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

Ms. Julia F. Kerst, University of Michigan

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

Ms. Hanna Pfershy, University of Michigan, Engineering Education Research

Hanna 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.

Mr. Robert Matthew DeMonbrun, University of Michigan

Matt DeMonbrun is a Ph.D. Candidate at the Center for the Study of Higher and Postsecondary Education (CSHPE) 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.

Dr. Cynthia J. Finelli, University of Michigan

Dr. Cynthia 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, 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.

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

Introduction

Despite recent research on the use of active learning and its benefits in the classroom, many engineering faculty have yet to implement these practices [9,13]. Previous studies at the University of Michigan have focused on identifying barriers for faculty who wish to implement active learning in their classrooms and outlining methods to overcome these barriers [6]. These findings pointed to several reasons why faculty are reluctant to adopt active learning, including: lack of familiarity with active learning practices, content and course restrictions, and the absence of a support structure for those attempting to change. Another commonly cited barrier was the fear of negative evaluations and their effects on tenure. Faculty are unsure how students will react to a different teaching style, and since tenure decisions depend in part on student evaluations of teaching, faculty are understandably concerned about the possible negative implications of adopting active learning [9].

After identifying common barriers to adoption of active learning, the third author of this paper created a faculty development program to proactively address some of those barriers [7]. This program, named the *Teaching Circle for Large Engineering Courses* (hereafter referred to as the *Teaching Circle*), provides a place for faculty to discuss ideas, fears, and successes, and a way for participants to support each other as they adopt active learning practices. The program was piloted during the Fall 2011 term and fully implemented during the following Winter 2012 term. Interested faculty apply for the program, with approximately six or seven participants selected each term. The *Teaching Circle* program consists of four monthly meetings held throughout the term where participants come together to discuss relevant readings about active learning in engineering and focus on implementation strategies. Two trained facilitators lead the discussions, and many topics (such as building rapport in large classes, influencing student motivation, and effectively continuing active learning practices) are touched upon during these sessions. After completion of the program, faculty are eligible for a $1,000 grant to further support their active learning endeavors.

Preliminary results from the *Teaching Circle* [1,7] were positive. The engineering faculty sampled had increases in six important teaching behaviors after *Teaching Circle* participation. Of these, four teaching behaviors demonstrated statistically significant performance increases. As well, faculty were observed using more active learning practices. It is unclear, though, how using these active learning practices will impact student ratings. We examine the effects of the *Teaching Circle* program on student evaluations of teaching to better understand how students respond to the use of new active learning practices and how programs like the *Teaching Circle* might affect this response. So, we ask: “Does the adoption of active learning strategies lead to a decline in students’ evaluations of teaching as is suspected by faculty, or does a faculty development program like the *Teaching Circle* mitigate these negative responses?

Student Resistance

Some studies have shown that speculations about student resistance to active learning may have merit. While findings about how strongly students resist faculty's use of active learning are inconclusive, it is widely accepted such resistance does exist in the STEM classroom [16]. This student resistance manifests itself in several different forms, including: passive resistance (refusal to participate in activities), partial compliance (completing work quickly and unenthusiastically), and open resistance (vocal complaints) [15]. Possible reasons for this resistance include the increased emphasis on higher-order thinking and individual exploration that active learning places on students as compared to the traditional classroom, the additional time students may be required to spend inside and outside of class preparing for the course, and the anxiety and unfamiliarity students feel about the new methods [16].

Faculty perceptions that student resistance may lead to poor negative evaluations is one of the barriers to adoption of active learning [6]. The threat of poor evaluations deters faculty from adopting active learning practices, since they in part determine tenure decisions. In addition, even for those who do successfully adopt active learning, this negativity from students can cause instructors to become discouraged and return to more familiar methods [5].

Student Evaluations

To understand the effects of the *Teaching Circle*, we analyzed student evaluations of teaching. The literature on the value of using student evaluations is mixed, however. Although Stark [4,14] asserts that student evaluations may not accurately measure teaching effectiveness, faculty rely heavily on them, especially during promotion and tenure. Active learning practices rely heavily on building rapport in the classroom, so it is important to understand how a student perceives the classroom environment [8].

Other literature defends the validity of student evaluations, claiming that students are qualified to judge teaching effectiveness accurately, do not score poor instructors higher, and are motivated through positive feedback [3,11]. And though there is some evidence that students score faculty differently based on instructor gender, ethnicity, or personality, in many cases researchers have found these biases to be small [2]. For our analysis and at University of Michigan, student evaluation data is an important motivator of faculty teaching practices.

