International Journal of Engineering Education Vol. 34, No. 4, pp. 1325–1334, 2018
Printed in Great Britain

0949-149X/91 $3.00+0.00

The Long-Term Impact of a Faculty Development Program on Student Evaluations of Teaching*

MATT DEMONBRUN
School of Education, University of Michigan, 610 E. University Avenue, Room 2117, Ann Arbor, MI 48109, USA.
E-mail: mdemonbr@umich.edu

JULIA KERST, HANNA PFERSHY and CYNTHIA J. FINELLI
College of Engineering, University of Michigan, 4227 EECS Building, 1301 Beal Avenue, Ann Arbor, MI 48109, USA.
E-mail: jfkerst@umich.edu, hpfersh@umich.edu, cfinelli@umich.edu

Although participation in faculty development programs has been linked to improvements in teaching and positive outcomes in student learning, prior research has yet to uncover how these outcomes might lead to increases in student evaluation of teaching scores. Student evaluation of teaching scores are frequently used in tenure and promotion decisions, and prior research has found that these scores often decline over time. Thus, it is important to better understand how faculty development programs might work to improve these scores. In this study, we examine whether student evaluations of teaching scores do decline over time, and the potential impact of a faculty development program on these evaluations.

Keywords: student evaluations of teaching; faculty development programs; longitudinal studies; quasi-experimental research

1. Introduction

Faculty development programs have long been used by colleges and universities to improve the skills necessary to be more effective instructors in academia [1, 2]. Findings from research on faculty development programs have indicated a number of positive outcomes from their use including—but not limited to—improvements in teaching that are tied to enhanced student learning and achievement in the college classroom [2, 3]. For example, in a large-scale faculty development program focused on improving students' writing skills across disciplines, researchers found that the average student's writing scores significantly improved for those faculty who participated in the program [2].

Although students may benefit from improved teaching encouraged by faculty development programs, less is known about how these improvements in teaching and learning may translate to increases in student evaluation of teaching scores. In other words, is there a significant positive relationship between participation in a faculty development program and improved scores on student evaluations of teaching? While some researchers have reported small gains in student evaluation of teaching scores after participating in faculty development programs, a majority of the research has not investigated how these scores compare to faculty who do not participate in these programs and how these scores might change over time. This research study attempts to address this gap in the literature by examining the benefits of one such program, aimed at the widespread implementation of evidence-based teaching practices (EBTP), on faculty in the College of Engineering at a large public research university.

1.1 The Teaching Circle for Large Engineering Courses

Recent reports have called for the increased dissemination of EBTP in college classrooms across the United States, particularly in the field of engineering. Reports from the President's Council of Advisors on Science and Technology, the American Society of Engineering Education, and the National Academy of Engineering have all discussed the importance of the use of EBTP in the classroom to prepare future engineers for the workforce [4–6]. These reports point to the benefits of these practices in promoting retention and success in engineering courses [7–11] and the effectiveness of them in educating an increasingly diverse student population, particularly for women and students of color [10, 12]. EBTP that have been shown to positively impact student learning include establishing rapport between students and the instructor [13, 14], using real-world examples [15–18], identifying and addressing common misconceptions [19], and incorporating a student-centered approach to teaching [20–22].

To promote adoption of EBTP by engineering instructors at the University of Michigan, faculty developers at the university's Center for Research on Learning and Teaching in Engineering designed the *Teaching Circle for Large Engineering Courses* [23, 24]. The *Teaching Circle* is designed to leverage many of the factors known to motivate instructors

* Accepted 14 March 2018.

to adopt EBTP while lowering some of the barriers [25]. The program provides an opportunity for instructors to network and learn from others, offers personalized support during the process of change, and features a safe environment where instructors can discuss successes and failures.

The *Teaching Circle* is a faculty learning community [26], where faculty meet to discuss evidence-based teaching and identify practical strategies for success during four monthly, two-hour meetings over one term. The sessions are co-facilitated by a professional faculty developer and an engineering professor with experience in utilizing EBTP. Topics of the four sessions include: building rapport to make a large class more welcoming, integrating active learning into the classroom, understanding and influencing student motivation, overcoming student mis-/pre-conceptions about course content, and implementing instructional technology in the classroom (e.g., screencasts and classroom response systems). During each session, faculty discuss readings that summarize recent research on each topic with other faculty colleagues and facilitators. These sessions also allow for faculty to interact with the program facilitators, each other, and senior faculty to share best practices and experiences with EBTP implemented in their classrooms.

Outside of the four monthly meetings, faculty selected for the *Teaching Circle* are expected to participate in a midterm student feedback session [27] in one of their current courses to evaluate how participation might impact their efforts to implement EBTP. Participants also receive monetary support—upon completion of the program, the University of Michigan College of Engineering provides funds of $1,000 per participant, which instructors can use to support their teaching efforts. The *Teaching Circle* was piloted during the Fall 2011 term, was fully implemented during the Winter 2012 term, and has been offered once or twice every year since then. Interested instructors apply to the program and, due to space and funding constraints, six or seven instructors participate in the program each term, selected to ensure diversity of gender, faculty rank, and discipline within each cohort.

