve those environments. Indeed, faculty in this study pointed out that effective teaching has to do with delivery and their knowledge of their respective subjects and content. But focusing only on what the teacher does overlooks students' roles in co-creating an effective learning environment. That is, faculty noted that while they can perform certain effective behaviors, students are still responsible for engaging in effective learning practices too. For example, prior research has shown the importance of self-regulation in effective learning behaviors in engineering contexts [37, 38]. Understanding what kinds of learning students engage in is perhaps just as important for understanding the effectiveness of an instructor because it highlights the kinds of environments and behaviors they promote for their students.

Much of the SET process is centered on the things faculty do. But just as communication involves both talking and listening, effective teaching also requires effective learning. Our findings show that faculty recognize both the role of implementing effective practices as well as the ways in which students might meaningfully engage in a learning environment. Thus, while engagement is related to the instructor's effective teaching, student behaviors play a vital role in effective learning. We therefore recommend that SET practices incorporate the broader perspectives associated with both faculty and student behaviors. SET might leverage data from existing instruments that probe student engagement in class (e.g., [39]). Rather than asking questions only about teaching, SET might also ask about student learning. For example, student-centered questions might ask about attendance, how many hours students dedicated to the course, amount of time spent in office hours, etc. In line with recommendations by Lutz et al. (2018) and Osborne (1998), such questions will enable instructors to better understand the context surrounding their evaluations and offer more concrete support for decisions to make pedagogical changes.

Relatedly, in order to collect more student data

and help transform faculty teaching, instructors also noted the use of mid-course evaluations. Since SET are typically conducted during the end of the course, faculty expressed concern that students lack incentive to perform the evaluation. In order to compensate, faculty often supplemented with improvised feedback mechanisms. In doing so, they position current students as stakeholders in instructional improvements rather than individuals making recommendations for future cohorts of students. Mid-course evaluations give faculty the opportunity to respond to the individual needs of a given class in real-time and therefore offer a greater incentive for students to participate. Moreover, when combined with end-of-course SET scores, faculty can better understand the impact of their instructional changes over time. In agreement with several participants, we believe the use of mid-course feedback is useful and should be incorporated into faculties' formative feedback processes.

5.3 Data Use and Transparency Among Stakeholders

Lastly, faculty in this study recognized the importance of student evaluation and the affordances of both qualitative and quantitative data for improving pedagogical approaches, but they also had varying beliefs on how it may be used by different stakeholders. Consistent with existing literature [12], participants were often concerned with how administrators may use SET data for personnel decisions, such as promotion and tenure. A potential issue with this process is that in order to recognize and appreciate student insight, faculty should be comfortable with receiving negative reviews. When faculty are uncertain how administrators use evaluation data, it may lead them to wanting only positive comments. As noted by a participant in this study, faculty "savor the nice [comments]." But positive comments might also be less informative to guide pedagogical improvements. Thus while, negative comments or critical evaluations can help drive pedagogical improvement, that same data could be used against faculty during promotion and tenure reviews. Such implications are especially important in light of literature surrounding student resistance to active or student-centered learning [40]. For example, faculty may avoid the use of RBIS to avoid negative evaluations, despite being these practices being well-established means of facilitating effective teaching.

Faculty in this study also noted substantial confusion as to how their evaluation data were used by the different relevant stakeholders – in particular by university administrators. While they understood that SET data contributed to institutional decisions around promotion and tenure, the ways in which it contributed were often obscure. The disconnect in how evaluation data is used by stakeholders may cause tensions between what is helpful for improving teaching on the one hand, and what might be detrimental for faculty careers on the other. To address these concerns, deans and other administrators should consider clarifying with faculty how their evaluation data are used to make personnel decisions and the heuristics used to reach those conclusions. Greater transparency may help foster the use of RBIS in teaching while curbing behaviors used solely to boost evaluation scores (e.g., inflating grades, trying to be more charismatic). Administrators may want to consider studies like that by Nasser & Fresko (2002) and survey their faculty on how they believe evaluation data should be used to better guide pedagogical improvements and support student learning.

