Instructional Profiles: Exploring Peer-Observation at an Engineering College

Dr. Tareq Daher, University of Nebraska-Lincoln

Tareq Daher earned his Bachelors in Computer Science from Mutah University in Jordan. He pursued a Master's of Instructional Technology at the University of Nebraska –Lincoln while working as the coordinator for the Student Technology Program on the UNL campus. Currently, Dr. Daher works as the Director for the Engineering and Computing Education Core at the University of Nebraska – Lincoln. Dr. Daher collaborates with engineering faculty to document and research the integration of innovative instructional strategies and technologies in their classrooms and designs and delivers professional development programs for faculty in the college.

Dr. Jody Koenig Kellas, University of Nebraska-Lincoln

Jody Koenig Kellas (Ph.D., University of Washington, 2002) is a Professor of Communication Studies at the University of Nebraska-Lincoln (UNL) in the area of interpersonal, family, and health communication. She is trained in both quantitative and qualitative methods and has published over 50 articles and book chapters, as well as an edited book on storytelling in the family. She has overseen more than a dozen research projects and is a leader in the communication field on narratives and storytelling in the family. The goal of her research program and research lab – Narrative Nebraska – is to study the ways in which narratives and storytelling can be translated to help individuals and families understand, negotiate, and improve communication and coping within and outside of healthcare contexts. She is also co-director of UNL's Peer Review of Teaching Project – an interdisciplinary, faculty-led, campus-wide professional development project dedicated to helping faculty demonstrate the scholarly work of teaching and learning.

Dr. Wayne A. Babchuk, Department of Educational Psychology, University of Nebraska-Lincoln

Wayne A. Babchuk is an Associate Professor of Practice in the Department of Educational Psychology (Quantitative, Qualitative, and Psychometrics Program), the Department of Anthropology, and the Department of Sociology at the University of Nebraska-Lincoln (UNL).. He also serves as Dean's Fellow of the College of Arts and Sciences Teaching Academy at UNL and as a Research Associate for the Kalahari Peoples Fund of southern Africa. As a research methodologist, he teaches and conducts research on the history, epistemology, application, and instruction of qualitative research across disciplines, research ethics, grounded theory, ethnography, grounded ethnography, and mixed methods. He is also involved in several other research tracks including faculty teaching and evaluation strategies, interdisciplinary collaboration, teaching applied anthropology, Kalahari San land and resource rights, research to practice links in minority health care, and student and instructor perceptions of the impact of social media on student success. With a broad and diverse background in both education and the social sciences, he strives to bring a holistic and interdisciplinary approach to all aspects of teaching, research, and service.

Dr. Lance C. Perez, University of Nebraska, Lincoln

Dr. Lance C. Pérez received his B.S. in Electrical Engineering from the University of Virginia, and his M.S. and Ph.D. in Electrical Engineering from the University of Notre Dame. He is currently the Omar H. Heins Professor of Electrical and Computer Engineering at the University of Nebraska-Lincoln where he directs the Perceptual Systems Research Group. His research interests include information, video and signal processing, engineered healthcare and engineering education. He was appointed Dean of the College of Engineering in May 2018.

Instructional Profiles: Exploring Peer-Observation at an Engineering College

Abstract—This evidence-based research paper explores engineering faculty's instructional profiles emerging from COPUS observations based on the work of Stains et al. (2018). Multiple peer-observations of instructors teaching undergraduate classes within a College of Engineering at a large, Midwestern research-intensive institution were conducted. Faculty and graduate student paired teams conducted the observations. Upon completion of the classroom observations, researchers conducted exit interviews during which the results of the TPI, COPUS, and instructional profiles were shared with faculty. Follow-up semi-structured interviews were conducted with instructors to explore their experiences with the COPUS protocol, ascertain their perceptions of a new teaching evaluation system, and to gain insight into the instructional profiles of engineering faculty. Our analysis holds important and timely implications for how engineering courses are structured, evaluated, and viewed by faculty and administrators.

Keywords— COPUS, Peer-Observation, Engineering Faculty Experiences, Instructional Profiles, Teaching Practices Inventory.

