

# **Analyzing Student Achievement to Measure the Effectiveness of Professional Development for Active Learning Strategies in the Engineering Classroom**

**Sarah Hoyt**  
**Arizona State University**

**Lydia Ross**  
**Arizona State University**

**Eugene Judson**  
**Arizona State University**

**Stephen Krause**  
**Arizona State University**

**Lindy Mayled**  
**Arizona State University**

*When paired with a classroom observation rubric, like the Reformed Teaching Observational Protocol (RTOP), student achievement can provide evidence that active learning is present and impacting how students perform and interact with their coursework. This article also builds upon past findings as they relate to student achievement and how the data associated with this five-year project aligns with the research that student achievement increases in classrooms that utilize active learning and engagement-style strategies. Results show an increase in persistence along with an overall positive shift in grade distribution.*

## **INTRODUCTION**

Over the past several decades, higher education has seen a shift in teaching methodology. Faculty coming into colleges and universities today not only have to conduct research in their field, but they must also be familiar with pedagogy and how to best prepare students for their careers. One method that has proven widely useful in contextualizing course content to create relevance for students is active learning. Active learning promotes greater learning through engaging students by taking a more student-centered approach in teaching. Because this is such a widely researched topic, faculty may have access to a lot of information that many don't have the time to sift through. As a result, many administrative teams have begun incorporating professional development opportunities to introduce faculty to active learning strategies. This paper analyzes a faculty development program designed to indoctrinate engineering faculty

with active learning skills in order to determine if a connection exists between professional development and student achievement.

This large-scale faculty development program was implemented at a large, research-focused Southwestern university and had participants from 7 different disciplines from its engineering college: aerospace, biomedical, chemical, civil, construction, materials, and mechanical. The five-year program, ending this year, was funded and supported through the National Science Foundation's Improving Science Education (IUSE) initiative.

The goal of this paper is to share student achievement data and trends as they relate to active learning in the classroom and to provide a discussion on how active learning affects student achievement. The overall goals of the project, increasing the awareness and use of student-centered, or active learning, pedagogical practices amongst a multi-disciplinary group of engineering faculty, along with a complete analysis on its findings, can be found in previous papers published in earlier years.

## **REVIEW OF RELATED RESEARCH**

### **Professional Development**

Faculty development has been proven to be effective in improving instructional outcomes and professional practice in teaching. Those studies have also shown that faculty development opportunities can boost student learning (Rutz, Ciner, Marfleet, 2010). Most faculty development opportunities follow a workshop model, even though this has been shown to be widely ineffective (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). This is often due to a dissemination of material only, with few opportunities for faculty to interact or contextualize the content as it relates to their profession. Applying a more authentic approach to faculty development, as outlined by Ann Webster-Wright (2010) can help faculty become more self-directed and use faculty development opportunities to scaffold their learning as it pertains to their craft. This self-efficacy helps ignite a change process, similar to that described by EM Rogers' Model of Diffusion, discussed more fully below.

### **Active Learning**

Active learning is an evidenced-based pedagogical tool that shifts learning from teacher-centered to student-centered. The strategies that fall under this umbrella help instructors engage their students through different means with the ultimate goal of increasing learning and improving student outcomes. Active learning provides an alternative to traditional lecture-based instruction that has long been identified as largely ineffective for a majority of students. As such, there is ample research that supports greater student outcomes when active learning is embedded into classroom instruction, including a widely-cited meta-analysis of over 200 studies by Freeman et al. (2014) that found students who were enrolled in courses that utilized active learning illustrated higher learning and comprehension on concept inventories. This is similar to the work of John Hattie (1996) who identified different factors in education that had the most significant effect on student achievement; those strategies with the highest percentage of impact with regard to classroom and instructional practices were all examples of active learning. Another compelling case for active learning is found in Prince's (2004) work, where he reviews the evidence for how effective active learning can be in the engineering classroom.

