Exploring Faculty Beliefs About Teaching Evaluations: What is Missing from Current Measures?

Dr. Benjamin David Lutz, Oregon State University

Ben Lutz is a Postdoctoral Scholar in Engineering Education at Oregon State University. His research interests include innovative pedagogies in engineering design, conceptual change and development, school-to-work transitions for new engineers, and efforts for inclusion and diversity within engineering. His current work explores how students describe their own learning in engineering design and how that learning supports transfer of learning from school into professional practice as well as exploring students' conceptions of diversity and its importance within engineering fields.

Ms. Allyson Jo Ironside, Oregon State University

Ally Ironside is a recent graduate from LeTourneau University where she studied Water Resources in Civil Engineering. She is currently fusing her technical background with her passion for education in pursuing a doctoral degree in Civil Engineering while conducting research in Engineering Education at Oregon State University. Her research interests include the adoption of teaching best practices in engineering and the personal epistemology development students.

Dr. Shane A. Brown P.E., Oregon State University

Shane Brown is an associate professor and Associate School Head in the School of Civil and Environmental Engineering at Oregon State University. His research interests include conceptual change and situated cognition. He received the NSF CAREER award in 2010 and is working on a study to characterize practicing engineers' understandings of core engineering concepts. He is a Senior Associate Editor for the Journal of Engineering Education.

Ms. Dominga Sanchez, Oregon State University

Exploring Faculty Beliefs about Teaching Evaluations: What is Missing from Current Measures?

Abstract

In this research paper, we explore faculty perceptions about what information or data is missing from current evaluation approaches. Faculty evaluation of teaching is ubiquitous across engineering education, and the results of those evaluations play critical roles in institutional decision making. And while numerous studies have explored faculty perceptions of existing measures, relatively less work has asked faculty about what kinds of additional data or information might improve their teaching. To address this gap, we asked engineering 20 faculty members across eight engineering departments what they perceived to be missing from current evaluation practices and measurements. Inductive coding approaches revealed three major areas in which additional information would help improve their teaching. Faculty noted 1) the importance of soliciting additional student feedback beyond traditional student evaluations at the end of the semester; 2) the need for more data regarding student retention and transfer of concepts learning in class; and 3) the potential for soliciting additional peer feedback from colleagues and educational researchers. At the same time, some faculty were satisfied with current approaches to teaching evaluation and did not perceive anything to be missing. Findings point to the opportunity to collect more in-depth, qualitative feedback regarding faculty teaching effectiveness. In particular, expert consultation and creating more spaces to solicit written comments from students might help faculty obtain evaluation data that can both aid in both institutional and pedagogical decision making.

Introduction

Faculty evaluation of teaching plays a critical role in engineering education, and while data collected through evaluation and assessment processes are ostensibly designed to offer feedback to faculty regarding their effectiveness as instructors, they most often play central roles in discussions of promotion and tenure. Further, despite nearly a century of research surrounding the validity, reliability, efficacy, etc. of evaluations of teaching (Andersen & Miller, 1997; Marsh & Roche, 1997; Pounder, 2007; Spooren, Brockx, & Mortelmans, 2013; Uttl, White, & Gonzalez, 2017), higher education remains generally divided on whether such data belong in merit-based discussions in general. Despite myriad studies across contexts, time periods, and populations, there is no clear consensus on how or if evaluation of teaching should occur, especially as it pertains to student evaluations.

If faculty do not see their evaluation results as meaningful or helpful for their improvement, it is only logical they will not use them to make decisions about if or how to improve their teaching. It is thus critical that faculty obtain evaluation results that not only provide them with an accurate depiction of their teaching effectiveness, but that offer insight into strategies for growth and development. And while decades of research have investigated faculty perception of *existing* evaluation and assessment practices, less has asked faculty what they would like to know in order to improve their teaching. To that end, we pose the following research question:

*What kind of information or data do faculty perceive as meaningful and impactful that is **missing** from current evaluation practices?*

To answer this question, we present a qualitative analysis of 20 interviews with engineering faculty at a predominantly white research university in the Pacific Northwest. Specifically, we explore faculty responses to the question "Is there any information that would help you improve your teaching but that you don't currently have access to?" We focus on this question because, as noted, most research to date has focused on faculty interpretations of evaluation data that already exists or that they have already been evaluated with. Asking participants to imagine data or information they would like to have is useful because it introduces the opportunity to develop practical instruments that provide faculty with vital information. Including faculty in the process of developing their own evaluation tools and metrics increases the likelihood of their usefulness in terms of both faculty development and promotion and tenure decisions.