I. INTRODUCTION

A growing body of research stresses the importance of implementing empirically based instructional strategies in STEM Education (Hora, M. T., & Ferrare, J. J., 2013; Stains, M., & Vickrey, T., 2017; Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. ,2002). Research unequivocally indicates that strategies that engage students cognitively and with one another lead to improvements in learning in a number of domains (see Stains et al., 2018). In the context of engineering education, Smith, Sheppard, Johnson, and Johnson (2005) overview the effectiveness of and call for pedagogies of engagement, asserting "To maximize students' achievement, especially when they are studying conceptually complex and content-dense materials, instructors should not allow them to remain passive while they are learning" (p. 14). Towards this end, Smith, Jones, Gilbert, & Wieman (2013) developed a peer-observation instrument and process effective for collecting information about Science, Technology, Engineering, and Mathematics (STEM) teaching practices. The Classroom Observation Protocol for Undergraduate STEM (COPUS) was developed as a means to document and characterize both instructor and student behaviors in the classroom. The data gathered in COPUS can be used to document teaching practices employed by instructors, departments, and colleges. Additionally, the data can be used as an objective source of formative feedback for potential peer mentoring professional development programs or self-evaluation. Smith et. al. (2013) created the validated COPUS with 25 codes (e.g., instructor lectures, instructor writes, student ask question, clicker questions, etc.) that observers mark within 2-minute intervals. Observers can be trained to use the protocol during a 1.5 hour period, reducing the substantial training required by other commonly utilized protocols such as Teaching Dimensions Observation Protocol (TDOP) and Reformed Teaching Observation Protocol (RTOP) (Smith, et al., 2013).

Since the introduction of COPUS, researchers have further validated the tool and confirmed that it characterizes instructional practices in the classroom (M. K. Smith, E. L. Vinson, J. A. Smith, J. D. Lewin, M. R. Stetzer, 2014; G. L. Connell, D. A. Donovan, T. G., 2016; T. J. Lund *et al.*, 2015). Stains et. al (2018) investigated the teaching practices in over 2000 STEM classes across 25 universities and 500 faculty members observed using COPUS. They conducted a latent profile analysis that created 8 clusters based on four instructor behaviors and four student behaviors. Their latent profile analysis revealed three primary instructional profiles, including Didactic (the majority of time is spent lecturing), Interactive Lecture (student-centered activities such as group work accompanying lecture), and Student Centered (more consistent use of group work and other student-centered interaction).

The current analysis is part of a larger study assessing the efficacy of an evaluation of teaching system that moves beyond student evaluations as a primary means of understanding teaching effectiveness. The goal of the current study is to use data gathered from peer observations of teaching using the COPUS instrument to identify instructional profiles that will be used to engage faculty in critical analysis of their about their teaching methods, successes, and shortcomings. This process is both evaluative and developmental.

II. RATIONALE

The limitations of end-of-semester students' evaluation of teaching (SETs) as a sole or primary assessment tool have been well documented in the literature on teaching and learning across disciplines. SETs as the primary means of assessing teaching and learning effectiveness have been critiqued on the basis of gender bias with research indicating males and females are evaluated differently (Anderson & Miller, 1997), inconsistent response rates with in-class evaluations higher than online (Dommeyer, Baum, Hanna, & Chapman, 2004), the effect of instructor appearance/attractiveness (see Gray & Bergmann, 2003; Riniolo, Johnson, Sherman, & Misso, 2006), inconsistency of item effectiveness and lack of item multidimensionality (see Marsh & Roche, 1997), racial bias (McPherson & Jewell, 2007), and the prejudicial effects of grading leniency, class size, and instructor enthusiasm (Saroyan & Amundsen, 2001). Wieman (2015) summarizes this research by arguing, "…The data indicate that it would be nearly impossible for a physically unattractive female instructor teaching a large required introductory physics course to receive as high an evaluation as that of an attractive male instructor teaching a small fourth-year elective course for physics majors, regardless of how well either teaches" (p. 9). Wieman goes on to suggest that this makes criteria for evaluating teaching effectiveness particularly challenging. Moreover, research on faculty perceptions suggests, that faculty do not necessarily change course content based on student evaluations and that those who receive higher SETs tend to endorse their validity more than those who receive lower SETs (Nasser & Fresco, 2002). The controversy surrounding and shortcomings of SETs puts department and university administrators in a difficult position when it comes to making decisions about merit and annual reviews, salary increases, and promotion and tenure, particularly because, despite their limitations, SETs are still a primary means of evaluating teaching in the face of such significant decisions. Not surprisingly, in a previous study focusing on the culture of teaching across 14 STEM departments at our host institution, we found student evaluations was the primary teacher evaluation strategy employed as a means to this end (**Authors**, 2016).