While active learning is not a new strategy, it has not been as thoroughly adopted at the higher education level compared to its permeation into K12. Despite this slow adoption, emerging studies are starting to support the idea that infusing curriculum with active learning strategies can have a positive impact on student outcomes at the university level (Felder & Brent, 2016; Freeman, et al, 2014; Prince, 2004). In response, this grant was developed to create a robust faculty development program aimed at teaching engineering faculty how to utilize active learning in their classes. This was done through a series of workshops, coupled with classroom observations and instructional coaching, ultimately resulting in a Community of Practice (CoP) to help sustain the overall program.

## CONCEPTUAL FRAMEWORK

As aforementioned, professional development is a fundamental part of education as it allows faculty to learn about changes in pedagogy and receive instruction on implementing best practices into the classroom. Research shows, however, that faculty respond to the change process differently when they participate in faculty development. One study in particular, conducted by Everett Rogers, examined how people react or process change when introduced to a new innovation (2003). Called the Model of Diffusion of Innovation, Rogers' research focused on a five-stage model that describes how people adopt an innovation:

1. Awareness or knowledge - an individual is exposed to an innovation
2. Persuasion or Interest - interest in the subject grows and individuals seek out further information about the innovation
3. Evaluation and Decision - individual either adopts or rejects the innovation
4. Implementation and Trial - innovation is tested by an individual
5. Confirmation or Adoption - individual continues and sustains use of the innovation

Faculty typically begin this model, but fail to move past the implementation and trial phase (Rogers, 2003). There are a number of contributors: time is listed as the most used reason for not implementing an innovation in faculty development. Another important feature of the successful professional development program is the utilization of other faculty; faculty report that when they are trained by colleagues who are familiar with their working environment and discipline, they are more responsive to the professional development opportunity. Similarly, conducting professional development that is flexible and is respectful of their schedules is imperative for faculty participants while utilizing the innovation to be learned in the actual development is equally essential for sustainability; participants need to see how the innovation can be run in a contextualized setting. Another component that needs to be modeled for faculty participants is the natural shift from learning to implementation; in other words, taking the learned innovation from classroom into practice. Finally, participants should be provided with ample and robust opportunities to interact with the innovation across multiple platforms and in formal and informal situations.

## STRUCTURE AND DATA COLLECTION

### Sample and Initial Program Structure

The JTFD project aimed to increase active learning practices in engineering classrooms through a multi-year professional development program. The JTFD program, which was funded through the National Science Foundation's Improving Undergraduate Science Education grant, utilized a train-the-trainer model, where the project team trained pairs of disciplinary leaders, who in turn administered the program to groups of faculty from their own discipline. The program was structured so participants attended 8 biweekly workshops in the fall semester and six communities of practice (CoP) sessions in the spring semester. In the following years, faculty members could participate in eight continuing communities of practice sessions, which were voluntary. Faculty were also afforded the option for individual coaching calls and emails in the case of scheduling conflicts or desired follow-up. Individualizing part of the program provided opportunities for supported practice to strengthen new practice implementation.

In total, 82 faculty members completed the program. Overall, faculty from 7 disciplines participated in JTFD: aerospace, biomedical, chemical, civil, construction, mechanical, and materials. A detailed breakdown of participants' demographic information is provided in Table 1.

**TABLE 1**  
**PARTICIPANTS' GENDER BY DISCIPLINE**

<b>Discipline</b>	<b>Number</b>	<b>Female</b>	<b>Male</b>
Aerospace/Mechanical	15	2	13
Biomedical	20	4	16
Chemical	8	5	3
Civil	19	9	10
Construction	10	1	9
Materials	10	2	8

### **Classroom Observations**

Classroom observations are a form of formative feedback that can also be used to measure the effectiveness of a faculty development program. Although there are different forms of classroom observations, this study utilized trained observers with extensive pedagogical background. Classroom observations were scheduled with the participants to evaluate the presence of active learning in the observed classroom. Observations were conducted at three time points during the JTFD program: before they started the program at the start of the fall semesters (pre), during the program at the start of the spring semester (mid), and at the end of the spring semester, which coincided with the end of the program (post). Two distinct observations were conducted for each time point. Following the initial program, a third round of observations was conducted in the spring of 2019. The round was open to JTFD participants and was completely voluntary. Twenty-six faculty responded to the opportunity for an additional observation and coaching session.