Literature Review

Two major strands of literature are germane to the present study: current faculty evaluation practices in STEM and faculty perceptions of evaluations of teaching. Faculty evaluation is an important component of higher education, engineering included (Pitterson, Brown, Villanueva, & Sitomer, 2016). Faculty are evaluated based on their "effectiveness" as instructors through a range of methods, all of which have advantages and drawbacks. But these different approaches are often used to inform discussions of merit, promotion, tenure, and other personnel decisions within academia (Schmelkin, Spencer, & Gellman, 1997). As one might imagine, some approaches can be controversial, and the validity and reliability of teaching evaluations—especially student evaluations—is still a heated debate after nearly a century of research (Uttl et al., 2017). The following sections will highlight some current work in faculty evaluation of teaching with a specific focus on student evaluations of teaching (SET) and describe current issues with common evaluative approaches. Ultimately, we will argue that additional exploratory work is needed to better understand what faculty need to improve their instruction.

Faculty Evaluation in Engineering

Virtually all universities and colleges perform some kind of faculty evaluation, but how and in what ways that evaluation takes place can be substantially different across contexts. Villanueva, Brown, Pitterson, Hurwitz, and Sitomer (2017) conducted a broad survey of the current state of faculty evaluation practices in engineering, and they identify some common trends. More specifically, they found that across a range of institutional settings and types, the primary mode of evaluation is either end-of-term student evaluations of teaching (SET) or faculty observations, or some combination of the two (Villanueva et al., 2017). Other less common methods included mid-term evaluations, graduate and alumni interviews, and evaluation of classroom materials (e.g., syllabus, homework, exams). Though each form of evaluation might have its individual drawbacks and advantages, combining approaches generally helps to mitigate the weaknesses associated with any single method and paint more complete accurate picture of teaching effectiveness.

Although many faculty reported using some combination of measures to conduct evaluation, the way these measures were used were highly variable. For instance, SET is often mandatory and

used for summative purposes, while mid-terms evaluations were typically conducted voluntarily and used for formative assessment. Some universities used SET for formative assessment and peer observations for summative assessment, while others did the opposite. These findings are interesting because while they do point to significant overlaps and similarities in evaluation practices at a macro-level (i.e., most schools use SET and/or peer observations with other supplementary forms of data), the way these evaluations actually take place at a micro-level (i.e., what is evaluated, how the evaluations are used) can be drastically different. It is as if everyone is doing the same thing, while at the same time doing it differently.

Student Evaluation of Teaching (SET)

Even though SET is one of many different methods, it remains the primary mechanism through which faculty are provided information about the effectiveness of their teaching. Moreover, it often carries significant weight in decisions surrounding promotion and tenure (Zabaleta, 2007). It is therefore important to elaborate on the role of SET in both higher education as well as engineering specifically, especially as it pertains to the way they have been and currently are used. Despite nearly a century of research, faculty in higher education remain divided in regards to the use of SET (Uttl et al., 2017).

Typically, SETs take place within the last few weeks of a course, before grades are released. SET forms might be filled out online or with paper and pencil and typically take the form of a set of Likert-type questions that probe student perceptions of the instructor, classroom environment, learning outcomes, and an evaluation of the course and instructor. Students are often asked to rate the degree to which they agree (e.g., from "strongly disagree" to "strongly agree") with statements such as "Overall, my instructor's teaching was effective." These scores are typically averaged and used for comparisons across departments, colleges, and potentially the entire university. In some cases, SET forms will include an open-ended text box where students can elaborate or express concerns not addressed in the quantitative questions. Given the pervasiveness of SET and their influence in organizational and personnel decision making, it is perhaps unsurprising that credibility or validity of such measures have been the subject of much scrutiny (Spooren et al., 2013).