Evidence from the first two years of the program showed that it positively influenced participants' teaching attitudes and behaviors in two ways [23, 24]. First, as measured by the Approaches to Teaching Inventory [28], participants in the *Teaching Circle* showed a slight increase in their self-reported use of student-focused teaching behaviors when compared to colleagues who applied to the program but did not participate. Second, as measured by the Teaching Behaviors Inventory [29], participants demonstrated significant increases in three of six self-reported constructs used to evaluate effective teaching when compared to peers in the control group: the use of nonverbal behavior to seek student interest and engagement (enthusiasm), explanation of concepts and principles in learning material (clarity), and strategies to encourage student participation in class (interaction) [24]. Additionally, instructors' use of different ways to organize subject material for the course (task organization) significantly increased in terms that followed their participation in the *Teaching Circle*.

While the *Teaching Circle* was specifically designed to overcome some of the barriers to adopting EBTP (e.g., providing easy-to-understand syntheses of the research literature, decreasing the time required to learn about EBTP, and establishing an environment for instructors to network with and support one another), its effect on other barriers has yet to be explored. In particular, one significant barrier is instructors' concerns that adoption of EBTP could result in lower student evaluation of teaching scores, but little is known about the relationship between a faculty development program like the *Teaching Circle* and student evaluation of teaching scores. Thus, this study examines whether student evaluations of teaching scores do decline over time, and the potential impact of a faculty development program on these evaluations.

1.2 Student evaluation of teaching scores

There is some merit to the use of student evaluation of teaching scores. Historically, many scholars have used student evaluation of teaching scores at the university level to measure student satisfaction with self-reported learning in the course [30–33]. But, an increasing amount of research indicates that student evaluation scores may not be the best proxy for course effectiveness. For example, conclusions drawn from analyses of student evaluation scores can be biased by course size [34–36], grading expectations and leniency [37–39], instructor characteristics [40–46], and years of teaching experience and academic rank [47–49].

Research examining student evaluations of teaching longitudinally suggests that scores tend to gradually decline over time [50] or remain stagnant [51, 52] as years of teaching increase. Others have suggested that these results vary by academic department, with faculty in some departments demonstrating notable increases in their scores while other departments experienced decreases [50]. One study [51] found that evaluation scores declined for all instructors except for those who began their careers with much lower scores than their peers, and even then, those instructors' evaluation scores only slightly improved to levels still far below their peers.

Regardless of the merit in using student evaluation of teaching scores, administrators and other instructors routinely use these evaluation scores for decisions about merit increases and tenure and promotion decisions [53, 54] due to the simplicity and ease of use for assessing perceptions of student satisfaction with the course and instructor [45]. Student evaluation of teaching scores are often the most commonly-used method to evaluate teaching effectiveness, particularly for faculty tenure and promotion packages [55–58], they can play a direct role in the selection criteria for teaching awards [57, 59], and administrators often use them in decisions on how to improve the quality of teaching in a program or course [60–63].

Given that evaluation scores are used frequently in tenure and promotion decisions [53, 54] and that instructors have concerns that using new teaching practices, particularly those that feature EBTP, will result in lower student evaluation of teaching scores [20, 22, 23, 25–29], it is easy to understand how this could be a barrier to the implementation of such practices. Additional research is needed to determine whether such decreases, if they do exist, can be lessened or reversed by participating in a faculty development program. Thus, this research seeks to answer the questions:

Do student evaluation of teaching scores decline over time, as suggested by the literature? How does participation in a specially-designed faculty development program focused on implementing EBTP mitigate these declines?

2. Methods

2.1 Sample

Our sample consists of all instructors in the University of Michigan College of Engineering who applied to participate in the *Teaching Circle*; as of the winter term in 2016, there were 71 such instructors. Forty-one of the 71 instructors applied and participated in the *Teaching Circle*, with 10 applying at least twice before participating in the *Teaching Circle*, and 30 applied but were not invited to participate due to program constraints. The 41 instructors who participated in the *Teaching Circle* comprise the intervention group for our analysis, while the other 30 instructors who did not participate comprise the control group.

Data regarding the gender, department, and rank of instructors in the intervention and control groups are provided in Table 1. Note that the program intentionally included one or two female instructors in each cohort (so 63% of the female applicants were selected to participate in the program, but only 55% of the male applicants were selected). In addition, the program was piloted with faculty in Chemical Engineering, but later cohorts were selected to balance the composition of participants across each engineering department. A purposeful effort was also made to support assistant professors, which led to an overrepresentation of this rank in the intervention group, and although many non-tenure track faculty (Lecturers) applied to participate, because the program was designed to support tenured/tenure-track faculty, applications from tenured and tenure-track faculty were given priority.

2.2 Measures

To investigate the potential decline in student evaluation of teaching scores over time and the impact of participating in a faculty development program on these evaluations, we studied end-of-term student evaluation scores for instructors in both the intervention and control groups of our sample. Because the first four items on the University of Michigan's student evaluations of teaching instrument are required for every course, and because these four items are key metrics about teaching used for promotion and tenure, we studied scores to those four items (see Table 2).