Each classroom observation was conducted by two trained observers. Following the observation, the observers met with each faculty member to discuss each person's goals as they pertain to the RTOP rubric. An individualized follow-up plan was devised among the participants and observers and was also used to guide subsequent observations and instructional coaching. In order to examine how the active learning strategies that were presented in the workshops impacted student achievement, course-level data for undergraduate engineering classes between the academic years of 2012-2013 (before the faculty development program) and 2017-2018 (after the faculty development program) was gathered.

Each observation was conducted by two trained observers with extensive pedagogical expertise and experience sat through classes and identified the presence or absence of active learning activities using the Reformed Teaching Observational Protocol (RTOP) Evaluation Framework, which quantifies the extent of active learning in the classroom. Given this scale, a classroom can be characterized on a scale from 0 to 100, where 0 indicates traditional, lecture-based and 100 indicates a reformed, student-driven class. Lecture-based classes fall within a 0-29-point range, active lecture classes between 30 and 49, and active learning classes constitute a score of 50 to 100.

### **Reformed Teaching Observational Protocol (RTOP)**

Classroom observations were conducted using the Reformed Teaching Observational Protocol (RTOP), which quantifies the extent to which faculty use active learning practices in the classroom (Sawada et al., 2002). The RTOP tool measures what is identified within the observed course; there are no outside assumptions made about the level of active learning integration in previous or subsequent courses. Because of this, a faculty member can have different scores for different observations if active learning activities are built into different lessons. What can be deduced, however, is that if a faculty member is consistently scoring in the range of active learning, then they are regularly utilizing active learning practices within the classroom. Each faculty member was given at least six RTOP observations; instructional coaching sessions followed the observations. The RTOP has five dimensions (see Appendix A for a full version of the ROTP instrument):

- *Lesson design & implementation*: These items assess structure and delivery of class materials. For example, one item measures whether instructors’ draw on prior knowledge of students and the role of students in the learning process.
- *Propositional knowledge (content)*: This construct examines how course material is presented during the class session. For instance, these items focus on the subject matter being taught, and the ways in which the instructor includes key concepts.
- *Procedural knowledge (content)*: These items measure how students engage with course materials. Specifically, these items are focused on assessing the ways that students talk about or characterize the phenomena being taught in the class and whether they are reflective about their learning in the course.
- *Communicative interactions (culture)*: The communicative interactions section focuses on the types of interactions that occur in the classroom. These items examine if classroom culture is inclusive and what types of communication are encouraged in the classroom.
- *Student-teacher relationships (culture)*: This final construct examines the relationship between students and teachers in the classroom. For example, one item focuses on whether the teacher encouraged active participation in the classroom.

Each dimension has five items on which observers rate faculty on a five-point Likert scale, for a total score of 100 points. Given this scale, a classroom can be characterized on a scale from 0 to 100, where 0 indicates a traditional, lecture-based class and 100 indicates a reformed, student-driven class. Lecture-based classes fall within a 0-29-point range, active lecture classes between 30 and 49, and active learning classes constitute a score of 50 to 100.

### **Student Achievement Data**

We collected student achievement data over a multiple year period spanning the JTFD program. From university repositories, we pulled course-level data for all undergraduate engineering courses at the institution from fall 2013 to spring 2019. Data for each course included: discipline, course number, instructor name, and final grade count by letter grade. Longitudinal information from academic years 2012-2013 (pre-intervention) and 2017-2018 (post-intervention) was selected to account for implementation lag time.

The original cleaning process for the overall program evaluation involved merging grade distribution data for cross-listed courses to determine if the discipline, course number, and instructor were aligned. Then, the team removed any +/- grade letters to help streamline them into their letter category: A-/A+ were categorized into the A group, etc. We followed this classification system for A, B, C, D, E, and W. The cleaning process for this examination included removing any course that didn’t have an RTOP score associated with the pre and post, as well as courses that didn’t match those that were observed. Similarly, courses that had higher than 75% As in the pre-time period were removed. We also removed any courses with fewer than 10 A-E grades. After all of these classifications, we ended with a sampling of 108 courses.