Every class taught at University of Michigan is evaluated by students at the end of the term. These evaluations are completed online voluntarily by the student before the grades are final. The evaluation questions vary from class to class, but every course has four required questions and multiple other questions chosen by the faculty, department, and college. Students use a rating scale of 1 (strongly disagree) to 5 (strongly agree) for all questions. For our analysis, we selected 12 questions that: (a) we believed measured the outcomes associated with the *Teaching Circle* and (b) had sufficient responses amongst all groups for use in the analysis. These questions are listed in Table 1.

We selected these 12 questions for the following reasons:

- Question 1 gives an overview of what students thought of a course.
- Question 2 is tied directly to our research study, as the goal of the *Teaching Circle* is to support individual instructors as they improve the experience and learning of students.
- Questions 3, 15, and 23 focus on how much students feel that they learned in class, and that is likely to change with varied teaching practices. For instance, active learning is designed to improve student knowledge and retention [5], and while grades can show this, it is also important to take into account how much a student feels that they have learned.
- Question 4 is heavily based on the student's interests and feelings on the course material. For example, if students have a strong desire to take a course, this positive feeling may translate to better ratings from students.
- Questions 201 and 207 look at how students interpret the active learning styles of teaching. They may find the student-focused work more difficult to understand and could interpret the teaching style as unclear explanations. Additionally, students may believe that the faculty lacks knowledge in the subject, as the student is more responsible for their own learning, rather than being taught it explicitly.
- Question 216 focuses on how the faculty handled student questions. In active learning practices, like the ones the *Teaching Circle* promotes, students are asked to interpret more material compared to a traditional classroom [12]. Therefore, more questions may arise than traditional classroom settings.
- Question 217 looks at rapport in the classroom. An important aspect of the *Teaching Circle* is that it is aimed for large, lecture-style classrooms, where it is difficult to build rapport. Rapport in a classroom is very beneficial, especially in large lectures [8].
- Questions 229 and 230 focus on instructor preparedness. One of the major barriers to bringing active learning strategies cited previously was lack of time to prepare new lesson plans. When transitioning to a new teaching style, faculty may struggle with preparation of new lectures, homework, and other materials. The *Teaching Circle* aims to mitigate that by providing a support structure where faculty can discuss implementation strategies and have designated time to work.

Table 1: Evaluation Questions Analyzed

Number	Question Text
1*	*Overall, this was an excellent course.*
2*	*Overall, the instructor was an excellent teacher.*
3*	*I learned a great deal from this course.*
4*	*I had a strong desire to take this course.*
15	*I increased my ability to apply math and science knowledge to engineering problems.*
23	*I increased my ability to formulate, and solve engineering problems.*
201	*The instructor gave clear explanations.*
207	*The instructor appeared to have a thorough knowledge of the subject.*
216	*The instructor acknowledged all questions insofar as possible.*
217	*The instructor treated students with respect.*
229	*The instructor used class time well.*
230	*The instructor seemed well prepared for class meetings.*

*Questions required by the university

Methods

As of 2016, there have been 41 participants in the *Teaching Circle*. These 41 comprise the *intervention* group for our study, and they represent multiple departments and they span all ranks. Additionally, there were 40 faculty who applied to the program but were not originally accepted into the *Teaching Circle* due to program constraints. These faculty serve as our *control* group. Ten of these *control* faculty went on to apply in subsequent years, and were accepted later. We studied these faculty at two different points in time: they are included in the *control* group until acceptance into the program, and they are included in the *intervention* group afterwards. Demographics for our *control* and *intervention* groups are described in Table 2.

Table 2: *Teaching Circle for Large Engineering Courses* Demographics

	Control (N=40)	**Intervention (N=41)**
Gender		
Female	10	14
Male	30	27
Department		
Aerospace Engineering	1	2
Atmospheric, Oceanic, and Space Science	3	2
Biomedical Engineering	1	4
Civil and Environmental Engineering	2	4
Chemical Engineering	3	7
Computer Science Engineering	7	5
Electrical and Computer Engineering	3	1
Industrial and Operations Engineering	2	3
Mechanical Engineering	10	9
Material Science Engineering	3	1
Technical Communication	5	3
Rank		
Lecturer	17	5
Assistant Professor	6	18
Associate Professor	3	5
Professor	14	13

We compiled student evaluation data for every class taught by each of our faculty participants (including both the *control* and *intervention* groups) from Fall 2008 through Winter 2016. These classes ranged from large lecture classes to small graduate seminars to independent studies. However, since the *Teaching Circle* focuses on large engineering undergraduate courses, we eliminated all graduate-level courses and all courses having fewer than five responses. We labelled the term during which the faculty participated in the *Teaching Circle* (or applied, if they did not participate) as term 0. Then, we then assigned numbers sequentially, referring to only the terms in which a faculty member taught. The first term a faculty member taught prior to *Teaching Circle* application is term –1, and the first term a faculty member taught after is term +1. Altogether, our study includes terms from –10 to 8. If a faculty taught more than one class in a particular term, we computed a weighted average for the data based on the number of

responses in each class. As such, larger engineering courses have a heavier weight in the score compared to smaller discussion sections.