The emphasis of the *Teaching Circle* was on improving instruction in large, lecture-based undergraduate engineering courses, so we compiled student evaluation scores for every undergraduate, lecture-based, engineering course that was taught by one of the 71 instructors in our sample from Fall

Table 1. Teaching Circle Demographics

	Intervention (N = 41)	Control (N = 30)
Gender		
Female	14	8
Male	27	22
Department		
Aerospace Engineering	2	1
Atmospheric, Oceanic, and Space Science	2	3
Biomedical Engineering	4	1
Civil and Environmental Engineering	4	2
Chemical Engineering	7	1
Computer Science Engineering	5	5
Electrical and Computer Engineering	1	3
Industrial and Operations Engineering	3	2
Mechanical Engineering	9	7
Material Science Engineering	1	2
Technical Communication	3	3
Rank		
Lecturer	5	15
Assistant Professor	18	2
Associate Professor	5	2
Professor	13	11

Table 2. Items from the Student Evaluations of Teaching Instrument[a]

Item Text	Abbreviation
Overall, this was an excellent course.	Was excellent course
Overall, the instructor was an excellent teacher.	Was excellent teacher
I learned a great deal from this course.	Learned a great deal
I had a strong desire to take this course.	Had desire to take course

[a] Responses based on a 1–5 Likert scale, where 1 = strongly disagree and 5 = strongly agree.

2008 through Winter 2016. We excluded all laboratory courses, independent study courses, small-group multidisciplinary design courses, courses having fewer than five students, and graduate level courses.

Each term (x_T) was assigned an integer value representing the time period relative to the application term. Thus, $x_T = 0$ is the term in which instructors first applied to participate in (control group) or did participate in (intervention group) the *Teaching Circle*. The term immediately prior to the application term is $x_T = -1$, while the term immediately following the application term is $x_T = 1$, and so on. Terms in which an instructor did not teach were ignored. For our sample of 71 instructors, the term value ranged from –10 to 8. In other words, we have data for as many as 10 semesters before and 8 semesters after the application term.

For each instructor and each of the four student evaluation items, we computed the average evaluation score for every term in which the instructor taught. (If an instructor taught more than one course in a given term, we computed a weighted average of scores based on the number of student enrolled in the course.) Overall, our data set contains 769 total sets of evaluation data, where one set of evaluation data represents one instructor's average scores for one semester. When examined longitudinally, the number of terms taught per instructor ranged from 7 to 18 with a mean/standard deviation of 10.9 ± 3.5.

2.3 Analysis

We applied a Difference-in-Differences (DID) model [64] to compare the long-term impacts of the *Teaching Circle* on student evaluation of teaching scores. The DID model is a quasi-experimental design that utilizes longitudinal data from intervention and control groups to estimate a causal effect of a treatment through the use of a counterfactual. It is typically used to estimate the impact of a specific intervention at a point in time. In our case, we include three independent variables: (1) x_T, a continuous variable which models the term relative to application to the *Teaching Circle*, (the intervention point is $x_T = 0$), (2), x_{Int}, a dichotomous variable indicating whether or not an instructor was in the intervention group, and (3) x_{T*Int}, the DID coefficient (represented as an interaction between the time and intervention variables). Thus, our model in this study is represented by Equation (1):

$$y = \beta_1 x_T + \beta_2 x_{Int} + \beta_3 x_{T*Int} + \alpha + \varepsilon \qquad (1)$$

The coefficients on the right-hand side of Equation (1) represent the slope of the changes in scores across time for all instructors (β_1), the difference between the intercept scores (i.e., scores at term 0) for the intervention and control groups (β_2), the slope of the changes in scores across time for the intervention group after participation in the *Teaching Circle* (β_3), and the baseline average score for the control group (α). Epsilon (ε) is the error term which

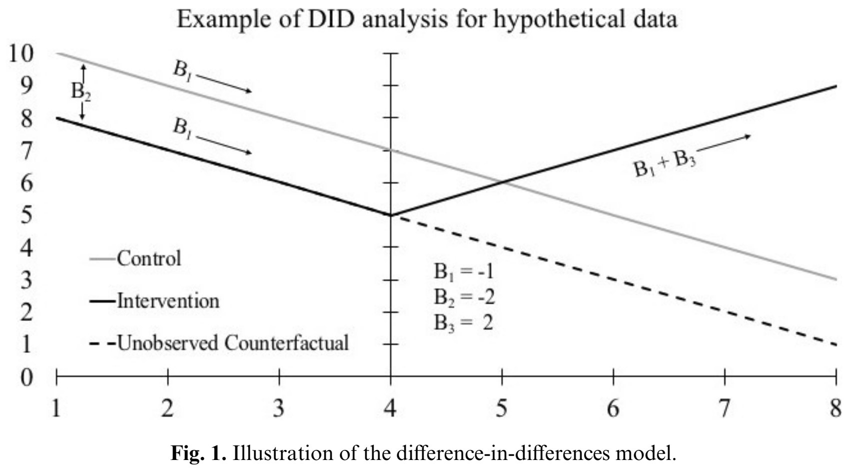

Fig. 1. Illustration of the difference-in-differences model.

represents the random error component not explained by the model. An example of how these coefficients are represented visually in the difference-in-differences model is presented in Fig. 1.

As illustrated in the hypothetical example of Fig. 1, β_1 represents the slope for the changes in scores across time for all instructors. For example, in Fig. 1, the scores for both the control and intervention groups (before the intervention) decline at a rate of approximately 1 point per time period (i.e., $\beta_1 = -1$). After the intervention point, this decline continues for the control group, and the declines that would be experienced by the intervention group are represented by the unobserved counterfactual (dashed line). β_2 represents the differences in the intercepts between the intervention and control groups. In Fig. 1, we can see that the scores at the initial time period start at 10 for the control group and 8 for the intervention group, a two point difference in scores for these groups (i.e., $\beta_2 = -2$). β_3 represents the differences in the slopes between the intervention and control groups *after* the intervention point. This coefficient is added to β_1 after the intervention point since β_1 represents the slope of change for all instructors, but β_3 represents the differences in these slopes for the intervention and control groups (i.e., $\beta_1 + \beta_3$ = one-point increase *per time period* for the intervention group after the intervention point).