Course grades were identified as administered either *before* (pre) or *after* (post) an instructor’s JTFD workshop participation. Students’ grades were examined to assess if there were significant differences before and after instructors participated in JTFD workshops. Courses which had more than 75% of the students earning an A prior to JTFD workshops were filtered out from analyses because it was judged that these courses, largely small project-based courses, were not representative of the targeted course types which have a wider distribution of grades. Data were further filtered to include only courses assigning typical A-F grades (i.e., not pass/fail courses) and with an enrollment of at least 12 students both pre- and post-JTFD participation. Data were then aggregated by course instructor and course name. Pre-to-post comparisons were made strictly based on the *same instructor* and *same course*. Data analysis comprised a multi-step process of descriptive statistics and comparison from pre- to post-time periods.

## RESULTS

Data analysis focused on three areas. First, we examined faculty RTOP scores to better understand the extent to which faculty members integrated active learning practices in the classroom. We also sought to determine whether faculty teaching practices changed after participating in the JTFD program. Next, we examined student achievement of faculty before and after their participation in the program. Lastly, we looked at the connection between RTOP and student achievement distributions to assess for possible connections between shift in teaching practices and student achievement.

### Classroom Teaching Practices: RTOP

Results from the RTOP were analyzed to determine if there were actual shifts in faculty teaching practices in the classroom after participating in the JTFD program. Table X below shows the distribution of RTOP scores across the program.

**TABLE 2  
DISTRIBUTION OF RTOP SCORES**

	<b>Pre</b>	<b>Mid</b>	<b>Post</b>
<b>Minimum</b>	30.50	34.50	36.00
<b>Lower Quartile</b>	45.81	46.50	56.00
<b>Mean</b>	58.44	58.34	66.50
<b>Upper Quartile</b>	71.50	66.06	78.50
<b>Maximum</b>	92.00	95.00	97.50

Overall, we observed shifts in the distribution of RTOP scores over the course of the program, with the average score increasing just under 10 points from the pre to post time point. We observed very little change from the pre- to mid-time points. In order to assess if these changes were significant, a series of paired samples t-tests were conducted (presented in table X below). The 3% average change in active learning practices from pre to mid was not significant ( $p > .05$ ). We did observe a significant increase from both mid- to post- and pre- to post-observation points ( $p < .05$ ). The average percent change was nearly identical for both of these time periods (approximately 12%), suggesting that the majority of the change happened during the latter half of the spring semester.

**TABLE 3  
CHANGES IN TOTAL RTOP SCORES**

<b>Time Period</b>	<b>Average Change</b>	<b>Average Percent Change</b>
Pre to Mid	1.51	2.58%
Mid to Post	7.36*	12.46%
Pre to Post	7.50*	12.68%

\* $p < .05$

Next, we examined the five dimensions of RTOP to assess for deeper change across specific pedagogical areas. Table X below shows the average change in the individual RTOP components across the course of the JTFD program. As with the overall RTOP scores, the majority of the shifts in the individual RTOP components occurred from the pre- to post-time points. The greatest shifts were observed in student-teacher relationships and communications/interactions.

**TABLE 4**  
**SUMMARY OF AVERAGES OF INDIVIDUAL RTOP COMPONENTS**

RTOP Component	Mean (SD)			Pre to Post Shift
	Pre	Mid	Post	
Lesson Design & Implementation	9.21 (4.19)	8.75 (4.07)	10.94 (3.83)	1.73
Propositional Knowledge	16.84 (1.84)	17.01 (1.54)	17.24 (1.78)	0.40
Procedural Knowledge	10.05 (3.97)	9.50 (3.60)	11.03 (3.52)	0.98
Communications & Interactions	10.37 (4.26)	10.45 (3.61)	12.39 (3.52)	2.02
Student-Teacher Relationships	11.96 (4.51)	12.65 (3.93)	14.79 (3.65)	2.83

As with overall RTOP scores, we conducted paired samples t-tests to assess for significance in change across the time periods for each individual RTOP component. There were no significance changes from pre- to mid-time periods ( $p > .05$ ). There were large gains from both mid- to post- and pre- to post-time periods. Notably, we noticed significant gains across all five RTOP dimensions (either from mid- to post- and pre- to post-time periods) ( $p < .05$ ). The greatest gains occurred in lesson design and implementation and student-teacher relationships.