Based on these critiques and challenges, scholars have suggested the need to take a more multidimensional approach to the evaluation of teaching (Saroyan & Amundsen, 2001; Stake & Cisnersos-Cohernour, 2000). In response, Wieman and colleagues have developed a variety of instruments for evaluating teaching based on research in STEM education which center around the argument that, "the teaching methods used by an instructor are a more accurate proxy for teaching effectiveness than anything else that is practical to measure" (Wieman, 2015, p. 7). Specifically, Smith, et al. (2013) developed a new method for evaluating teaching effectiveness grounded in the philosophy of evidenced-based teaching practices. Developing this new method involved identifying all the teaching practices relevant to teaching STEM courses and practices that have been empirically linked to better student learning outcomes and course completion rates. The use of evidence-based teaching practices in a course, is used as a proxy for teaching effectiveness in this system that consists of three parts. First, faculty complete the Teaching Practices Inventory (Wieman & Gilbert, 2014), a self-report measure on the types of teaching practices used in their classroom. The TPI is comprised of eight categories including (1) course information provided, (2) supporting materials provided, (3) in-class features and activities, (4) assignments, (5) feedback and testing, (6) other (e.g., new methods, assessments), (7) the training and guidance of teaching assistants, and (8) collaboration (Wieman, 2015). Next, for the elements that are difficult for faculty to objectively self-report – in-class activities – the COPUS classroom protocol enables trained observers to identify the behaviors of instructors and students during a given class period (see Wieman, 2015). Finally, an EPT score ("extent of use of research-based teaching practices") is created in which "points are given for each practice for which there is research showing that the practice improves learning" (Weiman, 2015, p. 12). According to Wieman, this type of assessment allows teaching effectiveness to be evaluated according to the preferred standards of validity, meaningful comparisons, fairness, practicality, and improvement.

The use of the TPI and COPUS has been well documented and validated in extant research. For example, Smith, Vinson, Smith, Lewin, and Stetzer (2014) examined the teaching practices used in 51 STEM courses across 13 departments of a large research university. Their findings based on use of the TPI and COPUS observation protocol revealed four quadrants of behavior that ranged based on their percentage of codes identified by the researchers as presenting (e.g., lecturing), guiding (e.g., answering and asking questions), administration (e.g., making announcements), and other (e.g., waiting on students to complete a task). The quadrants demonstrated a continuum from active-engagement strategies to pure lecturing. Findings also indicate that instructors are fairly accurate in their self-assessment of the strategies they use. Smith et al. (2014) argue that these findings should be used to guide professional development teaching programs.

Stains and colleagues (2018) expanded on how data emerging from peer-observations with COPUS can enhance faculty's understanding of the overall instructional techniques they use in their classrooms. They observed over 2000 STEM class sessions covering seven STEM disciplines. These courses were taught by 548 faculty members across 24 doctorate-granting universities. On average, each faculty member was observed 3.2 times. These researchers recognized that in order to determine what strategies were implemented in conjunction with or as a replacement of one another, they could not solely rely on the frequency of student and instructor behaviors captured by COPUS. Thus, they conducted a latent profile analysis based on four instructor behaviors and four student behaviors that they found to be of adequate heterogeneity. As a result, three groups of instructional profiles emerged. The first, Didactic (clusters 1 and 2), represents classrooms with the prevalence of instructors lecturing 80% or more of class time with little student involvement. The second, Interactive Lectures (clusters 3 and 4), represents classrooms in which instructors have selected to supplement their lecture with student-centered strategies. The third and final grouping, Student-centered (clusters 5, 6, and 7) represents instructors that largely rely on and integrate student-centered strategies into their instruction.

Together, the TPI, COPUS, and instructor profiles offer a multidimensional method for assessing and discussing teaching effectiveness. Based on the shortcomings of teaching evaluations, the benefits of implementing evidenced-based practices in teaching, and the opportunity to use such teaching practices as a proxy for evaluating the effectiveness of teaching, we set out to test the feasibility and effectiveness of this method of evaluation in a College of Engineering at a large research-intensive Midwestern university.

III. PUPROSE & RESEARCH QUESTIONS

A. Purpose

The purpose of this project was to evaluate the implementation of new approach to the evaluation of teaching in a College of Engineering at a Midwestern research intensive institution. Traditionally, teaching is evaluated on the basis of the end-of-semester forms completed by students enrolled in a faculty member's course. However, the soundness of making personnel decisions, including hiring and firing, promotion, and tenure based primarily on students end-of-semester forms has been questioned for many years because of serious drawbacks in using these forms as a means to evaluate teaching (Neath, 1996). The College of Engineering at our host institution is in the process of adopting an approach to the evaluation of teaching that not only utilizes students' end-of-semester forms but that also utilizes classroom observations conducted by peers and doctoral students who have received training in the Classroom Observational Protocol for Undergraduate STEM or COPUS.