2.4 Results

Before constructing estimates from out data for the DID model, we conducted a chi-squared test of differences between student evaluation scores of the intervention and control groups for the time prior the *Teaching Circle* to confirm whether our data satisfied the "parallel trends assumption" [65]. This assumption is a requirement of DID model to ensure that the trends of student evaluations for both the intervention and control groups are similar during the pre-treatment period. If the assumption is valid, one can confirm that changes between intervention and control groups after the treatment are a result of the treatment. The chi-squared test indicated that, for each of the four student evaluation items, the differences in trends over time before the intervention were nonsignificant ($p > 0.05$), so the parallel trends assumption was valid. These analyses indicate that there is no difference in pre-intervention teaching evaluation scores for instructors in our control and intervention groups despite differences in the overall composition of the groups. Thus, we turned our focus to the differences in the teaching evaluation scores between these two group after the intervention, rather than examining differences by demographic characteristics (e.g., race, gender, academic rank). To compare changes in student evaluation scores over time for the intervention and control groups, we estimated four separate DID models (one model for each of the four student evaluation items). We also clustered our standard errors for each panel unit (i.e., each faculty member in the dataset) to protect against autocorrelation and heteroscedasticity. The coefficients from the models, as represented in Equation (1), are presented in Table 3.

The β_1 coefficient (the slope of the changes in scores across time for all instructors) demonstrates that student evaluation scores decreased significantly over time for all four of the items. Specifically, instructors in the intervention and control groups experienced declines in their scores of 0.02 points/term for three of the items (*Was excellent course*, *Was excellent teacher*, and *Learned a great deal*; $p < 0.001$) and 0.03 points/term for the fourth item (*Had desire to take course*; $p < 0.05$). Thus, an instructor who started teaching ten terms before applying for the *Teaching Circle* would experience a predicted decline of 0.2 points for these first three items and 0.3 points for the fourth item by the time they applied for the *Teaching Circle*. Barring any intervention, these declines of 0.02 or 0.03 points/term would be expected to continue through the most recent term in our study (Term 8).

The β_2 coefficient (the difference between the intercept scores for the intervention and control groups) demonstrates that there were no significant differences in the student evaluation scores at $x_T = 0$ for the intervention and control groups, so there are no statistically observed differences in the teaching scores of these two groups leading up to the intervention. The β_3 coefficient (the slope of the changes in scores across time for the intervention group after participation in the *Teaching Circle*) does illustrate significant differences between the pre- and post-intervention slopes of all four student evaluation items for instructors in the intervention

Table 3. Results from Difference-in-Differences Analysis Across Items

	β_1 (T)	β_2 (Int)	β_3 (T x Int)	α (Constant)
Was excellent course	-0.02^{***}	–0.01	0.02^{**}	4.06
Was excellent teacher	-0.02^{***}	–0.02	0.02^{**}	4.25
Learned a great deal	-0.02^{***}	–0.01	0.03^{***}	4.22
Had desire to take course	-0.03^{*}	0.08	0.03^{*}	3.86

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

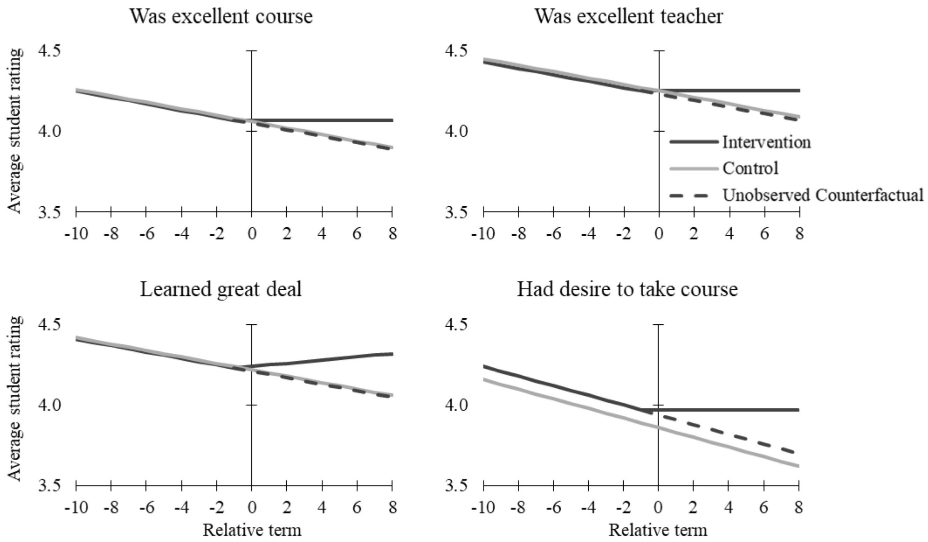

Fig. 2. Linear regression of actual evaluation scores for the intervention/control groups (solid grey and black lines, respectively) and predicted evaluation scores for the intervention group in the absence of the intervention (dashed line).

group. These differences were 0.02 points/term for two items (Was excellent course and Was excellent instructor) and 0.03 points/term for the other two items (*Learned a great deal* and *Had desire to take course*). Thus, instructors who applied but did not participate in the *Teaching Circle* continued to exhibit declining trends of 0.02 or 0.03 points/term, while instructors in the intervention group exhibited a trend that was mitigated for three of the four items (*Was excellent course*, *Was excellent instructor*, and *Had desire to take course*), and it became an increasing trend of 0.01 points/term for the remaining item (*Learned a great deal*). This lead to a 0.08-point increase in the scores of this item by the most recent term in this study.