**TABLE 5**  
**AVERAGE CHANGE IN INDIVIDUAL RTOP AREAS**

RTOP Area	Pre to Mid	Mid to Post	Pre to Post
Lesson Design & Implementation	-0.18	2.05*	1.43*
Propositional Knowledge	0.38	0.21	0.68*
Procedural Knowledge	0.32	1.43*	0.91
Communications & Interactions	0.43	1.78*	1.82*
Student-Teacher Relationships	1.27	1.89*	2.73*

\* $p < .05$

### Student Achievement

After removing courses that did not meet the criteria, we had 108 individual courses to focus our pre-/post-analysis on to assess the JTFD program effects. Table X shows the descriptive statistics of the average student grades before and after the program. There were minimal shifts in percentage of students in each letter grade group from pre to post, as well as with average GPA.

**TABLE 6**  
**STUDENT ACHIEVEMENT PRE AND POST**

	Percentage						Average GPA
	A	B	C	D	E	W	
<b>Pre</b>	41.84	37.12	15.25	3.07	2.71	5.57	3.12
<b>Post</b>	41.90	37.34	15.19	2.97	2.61	5.62	3.13
<b>Change</b>	0.06	0.21	-0.06	-0.10	-0.11	0.05	0.01

Data did not meet tests of normality and therefore a series of nonparametric Wilcoxon signed-rank tests were applied to detect changes in grades from pre- to post-JTFD participation. Pre-to-post mean GPAs and proportion of each grade type (i.e., A-F), as well as withdrawals and incompletes, per instructor/course were evaluated across all course levels, across undergraduate courses only, and per grade level (100-, 200-, 300-, and 400-levels). There were no significant changes in mean GPA ( $p > .05$ ). In examining the percentage of each grade type, there were slight shifts in letter grades, with a small increase in students receiving B's in the post-time period ( $p < .05$ ) and a decrease in students receiving D's ( $p < .05$ ).

Next, we examined shifts in student achievement by course size and level, since that can also be a predictor of student learning outcomes. Table X illustrates that in courses with fewer than 100 students, there was an overall positive GPA shift. All classes over 100 indicated a negative GPA shift. Similarly, courses at the 400 level also showed a positive GPA shift which, as mentioned earlier, could be indicative of the trend to include more problem-based instruction in upper-level courses.

**TABLE 7  
GPA SHIFT BY CLASS SIZE**

<b>Class Size Category</b>	<b>Number of Courses</b>	<b>GPA Shift (in points)</b>
0-100	55	+.21
101-200	27	-.09
201-300	9	-.08
301-400	4	-.04
401-500	6	-.08
> 500	6	-.06

**TABLE 8  
GPA SHIFT BY COURSE LEVEL**

<b>Course Level</b>	<b>Number of Courses</b>	<b>GPA Shift (in points)</b>
100	9	-0.03
200	20	-0.17
300	27	-0.06
400	31	0.13

### **Classroom Implementation & Student Achievement**

To further this analysis, we then examined classroom implementation of active learning in connection to student achievement through examination of RTOP and student grade data. For this analysis, we limited our RTOP data to only those that corresponded with classes for which we had student achievement data. To better understand these trends, we conducted basic descriptive statistics. Of the 108 courses, 61 of those courses had corresponding pre- and post-RTOP scores. We found that 66% ( $n=40$ ) of those courses had a positive shift in RTOP score; 19 of those classes had a decrease in RTOP score (31%); and 2 showed no change (3%).

Next, we examined changes in RTOP and GPA to assess for any possible relationships. There was no correlation between pre- and post-GPA and the respective RTOP score ( $p > .05$ ). We also did not find any significant correlations between RTOP change and GPA change ( $p > .05$ ).