Issues with SET

Despite nearly a century of research surrounding Student Evaluation of Teaching (SETs), their use and meaning remain controversial across a range of university settings (Uttl et al., 2017). Countless studies have explored the ways faculty make sense of, interpret, make use of, and generally engage with SET data (Centra & Gaubatz, 2000; Clayson, 2009; Laube, Massoni, Sprague, & Ferber, 2007; Pounder, 2007; Shevlin, Banyard, Davies, & Griffiths, 2000; Simpson & Siguaw, 2000; Zabaleta, 2007) and numerous issues have been raised surrounding both their accuracy in measurement and the factors that SET scores might be sensitive to. In fact, Aleamoni (1999) sixteen different "myths" about SET scores that have pervaded common perceptions of their meaningfulness for improving teaching. Proponents of SET see students as qualified evaluators with the most complete information (e.g., (Albanese, Schuldt, Case, & Brown, 1991)), while opponents argue that students are more often reporting on their general satisfaction, which

can be related to many factors unrelated to effectiveness in instruction (Shevlin et al., 2000; Simpson & Siguaw, 2000).

Common factors alleged to influence SET are demographic variables such as gender or ethnicity of the instructor (Andersen & Miller, 1997; Basow, Codos, & Martin, 2013; Centra & Gaubatz, 2000). Indeed, one can imagine that in engineering settings typically dominated by White masculine culture and values (Dryburgh, 1999; Tonso, 1996), faculty who do not resemble the normative traits (e.g., women, people of color) might be at a disadvantage when compared to their White male counterparts. For example, MacNell, Driscoll, and Hunt (2015) illustrated how simply changing the instructor name in an online course to a woman's name can result in lower scores—even when instruction is identical. More recently, Boring, Ottoboni, and Stark (2016) corroborated these findings and, to add to the complexity of the issue, demonstrated how even though gender bias exists, there is no statistical means to control for it. Less work has explored the links between race or ethnicity and SET scores, but some initial research in this space (e.g., (Basow et al., 2013; Merritt, 2008)) suggests that these interactions are equally complex and raise important questions about the ways SET is and should be used in higher education and engineering in particular.

The literature thus points to two critical gaps. First, while faculty evaluation and SET in particular were initially designed to provide feedback for faculty to improve their teaching, they have slowly moved away from that role and moved increasingly toward an indicator of accountability. Rather than providing meaningful data for faculty to reflect on, they are generally used by administrators to inform personnel decisions. It is therefore critical that researchers understand how to make faculty interaction with evaluation data more useful—especially considering their susceptibility to bias along lines of privilege and oppression. Moreover, most research in this space has investigated the use and perceptions of *existing* evaluation measures with relatively less focus on suggestions for new data. That is, fewer studies have asked engineering faculty to imagine new data, information, or stakeholders that might help them improve. Such a process would be able to leverage human-centered design principles to develop a more valid, meaningful approach to faculty evaluation.

Methods

The present work was conducted as part of a larger study of teaching effectiveness across STEM at the research site (Bouwma-Gearhart, Perry, & Presley, 2014; Koretsky, 2015; Koretsky et al., 2016). The research project is a broad institutional effort to cultivate interdisciplinary communities of practice focused on improving teaching practices in STEM. Faculty evaluation thus plays a critical role in enhancing teaching practices because the data collected and feedback provided are designed to help faculty improve as instructors. As a result, this research focuses on faculty perceptions of teaching evaluation data, in particular the kinds of data they would like to collect or stakeholders they would like to engage to improve their teaching.

Data Collection and Sample

Faculty were recruited through contact with school heads and subsequent email invitations to individual faculty inviting them to participate. Faculty were sent an email that explained the terms of the study and provided instructions to schedule an interview with the first Author.

Interviews were semi-structured and explored faculty beliefs about the use of teaching evaluations in the university. We asked faculty to describe the ways their teaching was currently evaluated, the way the university used the data, the way they used the data themselves, and what they thought was missing from current available data. Interviews were audio recorded, transcribed by a professional service, scrubbed of identifying information, and imported to MaxQDA (version 11.2) qualitative analysis software.

The current study focuses on participants' recommendations for additional data or information because it introduces the possibility of generating concrete, actionable data collection approaches. By listening to user needs, the output of evaluation measures is more likely to have practical value and therefore lead to meaningful changes in instruction.