These effects are demonstrated visually in Fig. 2. The unobserved counterfactual in Fig. 2 represents the predicted trajectory of instructors in the intervention group had they *not* participated in the *Teaching Circle*. Thus, it is simply the slope of the changes in scores across time for all instructors beyond the point of the intervention. It has the same slope and is parallel to the changes in scores for the control group.

3. Discussion

In response to the first part of our research question,

> *Do student evaluation of teaching scores decline over time, as suggested in the literature?*

we find that student evaluation scores for all four questions in our study steadily decreased over time for instructors in our control group, confirming prior research [49, 50]. These findings, derived from as many as 18 teaching terms, showed that average scores for faculty not in our intervention group declined from the first term (Term –10) to the last term (Term 8) by 0.36 points for *Was excellent course*, *Was excellent teacher*, and *Learned a great deal*, and by 0.54 points for *Had desire to take course*. These declines do provide some support for faculty's fears that, on average, student evaluation of teaching scores tend to decline over time. When examining our group of *Teaching Circle* participants prior to their involvement in our faculty development program, we found the same rate of decline in student evaluation scores from the first term (Term –10) to the term before the intervention (Term –1). The declines were 0.02 points/term for *Was excellent course*, *Was excellent teacher*, and *Learned great deal*, and 0.03 points/term for *Had desire to take course*. As illustrated in the unobserved counterfactual in Fig. 2, our analyses demonstrate that, like faculty in control group, these faculty in our intervention group would have continued to experience declines similar to the control group in their scores through the final term (Term 8) had they not participated in the *Teaching Circle*.

In response to the second part of our research question,

> *How does participation in a specially designed faculty development program focused on implementing EBTP mitigate these declines?*

we found that participation in the *Teaching Circle* did affect student evaluation scores in two different ways. For three of the four items (*Was excellent course*, *Was excellent teacher*, and *Had desire to take course*), participation in the *Teaching Circle* reversed the declining trend such that these faculty

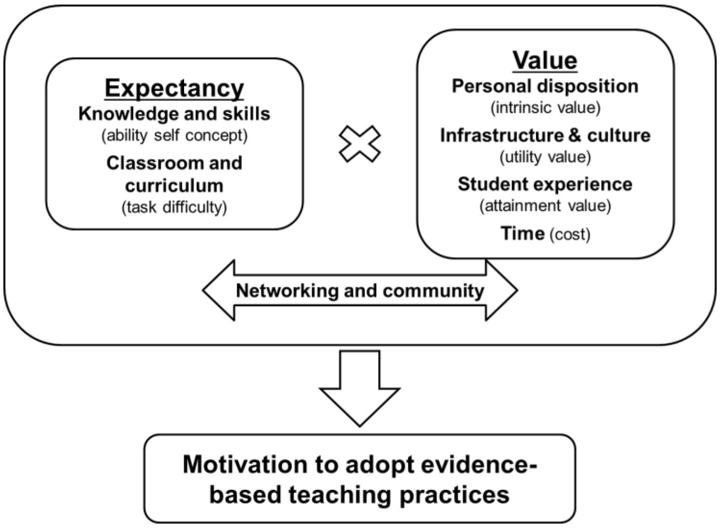

Fig. 3. Expectancy Value Theory-based framework for factors influencing instructors' decisions to adopt EBTP.

actually experienced stable scores. For the fourth item (*Learned a great deal*), participation in the *Teaching Circle* not only reversed the decline, but it resulted in a slight increase in student evaluation scores over time (rather than the predicted decline) to a rate of 0.01 points/term. While these positive results are marginal, they can add up over time. We found that faculty who participated in the *Teaching Circle* consistently had greater average scores for all four items by the end of the 8th term after applying than faculty in the control group.

The findings from our analyses indicate that, for engineering instructors, student evaluation of teaching scores do decline over time, which is consistent with prior research demonstrating that evaluation scores tend to decline or remain neutral over time [49–52]. Murray and colleagues indicated in their research study that the changes in evaluation scores over time was often grouped by academic departments. In other words, the increases or decreases in evaluation scores were consistent across all faculty in that department; however, they do not disclose which findings were attributed to which departments (i.e., the departments were blinded in the study). Thus, instructors and administrators should be cognizant of and prepared for these potential declines.

One way to circumvent these declining trends is to engage in professional development, as confirmed by the findings for our intervention group, which suggest that participation in our *Teaching Circle* program enabled faculty to "reset" their teaching scores. Our program was carefully designed to affect factors that influence faculty motivation to adopt EBTP. We grounded our program in research from the Expectancy Value Theory (EVT), which states that the motivation an individual has to engage in an activity depends on the interaction between expectancy (the degree to which the individual expects to succeed) and the anticipated value (including costs) associated with the action [66, 67]. Multiple sub-factors in EVT (e.g., task difficulty, attainment value, cost) were also instrumental in the development of this program, which are all illustrated in Fig. 3.