Results

Analysis 1: Linear Regression

Since the goal of the *Teaching Circle* is to promote lasting improvements, we studied the *trends* in the student ratings data over time by evaluating the slope of the course evaluation data over time for each individual faculty. That is, we took the average scores of each faculty's evaluation data for a given question across all terms and plotted a linear slope that best represented the changes in these average scores before and after the intervention. Scores before term 0 were coded as *Control Before* or *Intervention Before*, and scores after term 0 were coded as *Control After* or *Intervention After*. If a given faculty did not have at least two points of data for a given question, we did not count them in the analysis for that question. Table 3 shows the average slopes of these groups and the differences between and after application.

Table 3: Average Slope Values

	Control			Intervention		
Number	**Before**	**After**	**Difference**	**Before**	**After**	**Difference**
1	0.03	–0.08	–0.10*	0.03	–0.06	–0.09*
2	0.02	–0.10	–0.12*	–0.02	–0.07	–0.06
3	0.02	–0.08	–0.09*	–0.02	–0.05	–0.03
4	0.04	–0.03	–0.07	–0.06	–0.10	–0.04
15	0.01	–0.09	–0.10*	0.05	–0.01	–0.06
23	0.01	–0.12	–0.13*	0.04	0.01	–0.03
201	0.00	–0.15	–0.15*	–0.07	–0.01	0.05*
207	–0.01	–0.04	–0.03	–0.02	0.00	0.01
216	–0.01	–0.12	–0.11*	–0.01	–0.01	–0.01
217	0.02	0.03	0.01	–0.03	–0.02	0.01
229	0.00	–0.12	–0.12*	–0.03	0.01	0.04
230	–0.01	–0.06	–0.05	–0.01	0.02	0.03

*$p<0.05$

Using a paired t-test, we compared the numerical values of the differences in these slopes before and after term 0. We found ten statistically significant differences across eight of the twelve questions, most of which were across the *control* group before and after applying for the Teaching Circle For example, faculty in the *control* group experienced a significant decrease in their average slope values across Questions 1, 2, 3, 15, 23, 201, 216, and 229 ($p<0.05$). These differences in slopes ranged from –0.09 to –0.15. In other words, across each of these questions, faculty in the *control* group experienced significant declines in their average evaluation scores after their application to the *Teaching Circle*.

Declines were much less common for faculty in the *intervention* group. In fact, *intervention* faculty only experienced significant declines in their average slope values for one of the 12 questions, Question 1 (*Overall, this was an excellent course*; $p<0.05$). Additionally, these faculty

saw significant **increases** in their average slope values across Question 201 (*The instructor gave clear explanations*; $p<0.05$).

Analysis 2: Fixed Effects

Given the longitudinal nature of our dataset (student evaluation data for each faculty member across up to 18 consecutive terms), we conducted a fixed effects model to account for the correlations between each individual's evaluation scores over time. In other words, in a standard linear regression model, it is assumed that the errors between each of the observations in the dataset are independent of one another, but this assumption is not valid with time-series data. Instead, the fixed effects model allows us to account for the correlation in the faculty's student evaluation scores [10]. For this analysis, we only examined the four questions required on all student evaluations (Questions 1, 2, 3, and 4) as these questions were most frequently answered by students during each time interval. We compared the relationship between the *control* and *intervention* groups before and after applying to the *Teaching Circle*. The constant (mean score for the *Control Before* group) and coefficients are presented in Table 4.

Table 4: Fixed Effects Model

	Control		**Intervention**	
Number	**Before (Constant)**	**After**	**Before**	**After**
1, *Estimate*	4.05	0.01	0	0.11
2, *Estimate*	4.25	–0.02	0	0.14*
3, *Estimate*	4.21	0.00	0	0.12*
4, *Estimate*	3.92	0.03	0	0.04

*$p<0.05$

Our fixed effects analysis indicates significantly higher scores ($p < 0.05$) for only the *Intervention After* group. Across Questions 2 and 3, the *Intervention After* group scored significantly higher than the *Control Before* group, which averaged scores of 4.25 and 4.21, respectively. This included Question 2 (*Overall, the instructor was an excellent teacher*), the focal point of this analysis, in which faculty in the *Intervention After* group scored significantly higher than all other groups in the analysis. For example, after participating in the *Teaching Circle*, faculty in the *intervention* group received scores, on average, 0.14 points higher than both the *Control Before* and *Intervention Before* groups, and 0.16 points higher than the *Control After* group. Given the non-significant differences between the *Control Before* and *Intervention Before* groups, it can also be said that these faculty in the *intervention* group increased their scores for this question by 0.14 points after their experience in the *Teaching Circle*. Similarly, we found that faculty in the *intervention* group scored 0.12 points higher on Question 3 (*I learned a great deal from this course*) than the result of the groups. We found no significant differences across Questions 1 and 4.