5.4 Directions for Future Research

While this research has illuminated some critical aspects of faculty perceptions of the evaluation process, it has also opened up avenues for further critical research. First, related to the limitations in our demographics, future work should more intentionally explore perceptions of a wider range of faculty – in particular minoritized faculty in engineering. Literature has outlined various ways in which minoritized faculty might be at an inherent disadvantage when it comes to SET data, but we know relatively little about how these faculty perceive or make sense of that data when they do interact with it and how they make choices about responding to (or not responding to) student comments and feedback. Additional research should more specifically target engineering faculty from marginalized groups to further explore the implications of SET on instructional practices and examine if or how faculty perceptions differ across different demographics.

Moreover, future research should examine the degree of alignment between faculty and student perceptions concerning effectiveness. This work focused on faculty perceptions of their SET data but did not leverage student data from those individual instructors to corroborate or triangulate our findings. It is important that researchers examine the degree of alignment between faculty perceptions of SET data and the actual SET data that they receive from their institutions. For instance, do faculty who receive higher scores see more value in SET data? In contrast, when faculty receive less than desirable scores, to what do they attribute those scores? Future research should more closely examine the relationship between and potential alignment with faculty perceptions and actual SET data.

Finally, future work should expand on this research to explore faculty perceptions at a broader range of institution types and sizes. The present work was conducted at a high-research institution in which faculty merit is determined differently than, say, a small liberal arts college where the primary assignment for faculty is to teach courses. While we note that nearly all faculty in our research took pride in their pedagogical approaches and were concerned about their growth as educators, they also were aware of institutional demands related to research, funding, and publications and the ways those demands shape promotion and tenure decisions. Future research should therefore examine faculty perceptions of SET within different institutional settings to explore the ways in which beliefs and practices might differ.

6. Conclusion

Student evaluations play a critical role in institutional practices, from guiding pedagogical improvements to making administrative decisions, and faculty perceptions of these processes influence the way they go from data (i.e., evaluation) to improvement (i.e., pedagogy). Therefore, understanding faculty beliefs about these evaluation processes can inform the design of instruments that better guide and provide support for pedagogical changes. Based on interviews with 29 engineering faculty at a high-research activity university, we examined the ways in which faculty beliefs and practices about pedagogy relate to those concerning a range of student evaluation data. We implemented thematic analysis to explore faculty beliefs about pedagogy on the one hand and evaluation on the other.

Importantly, faculty in this study generally valued the use of student evaluation data for gaining insight about their teaching practices and guiding pedagogical improvements. But at the same time, they also noted a number of challenges related to using and making sense of that data. Faculty expressed confusion surrounding notions of ''effectiveness'' and described a range of challenges to creating an objective measure of it. Such findings point to the importance of leveraging existing scholarship to inform such measures and developing research-based instruments that offer more objective, equitable SET data. Further, there exists a tension in the ways data might be used for different stakeholders and for different purposes. While most participants recognize that student evaluation data is critical to informing pedagogical development over time, they also noted uncertainty surrounding how such data might inform decisions in settings beyond the classroom (e.g., promotion and tenure). Such findings point to the need to better understand the context in which student evaluation data is collected and then used by different interested parties. Student evaluation is likely to remain a topic of debate in engineering education, but a better understanding of how faculty use and make sense of that data can help inform institutional approaches to evaluation, offer insight to faculty, and guide pedagogical growth.

Acknowledgments – This material is based upon work supported by the National Science Foundation under Grant No. DUE 1347817. This material was also supported by funding from the California Polytechnic State University Summer Undergraduate Research Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or of Cal Poly.