Our faculty development program is driven by this theory in that it aims to increase faculty's expectancy by situating the national research on EBTP within the institutional context of the College of Engineering at the University of Michigan. This is accomplished by providing credible research evidence (increasing faculty's "ability self-concept") and by providing personalized support as faculty learn about EBTP (lowering the "task difficulty" perceived by faculty). Furthermore, we aim to increase "value" by offering the program only for those faculty requesting to participate (appealing to "intrinsic value"), institutionalizing the program and signaling administrative support for the program (increasing "utility value"), and emphasizing how using evidence-based teaching practices can improve the student experience (increasing "attainment value"). As a result, the *Teaching Circle* integrates a formal networking and community-building aspect by establishing a safe environment for participants to learn and practice (increasing "utility value" while decreasing "cost"). These features can be adapted by others wishing to support their faculty in implementing EBTP at their own institutions.

Despite the importance of our findings, there are some limitations to keep in mind. First, our work involved engineering instructors at one highly-intensive research university, and thus, our findings may not be generalizable to other contexts. Second, we do not have evidence about actual teaching practices enacted by instructors, and we are unable to tell whether instructors in the intervention groups implemented new teaching practices, and if so, how effective they were in implementing these practices. Therefore, we do not know whether these declines in student evaluation of teaching scores are related to incorporating new practices in the classroom or if evaluation scores decline regardless of which practices are used. Despite these limitations, this study provides valuable evidence for understanding how participation in a faculty development program can impact student evaluation of teaching scores.

4. Conclusions

This research study analyzed data from 41 instructors who participated in a faculty development program at the University of Michigan and 30 who applied but did not participate. The goal of the research was to determine if student evaluation of teaching scores declined over time, as suggested by prior research, and to understand whether or not participating in the *Teaching Circle* mitigated these declines. We found that student evaluation scores did decline over time for faculty who did not participate in the *Teaching Circle*, but the negative trend either stabilized or reverted to a positive trajectory for those faculty who did participate in the *Teaching Circle*.

The *Teaching Circle* was explicitly designed to address some of the barriers to adopting EBTP by, among other things, offering practical strategies for the successful implementation of EBTP and focusing on pedagogies relevant to the engineering context, such as using classroom technology in the engineering classroom and how to write concept questions for multiple-choice examinations. While our research does reinforce prior work revealing declines in student evaluation scores over time, it also points to the promise of faculty professional development in reversing these declines. We hope that continued research on the effectiveness of faculty development programs, like the *Teaching Circle*, can further demonstrate the positive effects of these types of programs and support instructors who wish to adopt EBTP in their classrooms in the future.

References

1. R. T. Blackburn, *Project for Faculty Development Program Evaluation: Final Report*, Ann Arbor, MI, Center for the Study of Higher and Postsecondary Education, 1980.
2. W. Condon, E. R. Iverson, C. A. Manduca, C. Rutz and G. Willett, *Faculty development and student learning: Assessing the connections*, Bloomington, IN, Indiana University Press, 2016.
3. T. A. Angelo and K. P. Cross, Minute Paper, in T. A. Angelo and K. P. Cross (eds), *Classroom Assessment Techniques: A Handbook for College Teachers*, 2nd ed., San Francisco, CA, Jossey-Bass, 1993, pp. 148–153.
4. President's Council of Advisors on Science and Technology, *Report to the president: Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*, Washington, D.C., Executive Office of the President, 2012.
5. L. H. Jamieson and J. R. Lohmann, *Creating a culture for scholarly and systematic innovation in engineering education: Ensuring U.S. engineering has the right people with the right talent for a global society*, Washington, D.C., American Society for Engineering Education, 2009.
6. S. Malcom and M. Feder, *Barriers and opportunities for 2-year and 4-year STEM degrees: Systemic change to support students' diverse pathways*, Washington, D.C., National Academies Press, 2016.
7. D. R. Barnett, *Academic and social integration of nontraditional students: The role of active learning strategies and sense of belonging in integration and persistence (Doctoral dissertation)*, 2014, Available: http://opensiuc.lib.siu.edu/cgi/view content.cgi?article=1940&context=dissertations
8. J. M. Braxton, W. A. Jones, A. S. Hirschy and H. V. Hartley III, The role of active learning in college student persistence, *New Directions for Teaching & Learning*, **115**, 2008, pp. 71–83.
9. J. M. Braxton, J. F. Milem and A. S. Sullivan, The influence of active learning on the college student departure process: Toward a revision of Tinto's theory, *Journal of Higher Education*, **71**(5), 2000, pp. 569–590.
10. D. C. Haak, J. HilleRisLambers, E. Pitre and S. Freeman, Increased structure and active learning reduce the achievement gap in introductory biology, *Science*, **332**(6034), 2011, pp. 1213–1216.
11. D. W. Johnson, R. T. Johnson and K. A. Smith, *Active learning: Cooperation in the college classroom*, Edina, MN, Interaction Book Co., 1991.
12. C. L. Colbeck, A. F. Cabrera and P. T. Terenzini, Learning professional confidence: Linking teaching practices, students' self-perceptions, and gender, *The Review of Higher Education*, **24**(2), 2001, pp. 173–191.
13. N. Fleming, Establishing rapport: Personal interaction and learning, in *IDEA Paper 39*, Manhattan, KS, Center for Faculty Evaluation and Development, Kansas State University, 2003.
14. N. A. Granitz, S. K. Koernig and K. R. Harich, Now it's personal: Antecedents and outcomes of rapport between business faculty and their students, *Journal of Marketing Education*, **31**, 2009, pp. 52–65.
15. L. E. Carlson and J. F. Sullivan, Hands-on engineering: learning by doing in the integrated teaching and learning program, *International Journal of Engineering Education*, **15**(1), 1999, pp. 20–31.
16. J. M. Case and D. M. Fraser, An investigation into chemical engineering students' understanding of the mole and the use of concrete activities to promote conceptual change, *International Journal of Science Education*, **21**(12), 1999, pp. 1237–1249.
17. L. D. Feisel and A. J. Rosa, The role of the laboratory in undergraduate engineering education, *Journal of Engineering Education*, **94**(1), 2005, pp. 121–130.
18. R. M. Felder, D. R. Woods, J. E. Stice and A. Rugarcia, The future of engineering education II: Teaching methods that work, *Chemical Engineering Education*, **34**(1), 2000, pp. 26–39.
19. J. D. Bransford, A. L. Brown and R. R. Cocking, *How people learn: Brain, mind, experience, and school.* Washington, D.C., National Academies Press, 1999.
20. N. Kober, *Reaching Students: What Research Says about Effective Instruction in Undergraduate Science and Engineering*, Washington, D.C., National Academies Press, 2015.
21. M. Prince, Does active learning work? A review of the