The significance of this project is to provide participating faculty members information and feedback on the participants' teaching as well as information and feedback about how the new approach to evaluating teaching in the College is working. This information and feedback can then be used to further improve upon the approach. Furthermore, faculty participants will be provided the opportunity to receive feedback based on their instructional profiles.

B. Research questions

The following research questions drove this analysis: (I) What types of instructional profiles emerge from the COPUS observation of engineering faculty? and (II) What are engineering faculty's experiences with the new system of teaching evaluation, including the self-report TPI, COPUS observation, and instructional profile analysis?

METHODOLOGY

This research employs an embedded mixed methods case study design with both qualitative and quantitative data collected and analyzed (Creswell & Plano Clark, 2017). We follow an instrumental case study design (Baxter & Jack, 2008; Creswell & Poth, 2018; Stake, 1995; Yin, 2014) with a quantitative strand. The case study incorporated an in-depth analysis of a bounded system, which in our study is the College of Engineering.

A. Participants

The participants in this study consisted of both observers and instructors. Faculty observers were recruited based on interest in the program, department chair approval and recommendation, and an established teaching record. Doctoral graduate student observers were recruited based on their acceptance into the College of Engineering Graduate Student Teaching Fellows program in the **(CENTER NAME REMOVED FOR BLIND REVIEW).** A total of 10 faculty peer observers and 6 graduate student observers were affiliated with the 7 engineering programs.

To recruit instructors, we distributed a college wide recruitment email seeking volunteers to all faculty in the college. The communication included information about the research study, requirements of participating as an observed instructor (teaching an undergraduate course, tenured or tenure-leading, and/or a professor of practice), reasons for conducting the research, and the procedures involved should they meet the requirements of the program. The 11 instructors were affiliated with six academic departments in the college.

B. Procedures

Selected observers were invited to an informational orientation session. The session covered the purpose of the study including a discussion on SETs, number of expected observations, required COPUS training, and post-program commitments. Following the orientation, faculty and graduate student observers attended a 1.5-hour training on using the COPUS. The training included participant introductions, exercise rationale, protocol and code introductions. Several guided practice opportunities were built into the training. Observers were presented with video lectures that they coded and discussed as a group and in pairs. Additionally, logistics of the program including gathering data, documentation and post-observation communications were discussed. After the observers were determined, faculty interested in participating as instructors were invited to an informational session. The session reiterated the items in the recruitment email and provided an opportunity for the faculty to ask questions about program expectations. Interested faculty indicated their interest by signing a consent form at the end of the session. Once the observers completed the training and the instructors were identified, the research team began the process of matching two paired observers with an instructor. The observers and instructors agreed on two class sessions to attend for a duration of 50 minutes each. Once the observations were complete, the observers submitted their completed COPUS protocols to the research team for analysis.

Following the observations, instructors were sent the Teaching Practices Inventory consisting of a 10-minute survey and were asked to self-report on the evidence-based practices they use as it relates to the course that was observed. Soon after, instructors were invited to an exit interview comprising of two parts. First, the instructors met with a research team member to discuss the results of their reported Teaching Practices inventory, observations using COPUS, and instructional profile analysis. In this meeting, instructors were presented with a personalized report that contains information about their observed class. Then, on a separate occasion, they met with another research team member to discuss their experiences with, or perceptions of, the new system of teaching evaluation.

C. Data gathering and analysis

Three instruments were used to gather data for this study. First, the COPUS was used to gather quantitative observational data and qualitative data in the form of the observer comments. Each instructor was observed twice, 50 minutes in each observation. The protocol captures data on student and instructor behaviors, as well as student engagement in two-minute intervals. Observed data includes behaviours such as students listening, asking questions, taking a test, and instructors demonstrating, presenting, lecturing, asking questions among others for a total of 25 recorded behaviours. The observational data analysis result in an inter-observer Kappa reliability value and a heat map (see Figure. 1) representing how time was spent in the observed sessions. The data is additionally used to determine the instructional profile. The second instrument is the Teaching Practices Inventory (TPI). The TPI contains instructor self-reported instructional strategies and can be quantitatively scored. Sample items include: supporting materials, course information, in-class features, in-class discussion and many others. Each category in the TPI is scored following a rubric developed by its creators. The rubric consists of a point system representing each of the items on the TPI The analysis of the TPI gauges the extent of use of research-based instructional practices, such that higher scores…[specify here and then we should also include means and standard deviations for our sample]. Third, we used an interview protocol to conduct semi-structured interviews that gathered qualitative data via audio recording about observed instructors perceptions of the TPI, COPUS observations, and instructional profiles and exit interview. Sample interview questions included open-ended questions about observed faculty members' assessments of the benefits, challenges, and potential barriers of the new system as well as closed-ended questions about its usefulness. Instructors were recorded as they were interviewed. The recordings were then transcribed and coded. The results section of this study highlights the themes that emerged.