Data analysis

In this preliminary analysis, we focus on responses to the question "Is there any information that would help you improve your teaching but that you don't currently have access to?" We focus on this question because although numerous studies regarding faculty beliefs of teaching evaluations have demonstrated concerns with the current approaches, relatively less work has explored the recommendations faculty might have for improving the process or data collected. This is important for two key reasons. For one, related research has demonstrated the positive effects of including faculty in the shaping of their own professional development (Barner, Brown, Lutz, & Montfort, 2018, forthcoming). Including faculty in the revision or enhancement of teaching evaluation positions them as stakeholders in their own success and increases engagement and motivation. Second, a deeper understanding of what faculty would like to know about their students also sheds light on other beliefs about teaching and learning more broadly. For example, faculty who are interested in having a one-on-one conversation with their students might have fundamentally different beliefs about teaching and learning than someone who would like to open their classroom up to more third-party observations and assessment. More specifically, the former might see teaching as a private, personal act while the latter might approach teaching from a more interactive, community-based model of pedagogy.

We performed inductive, descriptive coding consistent with recommendations from Miles, Huberman, and Saldaña (2014) for two reasons. First, because this research was exploratory, we intentionally did not develop any *a priori* instrumentation to guide our analysis. Given that we were asking faculty to imagine alternate information, data, or stakeholders in the evaluation process, entering analysis with predetermined categories could potentially limit the scope of our findings. Second, a goal of the larger project of which this research is a part is designed to improve teaching and learning practices *within the university*. Importantly, inductive methods allow for salient local factors to emerge from the analysis. As a result, we allowed codes to emerge progressively and we refined as we iterated through each interview.

Results

Inductive analysis resulted in four major descriptive codes that characterize the kinds of information or data faculty would like to receive but that either does not exist or that they don't have access to. Many faculty were interested in the possibility of acquiring more or richer information from students—both in terms of perceptions of the class and in meeting learning

outcomes. Faculty also noted the role and importance of third-party observations or consulting, such as an assessment from an educational researcher. Other faculty members were satisfied with current measures and were unsure about what kinds of additional data might help them improve their teaching. The inductive codes and their operationalized definitions are provided in Table 1 and described in more detail in the following sections.

Table 1: Overview of emergent codes, definitions, and examples from current data set.

Code	Definition	Example
More student perceptions	Interest in either collecting more student data or developing approaches for gathering richer data.	- Mid-term assessment - One-on-one conversations - More written comments
Learning and retention	Interest in how much of what was taught was retained and/or transferred to some new context over longer periods of time.	- Learning beyond the course - Function in the workplace - Process vs. product
Third party input	Interest in opening up classroom to third-party observations, assessments, and critiques.	- Sustained observations - Educational researcher assessment - Pedagogical advice
Not sure	Faculty are unsure about what else they might like to know about their teaching or ways to improve it.	- Satisfied with current system - Information is complete

The first two codes, *More student perceptions* and *Learning and retention*, represent faculty concern for how students react to, perceive, and ultimately learn from a course. Faculty in this study wanted to gain a deeper perspective of students' experiences by asking about, and were also interested the way these experiences transferred beyond a course. *Third party input* captures faculty interest in subjecting their classroom to critiques from other faculty both in and out of engineering. Lastly, some faculty were simply *Not sure* what kind of information was missing or if any was missing. The following sections elaborate on each descriptive code, supported throughout by participant quotes.

More student perceptions

Faculty reported wanting access to more student perceptions in two main ways: the volume of perception data and the richness of student responses. In terms of the amount of data collected, faculty discussed the potential to solicit additional feedback at times additional points during the course, including before the end of the term.

> *[Respondent]: Now, if we could decide a better SET set of questions and have a midterm SET, that would be useful. […] So you have a little bit of you get either a green light, a yellow, or red, like things are not going well.*
>
> *Ben: Yeah.*

[Respondent]: So at least you have still some time to correct things or to organize certain things where the students start liking it and they're feeling comfortable with you.

Participants noted that more student perception data collected earlier in the term would enable them to more effectively respond to issues that might arise throughout term. Relatedly, faculty perceived administering the SET early as empowering more students to answer, as they are potentially able to experience changes they recommend. Administering SET instruments only at the end of a term placed limits on both the number of suggestions as well as changes faculty were able to make in response to the unique needs of a particular class.

In addition to a greater quantity, participants also discussed the need for richer (i.e., greater quality) evaluation data. Written and verbal student comments often provide insight that are more specific or relevant to a class, and faculty in this study discussed the potential usefulness of introducing more opportunities to provide written and other forms of open-ended feedback, as opposed to quantitative, Likert-type responses typical of SET instruments.