References

1. V. Braun and V. Clarke, Using thematic analysis in psychology, *Qual. Res. Psychol.*, **3**(2), pp. 77–101, 2006.
2. K. A. Villanueva, S. Brown, N. Pitterson, D. Hurwitz and A. Sitomer, Teaching Evaluation Practices in Engineering Programs: Current Approaches and Usefulness, *Int. J. Eng. Educ.*, **33**(4), pp. 1–18, 2017.
3. G. Mohanty, J. Gretes, C. Flowers, B. Algozzine and F. Spooner, Multi-method evaluation of instruction in engineering classes, *J. Pers. Eval. Educ.*, **18**(2), pp. 139–151, 2005.
4. F. Guder and M. Malliaris, Online and paper course evaluations, *Am. J. Bus. Educ.*, **3**(2), pp. 131–138, 2010.
5. P. M. Simpson and J. A. Siguaw, Student evaluations of teaching: An exploratory study of the faculty response, *J. Mark. Educ.*, **22**(3), pp. 199–213, 2000.
6. W. R. McGowan, Faculty and student perceptions of the effects of mid-course evaluations on learning and teaching, *PhD Diss., Dep. Instr. Psychol. Technol. Brigham Young Univ.*, 2009.
7. J. O'Hanlon and L. Mortensen, Making teacher evaluation work, *J. Higher Educ.*, **51**(6), pp. 664–672, 1980.
8. C. Nygaard and D. Z. Belluigi, A proposed methodology for contextualised evaluation in higher education, *Assess. Eval. High. Educ.*, **36**(6), pp. 657–671, 2011.
9. L. M. Aleamoni, Student rating myths versus research facts from 1924 to 1998, *J. Pers. Eval. Educ.*, **13**(2), pp. 153–166, 1999.
10. M. Shevlin, P. Banyard, M. Davies and M. Griffiths, The validity of student evaluation of teaching in higher education: love me, love my lectures?, *Assess. Eval. High. Educ.*, **25**(4), pp. 397–405, 2000.
11. L. Layne, Defining effective teaching, *J. Excell. Coll. Teach.*, **23**(1), 2012.
12. F. Nasser and B. Fresko, Faculty views of student evaluation of college teaching, *Assess. Eval. High. Educ.*, **27**(2), pp. 187–198, 2002.
13. J. L. Osborne, Integrating student and peer evaluation of teaching, *Coll. Teach.*, **46**(1), pp. 36–38, 1998.

14. B. Lutz, A. Barlow, S. A. Brown and D. Sanchez, Exploring Faculty Beliefs About Teaching Evaluations: What is Missing from Current Measures?, 2018.
15. M. C. Hill and K. K. Epps, The impact of physical classroom environment on student satisfaction and student evaluation of teaching in the university environment, *Acad. Educ. Leadersh. J.*, **14**(4), p. 65, 2010.
16. D. E. Clayson, T. F. Frost and M. J. Sheffet, Grades and the student evaluation of instruction: A test of the reciprocity effect, *Acad. Manag. Learn. Educ.*, **5**(1), pp. 52–65, 2006.
17. M. Millea and P. W. Grimes, Grade expectations and student evaluation of teaching, *Coll. Stud. J.*, **36**(4), pp. 582–591, 2002.
18. B. Uttl, C. A. White and D. W. Gonzalez, Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related, *Stud. Educ. Eval.*, **54**, pp. 22–42, 2017.
19. D. E. Clayson and M. J. Sheffet, Personality and the student evaluation of teaching, *J. Mark. Educ.*, **28**(2), pp. 149–160, 2006.
20. A. Boring, K. Ottoboni and P. B. Stark, Student evaluations of teaching (mostly) do not measure teaching effectiveness, *Sci. Res.*, **10**, 2016.
21. K. G. Hendrix, Student perceptions of the influence of race on professor credibility, *J. Black Stud.*, **28**(6), pp. 738–763, 1998.
22. S. Secules and C. Turpen, The Construction of Competitive White Masculinity as Engineering Educational Culture, *Am. Anthropol. Assoc. Washington, DC*, 2017.
23. A. R. Bejerano and T. M. Bartosh, Learning masculinity: Unmasking the hidden curriculum in science, technology, engineering, and mathematics courses, *J. Women Minor. Sci. Eng.*, **21**(2), 2015.
24. K. M. W. Mitchell and J. Martin, Gender bias in student evaluations, *PS Polit. Sci. Polit.*, **51**(3), pp. 648–652, 2018.
25. M. B. Miles, A. M. Huberman and J. Saldaña, *Qualitative data analysis: A methods sourcebook*, SAGE Publications, Incorporated, 2013.
26. G. B. Rossman and S. F. Rallis, *Learning in the field: An introduction to qualitative research*, Sage, 2003.
27. M. Q. Patton, *Qualitative research*, Wiley Online Library, 2005.
28. B. L. Yoder, Engineering by the numbers. Washington, DC: American Society for Engineering Education, 2017.
29. S. Basow, S. Codos and J. Martin, The effects of professors' race and gender on student evaluations and performance, *Coll. Stud. J.*, **47**(2), pp. 352–363, 2013.
30. K. Andersen and E. D. Miller, Gender and student evaluations of teaching, *PS Polit. Sci. Polit.*, **30**(2), pp. 216–220, 1997.
31. A. Boring, K. Ottoboni and P. B. Stark, Student evaluations of teaching are not only unreliable, they are significantly biased against female instructors, *Impact Soc. Sci. Blog*, 2016.
32. R. M. Felder and R. Brent, Navigating the bumpy road to student-centered instruction, *Coll. Teach.*, **44**(2), pp. 43–47, 1996.
33. M. Prince, Does active learning work? A review of the research, *J. Eng. Educ.*, **93**(3), pp. 223–231, 2004.
34. S. L. Cutler, *How Static is the Statics Classroom? An investigation into how innovations, specifically Research-Based Instructional Strategies, are adopted into the Statics Classroom*, Virginia Polytechnic Institute and State University, 2013.
35. E. P. Douglas and C.-C. Chiu, Implementation of Process Oriented Guided Inquiry Learning (POGIL) in Engineering, *Adv. Eng. Educ.*, **3**(3), p. n3, 2013.
36. N. A. N. Azlina, CETLs: Supporting collaborative activities among students and teachers through the use of Think-Pair-Share techniques, *Int. J. Comput. Sci. Issues*, **7**(5), p. 18, 2010.
37. K. G. Nelson, D. F. Shell, J. Husman, E. J. Fishman and L. Soh, Motivational and self-regulated learning profiles of students taking a foundational engineering course, *J. Eng. Educ.*, **104**(1), pp. 74–100, 2015.
38. J. Zheng, W. Xing, G. Zhu, G. Chen, H. Zhao and C. Xie, Profiling self-regulation behaviors in STEM learning of engineering design, *Comput. Educ.*, **143**, p. 103669, 2020.
39. A. Barlow, S. Brown, B. Lutz, N. Pitterson, N. Hunsu and O. Adesope, Development of the student course cognitive engagement instrument (SCCEI) for college engineering courses, *Int. J. STEM Educ.*, **7**, pp. 1–20, 2020.
40. S. Tharayil, M. Borrego, M. Prince, K. A. Nguyen, P. Shekhar, C. J. Finelli and C. Waters, Strategies to mitigate student resistance to active learning, *International Journal of STEM Education*, **5**(1), pp. 1–16, 2018.