research, *Journal of Engineering Education*, **93**(3), 2004, pp. 223–231.
22. E. Seymour and N. Hewitt, *Talking About Leaving*, Boulder, CO, Westview Press, 1997.
23. O. S. Anderson and C. J. Finelli, A faculty learning community to improve teaching practices in large engineering courses: Lasting impacts, in *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference and Exposition 2014*, Indianapolis, IN, 2014.
24. C. J. Finelli and J. M. Millunchick, The teaching circle for large engineering courses: A cohort-based model for faculty development, in *Proceedings of the 2013 ASEE Annual Conference & Exposition*, Atlanta, GA, 2013.
25. C. J. Finelli, S. R. Daly and K. M. Richardson, Bridging the research-to-practice gap: Designing an institutional change plan using local evidence, *Journal of Engineering Education*, **103**(2), 2014, pp. 331–361.
26. M. D. Cox, Introduction to Faculty Learning Communities, *New Directions for Teaching and Learning*, **97**, 2004, pp. 5–23.
27. C. J. Finelli, T. Pinder-Grover and M. C. Wright, Consultations in teaching: Using student feedback for instructional improvement, in C. Cook and M. Kaplan (eds), *Advancing the Culture of Teaching on Campus*, Sterling, VA, Stylus, 2011, pp. 65–79.
28. K. Trigwell and M. Prosser, Qualitative variation in constructive alignment in curriculum design, *Higher Education*, **67**(2), 2014, pp. 141–154.
29. H. G. Murray, Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness, *Journal of Educational Psychology*, **75**(1), 1983, pp. 138–149.
30. H. W. Marsh and M. J. Dunkin, Students' evaluations of university teaching: A multidimensional perspective, in J. C. Smart (ed), *Higher Education: Handbook of Theory and Research*, **8**, Houten, Netherlands, Springer Netherlands, 1992.
31. H. W. Marsh and L. A. Roche, Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility, *American Psychologist*, **52**(11), 1997, pp. 1187–1197.
32. M. D. Svinicki and W. J. McKeachie, *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers*, Belmont, CA, Wadsworth Publishing, 2014.
33. H. G. Murray, Low-inference teaching behaviors and college teaching effectiveness: Recent developments and controversies, in R. P. Perry and J. C. Smart (eds), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, Houten, Netherlands: Springer Netherlands, 2007, pp. 145–200.
34. K. Bedard and P. Kuhn, Where class size really matters: Class size and student ratings of instructor effectiveness, *Economics of Education Review*, **27**(3), 2008, pp. 253–265.
35. S. H. Liaw and K. L. Goh, Evidence and control of biases in student evaluations of teaching, *International Journal of Educational Management*, **17**(1), 2003, pp. 37–43.
36. I. Neath, How to improve your teaching evaluations without improving your teaching, *Psychological Reports*, **78**(3), 1996, pp. 1363–1372.
37. A. G. Greenwald and G. M. Gillmore, Grading leniency is a removable contaminant of student ratings, *American Psychologist*, **52**(11), 1997, pp. 1209–1217.
38. V. E. Johnson, *Grade Inflation: A Crisis in College Education*, New York, NY, Springer, 2003.
39. H. W. Marsh and L. A. Roche, Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders?, *Journal of Educational Psychology*, **92**(1), 2000, pp. 202–228.
40. L. M. Aleamoni, Typical faculty concerns about student evaluation of teaching, *New Directions for Teaching & Learning*, **31**, 1987, pp. 25–31.
41. A. Boring, K. Ottoboni and P. B. Stark, Student evaluations of teaching (mostly) do not measure teaching effectiveness, *ScienceOpen Research*, **10**, 2016, pp. 1–11.
42. S. A. Basow, Student evaluations of college professors: When gender matters, *Journal of Educational Psychology*, **87**(4), 1995, pp. 656–665.
43. L. Carson, Gender relations in higher education: Exploring lecturers' perceptions of student evaluations of teaching, *Research Papers in Education*, **16**(4), 2001, pp. 337–358.
44. M. S. Chamberlin and J. S. Hickey, Student evaluations of faculty performance: The role of gender expectations in differential evaluations, *Educational Research Quarterly*, **25**(2), 2001, pp. 3–15.
45. P. B. Stark and R. Freishtat, An evaluation of course evaluations, *ScienceOpen Research*, **9**, 2014, pp. 1–26.
46. F. Mengel, J. Sauermann and U. Zolitz, Gender bias in teaching evaluations, *Journal of the European Economic Association*, **16**, 2018, pp. 1–50.
47. O. L. Liu, Student evaluation of instruction: In the new paradigm of distance education, *Research in Higher Education*, **53**(4), 2012, pp. 471–486.
48. A. P. Rovai, M. K. Ponton, M. G. Derrick and J. M. Davis, Student evaluation of teaching in the virtual and traditional classrooms: A comparative analysis, *The Internet and Higher Education*, **9**(1), 2006, pp. 23–35.
49. K. A. Feldman, Seniority and experience of college teachers as related to evaluations they receive from students, *Research in Higher Education*, **18**(1), 1983, pp. 3–24.
50. H. G. Murray, R. B. Jelley and R. D. Renaud, Longitudinal trends in student instructional ratings: Does evaluation of teaching lead to improvement of teaching?, New York, NY, American Educational Research Association, 1996.
51. A. C. Carle, Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach, *Computers & Education*, **53**(2), 2009, pp. 429–435.
52. H. W. Marsh, Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness, in R. P. Perry and J. C. Smart (eds), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, Houten, Netherlands, Springer Netherlands, 2007, pp. 319–383.
53. J. A. Centra and N. B. Gaubatz, Is there gender bias in student evaluations of teaching?, *The Journal of Higher Education*, **71**(1), 2000, pp. 17–33.
54. B. G. Davis, *Tools for Teaching*. Hoboken, NJ, John Wiley & Sons, 2009.
55. P. Abrami, S. d'Apollonia and S. Rosenfield, The dimensionality of student ratings of instruction: What we know and what we do not, in R. P. Perry and J. C. Smart (eds), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, 2007, pp. 385–456.
56. S. L. Benton and W. E. Cashin, Student ratings of teaching: A summary of research and literature, in *IDEA Paper #50*, Center for Faculty Evaluation and Development, Manhattan, KS, Kansas State University, 2012.
57. S. L. Benton and W. E. Cashin, Student ratings of instruction in college and university courses, in M. B. Paulsen (ed), *Higher Education: Handbook of Theory and Research*, Houten, Netherlands, Springer Netherlands, 2014, pp. 279–326.
58. J. Sojka, A. K. Gupta and D. R. Deeter-Schmelz, Student and faculty perceptions of student evaluations of teaching: A study of similarities and differences, *College Teaching*, **50**(2), 2002, pp. 44–49.
59. C. Ballantyne, Online evaluations of teaching: An examination of current practice and considerations for the future, *New Directions for Teaching & Learning*, **96**(103–112), 2003.
60. J. A. Centra, *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*, San Francisco, CA, Jossey-Bass, 1993.
61. P. Ginns, M. Prosser, and S. Barrie, Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students, *Studies in Higher Education*, **32**(5), 2007, pp. 603–615.
62. H. W. Marsh and L. Roche, The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness, *American Educational Research Journal*, **30**(1), 1993, pp. 217–251.
63. F. Zabaleta, The use and misuse of student evaluations of teaching, *Teaching in Higher Education*, **12**(1), 2007, pp. 55–76.