[Respondent]: I think I would like more detailed feedback from the students. Like that time I said I had a student or two who had really low scores, I would really like to know why that happened. The more written feedback they gave, the better.

Other participants suggested going beyond written comments and setting up opportunities for one-on-one conversations with students. By having more candid conversations, faculty believe that might be able to learn more from students.

[Respondent]: I guess if we could talk one-on-one with a bunch of students about that, and have them be candid.

Ben: What would you talk about? What kinds of questions would you ask them, what would you want to learn from those one-on-ones?

[Respondent]: I mean, first off, what'd you think of the class? Was it good, did it work, did it not work? And then have them not just go, "Yeah, it seemed okay." Actually go beyond that and have a bunch of follow-up questions.

Responses such as these illustrate the influence of qualitative feedback from students on faculty motivation to improve their instruction.

Learning and retention

Perhaps unsurprisingly, faculty beliefs about effective teaching were intertwined with their beliefs about effective learning. As a result, participants also noted the potential usefulness of collecting more robust data surrounding longer-term learning and retention of course content. Participants here were concerned with whether students were able to transfer and/or apply their learning from one class to another. The following quote illustrates the need to understand learning beyond time constraints of the course.

[Respondent]: Lots of time, I plant seeds. There are some students who are transformed by my classes and really develop better critical thinking skills. I wish I was better ... how do you evaluate critical thinking? These are kind of hard. These are the kind of things that I wanna get at. And then plant seeds around [my topic], and there are some students that just aren't ready to engage the topics in the way that I would hope they would engage. But they come back to me five years later and say, "I remember ... you know, this just happened to me and I remember that conversation, I remember that article you read in your class." It would be great if I had longitudinal data. That would be really fun.

Sometimes the resultant learning from a class might not immediately materialize, but it does not mean that learning has not occurred. The participant in the quote above recognizes the potential utility of longitudinal data to evaluate learning outcomes around critical thinking in ways that account for latent effects beyond the class. Other participants discussed ways performance data beyond the university and into their jobs.

[Respondent]: That's one of the big ones that I go for. I would like my students to be functional when they enter the workplace… Whether it's at a consulting firm, a public agency, a research laboratory, as a faculty member, I'd like them to be able to know about and in some detail the kinds of things that they'd be expected to do when they get there.

Many faculty see their roles as preparing students to contribute to their engineering profession of choice. Therefore, knowing more about how students transfer their learning to a workplace can potentially help faculty improve their instructional approaches to more adequately meet the needs of the discipline. Participants recognized that measuring and assessing learning is complex, and doing so within the timeframe of a single course presents limitations when it comes to understanding more distal effects of their instructional and pedagogical approaches.

Third party input

Though student evaluations of teaching tended to dominate conversations, faculty also discussed the value of peer observations and feedback. At the same time, however, they recognize that single, sporadic observations can offer relatively limited insight into ways to improve instruction. One suggestion participants made was to engage in more sustained observations in order to help the observer better understand the context of the course and therefore offer more useful feedback.

[Respondent]: I mean… peer evaluations are good. Maybe more of those?

Ben: Okay.

[Respondent]: Maybe more than once a term, that would be kind of like within the department. That might be something useful, I guess.

Engineering faculty in this study also recognized that they are disciplinary experts, and not necessarily learning scientists. As a result, they suggested the potential for inviting pedagogical or learning experts in to observe and offer feedback.

> *[Respondent]: You know I'm trying to think about if I take that to the extreme, you know I could imagine a situation where some sort of education expert, somebody like [researchers on this team], but like ... would sit in my class for an entire term and you know, effectively take the class and then at the end say here are some things that you could do to make this better, you know, you could do some active learning over here, you could maybe incorporate a design project, you could change the midterm like this, but I don't know that I know exactly what the specific information would be...*

The participant expresses an interest in critique or scrutiny from an engineering education researcher, but also struggles to specify the kinds of feedback they might offer. Participants were motivated to improve their instruction, but were not always sure how or in what ways those improvements should be made. Supporting faculty with both expertise and sustained contact could help improve instruction by facilitating both content- and context-specific feedback.