Ahmed Osman is a fourth-year Civil Engineering student at Cal Poly San Luis Obispo and serves as an undergraduate research assistant through the Department of Mechanical Engineering. He has worked with the two authors previously through the Critical Research in Engineering and Technology Education (CREATE) group at Cal Poly. His passion for research is driven by his interest to gauge implicit factors within engineering education to make the learning environment equitable and strive for diversity within engineering. His current work consists of thematic analysis to steer thoughts and deepen analysis of faculty practices and beliefs.

Eric Cuellar is a fourth-year industrial engineering student at Cal Poly San Luis Obispo. He is a member of the Critical Research in Engineering And Technology Education (CREATE) group in the mechanical engineering department. He is passionate about optimizing the creative capacity of engineering design teams. His research interests include the ideation process, metrics for creativity, idea selection methods, diversity in engineering, and equitable design practices.

Shane Brown is a professor in the School of Civil and Construction Engineering at Oregon State University. His research interests are in situated cognition, conceptual change, and social capital. He received his BS and PhD from Oregon State University and his MS from University of California at Davis, all in civil engineering. He is a recipient of several research and teaching awards, including the NSF CAREER award, best paper awards from the Educational Research and Methods and Civil Engineering Divisions of ASEE, and the ASCE ExCEEd National New Faculty Excellence in Teaching Award.

Benjamin D. Lutz is an Assistant Professor of Mechanical Engineering Design at Cal Poly San Luis Obispo. He is the leader of the Critical Research in Engineering and Technology Education (CREATE) group at Cal Poly. His research

interests include critical pedagogies; efforts for diversity, equity, and inclusion in engineering, engineering design theory and practice; conceptual change and understanding; and school-to-work transitions for new engineers. His current work explores a range of engineering education contexts, including the role of power in brainstorming activities, epistemological and conceptual development of undergraduate learning assistants, as well as the experiences of recent engineering graduates as they navigate new organizational cultures.