RESULTS

A. Research Question 1

Research question 1 asked about the types of instructional profiles that emerged in the observations and analysis of engineering faculty members' teaching. The results of the inter-rater reliability Kappa analysis showed strong agreement among observers (kappa = .79). The high inter-rater reliability is a strong indicator of the validity of the COPUS tool and training.

Our profile analysis shows high Didactic teaching among Engineering faculty with a total of 72% falling in Clusters 1 or 2. Specifically, 29% of faculty falling in cluster 1. Cluster 1 instruction is described as over 80% of class time spent on lecturing with minimal student engagement that is reduced to sporadic questions to and from the students. Additionally, 43% of faculty fall in cluster 2. Cluster 2 instruction is described as having clicker questions that are sometimes associated with group work. With regards to the Interactive Lecture profile, 14% of faculty were identified by Cluster 3 or 4. Specifically, 7% of faculty fell into Cluster 3 and 7% in Cluster 4. Cluster 3 represents the instructors that supplement their lectures with "Other group work" and student-centered strategies, whereas Cluster 4 represents instructors that supplement their lectures with clicker questions associated with group work and student-centered strategies. In the final group, Student-Centered, none of the instructors fell into Clusters 5 and 6, and 7% fell into Cluster 7 that is represented when an instructor applies student-centered strategies with a variety of group work that is less consistent in terms of usage.

A thorough and systematic review the heat-maps for each instructor show consistency between the recorded instructor and student behaviors and the clusters they fell into. As in the example presented in Figure 1, the instructor spent 100% of the class time lecturing and presenting and the students spent 100% receiving information and listening. Instructor 8 fell into the Didactic – Cluster 1 instructional profile as we would expect.

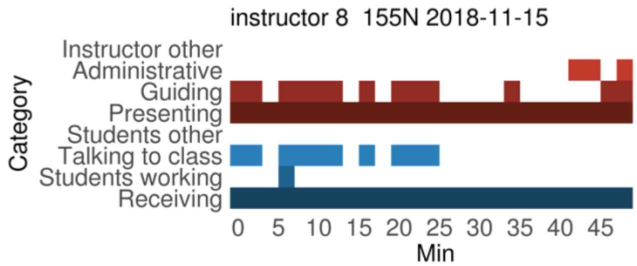

Fig. 1 Example of class session heat map instructor 8 falling in cluster 1

On the contrary, instructor 6 fell into cluster 7. When examining the heat map in Figure 2, the instructor has period of time without lecturing and time guiding the students. The students have periods of time where they are engaged in group work and answering questions. Engagement behaviors are evident through the observation.

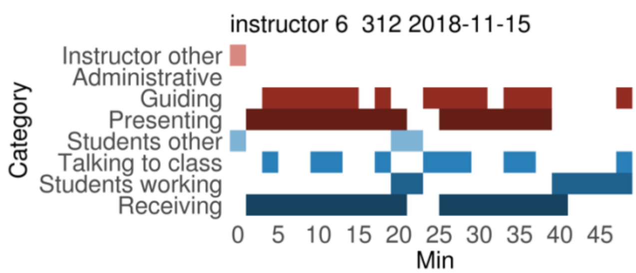

Fig. 2 Example of class session heat map representing instructor 6 falling in Cluster 7.

In short, and in response to Research Question 1, the analysis of heat maps and the latent profile analysis shows that the majority of instructors observed engaged in Didactic Teaching methods.

B. Research Question 2

Research question 2 asked about faculty members' experiences with, or perceptions of, the new system of teaching evaluation, including the self-report TPI, COPUS observation, and teaching profile analysis discussed in their exit interviews. The thematic analysis of participants final in-depth interviews revealed themes within the categories of their perceptions on the benefits, challenges, potential barriers to adoption, and recommendation for future implementation. Themes that emerged included:

Benefits. When asked to reflect on the benefits of the new system for evaluating teaching effectiveness, participants identified four primary themes: reflection, unbiased, systematic, and non-threatening.

First, participants articulated an appreciation for the opportunity to gain knowledge about and **reflect** on their teaching. For example, one participant said, "Just setting aside time to reflect on [teaching] helped" (Participant 1).