Conclusions and Implications

Our data shows that participation in the *Teaching Circle* positively improves student evaluation scores for some questions, and it significantly improves them in some cases. In 5 of the 12

questions analyzed, the *intervention* group showed positive improvement after participated in the *Teaching Circle* (Table 3). From our first analysis using linear regression, while the *control* group had significantly decreasing scores before and after term 0 for two questions (Questions 15 and 23), the *intervention* group did not reflect this decreasing trend in scores. In one other case, the *Intervention After* scores were significantly greater than those of the *Control After* group (Question 23). These results might indicate that the faculty who participated in the *Teaching Circle* received the support they needed to make changes to their curriculum without seeing a decrease in evaluation scores. Although we do not know whether or not the *control* group still attempted to implement active learning strategies after applying for the *Teaching Circle*, their scores did significantly decrease across many questions after this event.

From our second analysis employing fixed effects, we found similar significant relationships. We saw significant increases in the scores of the *Intervention After* group (as compared to *Control Before*) in 2 of the 4 required questions, resulting in between 0.12 and 0.14 point increases when compared to other groups. This means that, even after controlling for the increases in teaching evaluations over time, this *Intervention After* group averaged scores that were at least 0.10 point higher than their peers in the *control* group.

Overall, it appears the perception that active learning strategies negatively impact student evaluation scores may be unfounded. In many cases, the student evaluation scores of faculty who participated in the *Teaching Circle* increased after the program when compared to a *control* group. This is especially true of those questions that asked students about how much they learned in class and how the instructor handled their questions. Thus, we find that this barrier to implementation of active learning is not a barrier at all, and it could be seen as a reward.

While this study was conducted/performed/etc. at a large research institution, the results can be applied to other universities. The apprehension regarding student resistance and poor evaluations may be misleading when faculty are given appropriate support in their adoption of active learning practices. While we used a term-long faculty learning community for interested faculty teaching large engineering courses, there are certainly other ways to support faculty during the transition, such as creating smaller, department-based discussion groups or offering occasional workshops on implementing new active learning practices in the classroom. We hope that others will be able to adapt our ideas into their own context to support faculty in their attempt to improve teaching. Next steps for this research is to use continuing data from courses taught this past year to see if the trends do indeed continue, or analyzing additional evaluation questions.

References

1. Anderson, O. S., & Finelli, C. J. (2014). A faculty learning community to improve teaching practices in large engineering courses: Lasting impacts. *Proceedings of 2014 ASEE Annual Conference & Exposition*, Indianapolis, IN.
2. Barr, J., Benton, S., Li, D., & Ryalls, K. (2016). Response to bias against female instructors. *IDEA Editorial Note No. 2*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
3. Benton, S. L., & Ryalls, K. R. (2016). Challenging misconceptions about student ratings of instruction. *IDEA Paper No. 58.* Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
4. Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.
5. Felder, R. M., & Brent, R. (1996). Navigating the bumpy road to student-centered instruction. *College teaching*, *44*(2), 43–47.
6. Finelli, C. J., Daly, S. R., & Richardson, K. M. (2014). Bridging the research-to-practice gap: Designing an institutional change plan using local evidence. *Journal of Engineering Education, 103*(2), 331–361.
7. Finelli, C. J., & Millunchick, J. M. (2013). The teaching circle for large engineering courses: Clearing the activation barrier. *Proceedings of the 2013 ASEE Annual Conference & Exposition*. Atlanta, GA.
8. Fleming, N. (2003). Establishing rapport: Personal interaction and learning. *IDEA Paper No. 39*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
9. Henderson, C., & Dancy, M. H. Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research*, *3*, 1–14.
10. Maddala, G. S., & Lahiri, K. (1992). *Introduction to econometrics* (Vol. 2). New York City, NY: Macmillan.
11. Marsh, H.W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). New York City: NY: Springer Publishing.
12. Murray, H. G. (2007). Low-inference teaching behaviors and college teaching effectiveness: Recent developments and controversies. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 145-200). New York City, NY: Springer Publishing.
13. Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223–231.
14. Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research.*
15. Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice.* Hoboken, NJ: John Wiley & Sons.
16. Yadav, A., Lundeberg, M., Subedi, S., & Bunting, C. (2011). Problem-based learning in an undergraduate electrical engineering course. *Journal of Engineering Education, 100*(2), 253–280.