64. J. D. Angrist and J. S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ, Princeton University Press, 2008.
65. A. Abadie, Semiparametric difference-in-differences estimators, *The Review of Economic Studies*, **72**(1), 2005, pp. 1–19.
66. J. S. Eccles, B. L. Barber, K. Updegraff and K. M. O'Brien, An expectancy-value model of achievement choices: The role of ability self-concepts, perceived task utility, and interest in predicting activity choice and course enrollment, *Seeon Conference on Interest and Gender*, Kiel, Germany, 1998, pp. 267–280.
67. A. Wigfield and J. S. Eccles, Expectancy value theory of motivation,*Contemporary Educational Psychology*, **25**(1), 2000, pp. 68–81.

Matt DeMonbrun is a PhD candidate at the Center for the Study of Higher and Postsecondary Education in the School of Education at the University of Michigan. His research interests include college student development theory, intergroup interactions, and teaching and learning practices and how they relate to student learning outcomes in engineering education.

Julia Kerst is a third-year undergraduate student in Electrical Engineering at the University of Michigan where she is also the Internal Vice President of the Society of Women Engineers. She has been doing research in Engineering Education since May 2016, and she focuses on student responses to faculty trying new learning techniques.

Hanna Pfershy is a third year undergraduate student at the University of Michigan studying Biomedical Engineering. She started as a research assistant in Engineering Education Research in summer 2016.

Cynthia J. Finelli is Associate Professor of Electrical Engineering and Computer Science, Associate Professor of Education, and Director of Engineering Education Research at University of Michigan. Her research areas include student resistance to active learning, the impact of the classroom space on teaching and learning, the use of classroom technology to increase student learning and engagement, and faculty adoption of evidence-based teaching practices. She recently led an international initiative to develop a taxonomy for the field of engineering education research. Dr. Finelli is a Fellow of the American Society of Engineering Education, Deputy Editor of the *Journal for Engineering Education*, Associate Editor for the *IEEE Transactions on Education*, and past chair of the Educational Research and Methods Division of ASEE. She founded the Center for Research on Learning and Teaching in Engineering at University of Michigan in 2003 and served as its Director for 12 years.