Not sure

Finally, some faculty struggled to imagine any additional data or information that might help them improve their teaching. One participant noted the current evaluation "covers what they learn from the class," while another described how the combination of student evaluations and faculty observations complement each other.

> *[Respondent]: So the students give you feedback about what the people who come to evaluate you don't see. Like timely grading and all those things that the students care about. They're going to give you feedback about that.*

Responses in which faculty were unsure were rare, and they primarily expressed satisfaction with the current evaluation practices. Importantly, not being able to imagine an alternate form of evaluation in an interview setting does not imply poor teaching. Indeed, most faculty in this sample reported above-average SET scores and were satisfied with their evaluation results.

Discussion

The results suggest that faculty in this study are generally interested in collecting new or additional information in order to improve their teaching. Students, peers, and external reviewers were noted as potential sources of data. The kinds of information participants expressed interest in echo current discussions surrounding faculty evaluation, and the following sections highlight areas of support within existing literature.

Consulting an expert

Villanueva et al. (2017) noted that although most faculty reported being evaluated by an observer, they also expressed concerns regarding the perceived lack of training of the observers. Though many faculty are evaluated through peer observation, many of those observers might not be credentialed in ways that provide faculty with confidence in the results. Indeed, participants'

desire for *Third party input* corroborates this finding. Faculty in this study described their peer observations as incomplete or lacking the context needed to accurately evaluate their teaching. Including an educational researcher in the evaluation process, especially for extended periods of time, might increase faculty members' confidence in the results of the evaluation and make them more likely to make pedagogical changes in the future.

Probing for richer student data

A rich body of literature demonstrates faculty members' persistent concerns regarding the validity of student evaluations as a useful measure of effectiveness (Boring et al., 2016; Spooren et al., 2013; Uttl et al., 2017). Similarly, common concerns raised by participants in this study concerned whether the information provided by quantitative measures actually reflect one's effectiveness in instruction. Spooren et al. (2013) argue that the use of SET remains contentious, and the present findings related to *More student perceptions* as well as *Learning and retention* support this claim. A desire to collect more data suggests that current mechanisms do not provide enough to inform change, while the need to understand longer-term learning suggests a degree of uncertainty regarding their own measures of learning. Further, participants also noted the potential usefulness of qualitative, written evaluation responses or the opportunity to converse with students after the course. Such findings align with previous work surrounding faculty beliefs about students' qualifications to evaluate teaching (Schmelkin et al., 1997). By asking for more in-depth responses from students, our participants seemed to acknowledge that students can offer useful information, but perhaps only in certain contexts.

Implications and Future Work

Revision of SET items

One noteworthy finding of the present research concerns faculty beliefs surrounding student interaction with current SET instruments. While most participants eschewed the coarseness and ambiguity associated with quantitative items on SET forms, the optional written comments were often perceived as highly valuable. The quantitative measures are challenging to parse out because of all the other factors (discussed above) that can be conflated in that number. However, at the same time, most faculty believe that only outliers (i.e., those who are either very satisfied or dissatisfied) actually take the time to write comments. So although the written comments provide a rich form of feedback, faculty beliefs about which students are responding can affect the influence those comments have on their desire improve their teaching.

Thus, one way to encourage more handwritten, qualitative evaluation data would be to revise the structure of the SET instrument to promote written (i.e., typed) comments. Lee and Lutz (2016) developed survey items which they term *Anchored Open-ended* or AOE questions. The basic premise of an AOE question is that each item asks the student to respond a trinary scale—positive, negative, neutral—and provide an explanation for the response. In this way, the response to the trinary scale serves as the *anchor* to the accompanying open-ended response. For example, current SET questions at the research site ask students to rate their "instructor's contribution" to their learning on a scale from *very poor* to *excellent*, as shown in Figure 1.

	Very Poor	Poor	Fair	Good	Very Good	Excellent
My instructor's contribution to this course						
My instructor's contribution to my learning						

Figure 1: Example of current SET format and structure.

However, an AOE question might be reformatted to be structured as in Figure 2, in which the response is only positive, negative, or neutral, but requires an explanation for the reason the contribution is considered as such. This structure is fundamentally different from the approach currently used, which asks students to add general comments after responding to all the quantitative items.

	Positive	Neutral	Negative
My instructor's contribution to this course was			
Please explain your answer:			
My instructor's contribution to my learning was			
Please explain your answer:			

Figure 2: Example of AOE format for student evaluations.