This participant also described discovering the teaching profiles as "extremely eye opening." Another participant said "It is good to know what is going on in the classroom." Finally, another participant expressed gratefulness to bring his teaching into focus:

> *It was very helpful in inspiring, "Ok, yes! that's something I wasn't thinking about." I have certainly thought about it in the past. Its' taking a whole collection of things and say[ing] "Ok here's kind of a universe of things…if I think about all of these things instead of just one. I remember that! I remember that! I remember wanting to do that! Yes, I would like to be able to do this better." It just kind of pools all this stuff (Jennifer).*

Part of this benefit was enabled by **unbiased** observation. The combination of self-reporting on one's teaching practices using the TPI and seeing observer's neutral reports on the COPUS instrument empowered participants. For example, one participant said, "I think it…helps when you're not framing the questions. Because you frame questions typically around your hidden bias or your blind spots so when you have questions framed by a peer it helps you think outside your comfort zone" (Participant 8) Another participant appreciated the "unbiased feedback on how time is being spent" (Participant 6).

Faculty also liked the **systematic** nature of the process, noting is structure, nuance, and objectivity as benefits. One person said:

> *I like that there is a structure to it. And that just might be the engineer in me. I think it's nice to have people trained on the same procedure and also have certain structure and cadence… to evaluate as opposed to having someone do more subjective…which depends on their personality and their experience of teaching…[This] gets a little more fair" (Participant 1).*

Another participant noted that the process was more nuanced than SETs, while a third participant likened the process to a mathematical formula: "Sometimes you kind of forget yourself what you're doing but this will give you like a mathematical [formula] to see how I'm doing this, more or less" (Participant 2).

Finally, participants appreciated the **non-threatening** nature of the system, noting that observers were inconspicuous and easy to forget. Another said, "It's non-threatening. People are simply observing characteristics: here's what's happening in the classroom" (Participant 8).

Challenges. The main theme that emerged when faculty identified drawbacks or challenges of the new system was the **acontextuality** of the results. This emerged as a need for more context and more detail both about the instructional profiles and how to interpret the results of the TPI, COPUS observation, and profile in light of discipline and course diversity.

First, participants placed value on the instructional profiles, but felt they did not have enough information to interpret them. During their exit interview, researchers shared with faculty members the profiles that emerged from the cluster analysis and discussed the findings from the TPI and COPUS observations. They were also given references to articles on Stains et al.'s (2018) profile analysis for more information on each profile. Faculty clearly placed value on the clusters, but longed for more detail.

For example, one faculty member said,"

> *[It was] nice to know I wasn't in cluster 1 or 2, but how to interpret…?...I don't know that I want every class period to be cluster 7….[It's] not clear yet on the differences between the profiles other than*

student centered is better than interactive or didactic. I'm not exactly sure how to interpret these…and some of it, I'm not sure if every class is student-centered (Participant 6).

It was clear that this faculty member placed value on, but grappled with the profiles.

The idea that not all classes should conform to a particular profile was echoed in the second subtheme of acontextuality: the idea that different class types and different disciplines may need to be taught differently and that the profiles, therefore, must be interpreted within those contexts. Faculty noted that things may be classified differently in different fields (Participant 4), that two class periods was not enough to give an accurate picture (Participant 2) because different class period and weeks would differ (Participant 6, 8). These differences emerged in one faculty member's interview:

It's a good method but I mean the limitation I see does not capture…the variation between courses, you know, some courses are hardcore like engineering or some other so… if there's a course, where it's just more an information course, there's a lot… it will show up a good score. But if it's just like it dry materials course… you know what I'm saying. So to answer your question, I think it should be different version of it you know I'm saying that like not one size fits all" (Participant 2).

Similarly, another faculty member noted the limitations of acontextuality when stating:

It may be a lecture the full time, but if that puts you in cluster 1, 2 or 3, I don't think that makes you a bad teacher…then it becomes as ineffective as just going with student evaluations because you don't know if students are leaving with good quality knowledge…Just because every two minutes I'm talking or they're talking or they're engaged doesn't fully say, you know, how well they're actually learning the material (Participant 6).

In short, although all faculty felt the process was helpful or very helpful, they articulated the need to modify the system to be more contextual and nuanced for different courses and circumstances.

Barriers. The main potential barriers for implementing the new system into the College of Engineering included concerns about **resources** and the **potential resistance** of other faculty to receiving feedback on teaching. First, when asked what barriers they saw to implementing the system, faculty talked about time. They noted that faculty are already so busy, they may not have time to commit to multiple observations. They also discussed the potential difficulty in getting enough trained observers for this reason. At least one faculty member noted the potential for scarce resources to support such a program, particularly if administrators change or do not buy in to the program.