Given this approach, students would respond to a simplified scale, but provide a qualitative elaboration on the student perception of teaching effectiveness. However, this approach would also likely take more time for students to complete, and could cause survey fatigue if overused. But most modern survey tools support such a design, and the authors suggest future work should pilot this approach within evaluative settings to verify its utility.

Conclusion

Faculty are interested in collecting evaluation data in order to help them improve their teaching, but some current approaches might not accommodate such collection. We interviewed 20 engineering faculty across three schools at a major research institution in the Pacific Northwest and asked them to imagine data or stakeholders they perceive as currently missing in terms of improving their practice. Participants noted the need for more and richer student data, longitudinal data regarding conceptual understanding and transfer, and input from a third-party expert such as an educational researcher. These findings align with existing research and contribute to the debate surrounding the usefulness of faculty evaluation measures—student evaluation in particular. We argue that in order to improve faculty perceptions of teaching evaluation practices, future work should explore the potential of AOE questions for more meaningful approaches to faculty evaluation. Such approaches might provide faculty with concrete, useful suggestions, and empower them to make positive pedagogical changes.

References

Albanese, M. A., Schuldt, S. S., Case, D. E., & Brown, D. (1991). The validity of lecturer ratings by students and trained observers. *Academic medicine*, *66*(1), 26-28.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*(2), 153-166.

Andersen, K., & Miller, E. D. (1997). Gender and student evaluations of teaching. *PS: Political science & politics, 30*(2), 216-220.

Barner, M., Brown, S., Lutz, B., & Montfort, D. (2018, forthcoming). Engineering Faculty Interpretations of Pull-Oriented Innovation and Why Context Matters. *The International journal of engineering education*.

Basow, S., Codos, S., & Martin, J. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal, 47*(2), 352-363.

Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research, 10*.

Bouwma-Gearhart, J., Perry, K. H., & Presley, J. B. (2014). Improving postsecondary STEM education: Strategies for successful interdisciplinary collaborations and brokering engagement with education research and theory. *Journal of College Science Teaching, 44*(1), 40-47.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education, 71*(1), 17-33.

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

Dryburgh, H. (1999). Work Hard, Play Hard: Women and Professionalization in Engineering - Adapting to the Culture. *Gender & Society, 13*(5), 664-682.

Koretsky, M. (2015). Program level curriculum reform at scale: Using studios to flip the classroom. *Chemical Engineering Education, 49*(1), 47-57.

Koretsky, M., Bouwma-Gearhart, J., Dick, T., Brubaker-Cole, S. J., Sitomer, A., Fisher, K. Q., . . . Kayes, L. J. (2016). *Enhancing STEM Education at Oregon State University-Year 2.* Paper presented at the 2016 ASEE Annual Conference & Exposition.

Laube, H., Massoni, K., Sprague, J., & Ferber, A. L. (2007). The impact of gender on the evaluation of teaching: What we know and what we can do. *NWSA Journal, 19*(3), 87-104.

Lee, W. C., & Lutz, B. (2016). *An Anchored Open-Ended Survey Approach in Multiple Case Study Analysis*. Paper presented at the American Society for Engineering Education Annual Conference, New Orleans, LA.

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*(4), 291-303.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187.

Merritt, D. J. (2008). Bias, the brain, and student evaluations of teaching. . *John's L. Rev., 82*, 235.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Pitterson, N. P., Brown, S., Villanueva, K. A., & Sitomer, A. (2016). *Investigating current approaches to assessing teaching evaluation in engineering departments.* Paper presented at the Frontiers in Education Conference (FIE), 2016 IEEE.

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

Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. *Research in higher education, 38*(5), 575-592.

Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: love me, love my lectures? *Assessment & Evaluation in Higher Education, 25*(4), 397-405.

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education, 22*(3), 199-213.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598-642.

Tonso, K. L. (1996). The Impact of Cultural Norms on Women*. *Journal of Engineering Education, 85*(3), 217-225.

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

Villanueva, K. A., Brown, S., Pitterson, N., Hurwitz, D., & Sitomer, A. (2017). Teaching Evaluation Practices in Engineering Programs: Current Approaches and Usefulness. *International Journal of Engineering Education, 33*(4), 1-18.

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in higher education, 12*(1), 55-76.