Several faculty members also noted that while they themselves appreciated feedback, they worried that other faculty members would resist. For example, one foresaw potential challenges in faculty accepting feedback, noting, "I'm open to receiving feedback, but I imagine more seasoned folk might want to see some history behind it" (Participant 1). Another person talked about the risk of feeling vulnerable to the process:

I don't have a problem with that, but I can see other people, they might … feel this will expose, expose …what's wrong about what they do the course… (Participant 2).

Interestingly, all participants who discussed this potential barrier acknowledge that they themselves appreciated feedback and simply worried that others might be more closed to the process.

RECOMMENDATIONS

In light of the benefits, challenges, and potential barriers, faculty recommended a number of supplements to the new system, including:

1. Pairing the system with a community of scholars. This could include having discussion groups interesting in sharing ideas about teaching and their profile results. It could also include pairing this process with another faculty development program that first exposes faculty to evidence-based teaching strategies before evaluating them according to the TPI and COPUS items. In this way, faculty would have the chance to change their courses prior to COPUS implementation and then test their strategies and teaching practices.
2. Consider supplementing the in-person observations with videotaped records so faculty can go back and watch their own teaching practices in light of their profiles. One participant suggested that videotapes could be used in lieu of in-person observers to cope with barriers of resources, but another argued that you can only get the vibe of a class and know if it is a learning environment by being there in person.
3. Training and history on the process for faculty and administrators. Along with this, faculty believed it should be used in concert with teaching evaluations rather than in lieu of them.

In sum, faculty found the new process helpful, effective, and desirable, and offered many suggestions for ways to improve its feasibility in the College of Engineering.

CONCLUSION

Contemporary research underscores the need for implementing empirically based instructional strategies in STEM education classrooms that engage students more fully in their own learning. As a result, increasing attention has been devoted to identify and document pedagogical strategies currently employed by STEM educators to maximize student achievement and to provide alternative means for assessing teaching effectiveness. Along these lines, several peer observation protocols have been developed to better document teaching and learning such as the Classroom Observation Protocol for Undergraduate Education (COPUS), a tool that has terrific potential for contributing to our knowledge of instructional practices.

The current analysis is part of a larger, longitudinal study geared at assessing teaching effectiveness at large research-intensive Midwestern institution to ultimately contribute to our broader goals both to improve teaching and evaluation strategies in a College of Engineering. Research questions guiding this inquiry included (1) what types of instructional profiles represent engineering faculty?, and (2) What are engineering faculty's experiences with the new system of teaching evaluation, including a self-report TPI (discussed above), COPUS observation, and teaching profile analysis. Multiple peer observations by faculty and graduate student paired teams of undergraduate classes within the College were conducted. These were followed by exit interviews that discussed results of the TPI, COPUS, and instructional profiles with participant observers and instructors. Findings revealed salient themes focusing on the benefits, drawbacks or challenges, potential barriers to adoption, and recommendations for future implementation of this new multifaceted system of teaching evaluation.

Of particular note, the findings of the current pilot study underscore at least two fundamental needs moving forward. First, as revealed in the analysis of observed faculty interviews, in order to make the results of the new system meaningful, faculty need more information about the evaluative nature of their profiles. Research shows that student engagement is key to evidence-based practices of student learning (Fairweather, K. (2008)), which would position Interactive Lecture and Student-Led profiles as preferable to Didactic. Faculty interviews showed that they placed values on the clusters and the results of the profile analysis but that they needed more context and information to interpret their profiles and how to improve. Thus, in order to make the new system viable and meaningful, it needs to be contextually evaluative for instructors. Second, although the interactive and student-led profiles may be preferable in terms of evidence-based practices, the majority of participants in the current study fell into the Didactic profile. Our sample size, given the nature of the pilot, is very small. Thus, additional research is needed on the prevalence of profiles, but future use of the COPUS, TPI, and profile analysis would be strengthened by additional information about how to increase the use of evidence-based teacher and student practices in the classroom. For example, one participant suggested pairing the process with professional teaching development programs or cohorts to support teaching improvement.

Overall, these findings underscored both strengths of this process and ways of improving its feasibility within the College of Engineering, which we believe can be of value in other colleges of engineering and STEM disciplines as we work on our shared goal of improving teaching and learning in the classroom.

This paper is based upon work supported by the National Science Foundation under Grant No.1347814. Any opinion, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Andersen, K., & Miller, E. D, "Gender and student evaluations of teaching". PS: Political science & politics, vol. 30(02), 1997, pp.216-219.

[2] Baxter, P., & Jack, S., "Qualitative case study methodology: Study design and implementation for novice researchers. The Qualitative Report", vol. 13(4), pp. 544–559, 2008

[3] Creswell, J. W., & Poth, C.N.. "Qualitative inquiry and research design: Choosing among five approaches" (4th ed.). Thousand Oaks, CA: Sage, 2018.

[4] Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S,"Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations," Assessment & Evaluation in Higher Education, vol. 29(5), pp. 611-623, 2004.

[5] Gray, M., & Bergmann, B. R, "Student teaching evaluations," Academe, vol. 89(5), pp. 44-46, 2003.

[6] Hora, M. T., & Ferrare, J. J. "Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching," Journal of the Learning Sciences, vol. 22(2), pp. 212-257, 2013.

[7] Kim, H., Babchuk, W.A., Heaton, R.M., & Perez, L.C, "At the STEM of the problem: A qualitative case study of department chairs' perspectives on the culture of teaching. Proceedings of the 35st Annual Research-to-Practice Conference in Adult and Higher Education," University of Central Oklahoma. Albuquerque, New Mexico, 2016.

[8] G. L. Connell, D. A. Donovan, T. G. Chambers, "Increasing the Use of Student-Centered Pedagogies from Moderate to High Improves Student Learning and Attitudes about Biology," CBE Life Sci Educ vol. **15**, 2016.

[9] Fairweather, James. "Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education." Board of Science Education, National Research Council, The National Academies, Washington, DC (2008).

[10] Marsh, H. W., & Roche, L. A, "Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility," American psychologist, vol. 52(11), pp. 1187, 1997.

[11] McPherson, M. A., & Jewell, R. T. "Leveling the playing field: Should student evaluation scores be adjusted?," Social Science Quarterly, vol. 88(3), pp. 868-881, 2007.

[12] M. K. Smith, E. L. Vinson, J. A. Smith, J. D. Lewin, M. R. Stetzer, "A campus-wide study of STEM courses: new perspectives on teaching practices and perceptions," CBE Life Sci. Educ. vol. 13, pp. 624, 2014.

[13] Nasser, F., & Fresko, B, "Faculty views of student evaluation of college teaching," Assessment & Evaluation in Higher Education, vol. 27(2), pp. 187-198, 2002.

[14] Neath, Ian., "How to improve your teaching evaluations without improving your teaching." Psychological Reports 78, vol. 3, pp. 1363-1372, 1996.

[15] Riniolo, T. C., Johnson, K. C., Sherman, T. R., & Misso, J. A, "Hot or not: Do professors perceived as physically attractive receive higher student evaluations?," The Journal of General Psychology, vol. 133(1), pp. 19-35, 2006.

[16] Saroyan, A., & Amundsen, C, "Evaluating university teaching: Time to take stock," Assessment & evaluation in higher education, vol. 26(4), pp. 341-353, 2001.

[17] Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I, "Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol," School science and mathematics, vol. 102(6), pp. 245-253, 2002.

[18] Smith, K. A., Sheppard, S. D., Johnson, D. W., & Johnson, R. T, "Pedagogies of engagement: Classroom-based practices," Journal of engineering education, vol. 94(1), pp. 87-101, 2005.

[19] Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E, "The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices," CBE-Life Sciences Education, vol. 12(4), pp. 618-627, 2013.

[20] Stains, M., J. Harshman, M.K. Barker, S.V. Chasteen, R. Cole, S.E. DeChenne-Peters, M.K. Eagan, J.M. Esson, J.K. Knight, M. Laski, et al., "Anatomy of STEM Teaching in North American Universities," Science vol. 359 (6383), pp. 1468–1470, 2018.

[21] Stains, M., & Vickrey, T, "Fidelity of implementation: An overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. CBE—Life Sciences Education, vol. 16(1), rm1, 2017.

[22] Stake, R.. The art of case study research. Thousand Oaks, CA: Sage, 1995.

[23] Stake, R. E., & Cisneros-Cohernour, E. J,. Situational evaluation of teaching on campus. New directions for teaching and learning, vol. 2000(83), pp. 51-72, 2000.

[24] T. J. Lund et al., "The best of both worlds: building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. CBE Life Sci. Educ. vol. 14, ar18, 2015.

[25] Wieman, C., & Gilbert, S. The Teaching Practices Inventory: a new tool for characterizing college and university teaching in mathematics and science. CBE—Life Sciences Education, vol. 13(3), pp. 552-569, 2014.

[26] Wieman, C. A better way to evaluate undergraduate teaching. Change: The magazine of higher learning, vol. 47(1), pp. 6-15, 2015.

[27] Yin, R. K.,"Case study research: Design and methods" (5th ed.). Thousand Oaks, CA:Sage, 2014.