We then turned our analysis to those faculty who volunteered to be observed after their first year of the program, when the formal observation period occurred. These voluntary third observations are designated as “post-post” because they occurred one year after formal participation in the program. In total, 26 faculty volunteered for a “post-post” observation; 21 of those faculty had RTOP for all three observational points. Of those who showed a positive change in the RTOP score, 67% had positive GPA shifts in student



achievement. We did find a moderate, positive relationship between change in GPA and change in RTOP from pre- to post-post ( $r = .501, p < .05$ ).

## DISCUSSION

### Classroom Practices

Overall, we saw great improvement in use of active learning strategies across the participating faculty members. This was observed anecdotally through feedback faculty shared with us over the course of the program. We also observed this through the RTOP data collected over the course of faculty participation. The average RTOP score increased approximately 10 points, indicating significant growth in active learning strategies being implemented in the classroom. This shift suggests that the JTFD program successfully shifted faculty from the awareness/knowledge phase to implementation and/or the confirmation/adoption phase in Rogers' *diffusion of innovation*. We observed a 12% change in average faculty use of active learning strategies, which indicates substantial improvement in engagement of student-centered pedagogy in the classroom amongst program participants.

The average percent change was nearly identical from mid- to post, and pre- to post, indicating that nearly all change occurred during the spring semester. Since the majority of change occurred during the spring semester, this suggests that it takes quite a while for the change process to fully take effect. We observed a significant increase in awareness of active learning strategies between the pre- and mid-points (Author, 2017; Author, 2018; Author, 2019), which aligns with the early stages of Rogers' *diffusion of innovation*: awareness and persuasion. However, faculty did not move into the later stages of the *diffusion of innovation* until the second semester of their program participation.

When we broke down the RTOP scores by individual components, we found significant gains across all five dimensions. Just as with overall RTOP scores, substantial gains in each dimension only occurred from the mid- to post- or pre- to post-observation points. Not surprisingly, the largest increase was made in student-teacher relationships, which aligns directly with the main focus of the JTFD program, through centering of instruction on students during classroom time. Communication is a hallmark or contributing factor to student-teacher relationships. And so, in line with that, we also observed substantial average gains in communications and interactions in the classroom. We also noticed substantial gains in lesson design & implementation, which was a major focus of JTFD discussions and program structure, especially around learning objectives and implementing Bloom's taxonomy in the classroom.

### Student Achievement

The body of research overwhelmingly indicates the efficacy of active learning strategies, especially as it relates to student achievement. We observed substantial increase in use of active learning practices in our participants. So, in line with the literature, we expected to see gains in student achievement for JTFD program participants. However, we were surprised to find very minimal to no positive effect of the JTFD program. We conducted rigorous analysis through multiple linear regression, linear mixed effects models, and non-parametric comparisons of grade distributions before and after the program.

In this paper, we focused on comparisons of grade distributions. After very thorough comparisons of shifts in grade distributions, we observed a .21 percent average increase in students receiving a B letter grade and a 0.10 percent average decrease in students who received a D letter grade. Otherwise, we found no other significant effects on student achievement. Surprisingly, we only found a .02% increase in average GPA, or a negligible difference.

While these findings were disappointing, there are a number of possible explanations for the lack of change in student achievement. First, although many of the faculty progressed through the various stages of Rogers' *diffusion of innovation*, it is difficult to measure how far they advanced through each stage. It is likely that faculty members were still in the process of advancing through the innovation stages. Due to the prolonged amount of time it takes to fully advance through the various stages of the *diffusion of innovation* (Rogers, 2003), the JTFD faculty participants had likely not fully advanced to the final stage of adoption, where they practice sustained use of active learning pedagogy in the classroom. As faculty continue to

advance through the stages of innovation into the full, ongoing adoption phase, we would likely see an increase in persistent active learning practices across the entire semester, and as a result, we would expect increased student achievement.

Along with the slower advancement through the five stages of innovation, another possible explanation for the lack of change in student achievement is delayed implementation time and effects of active learning. First, many faculty likely implemented either smaller active learning practices in the classroom, so not creating a fully student-centered environment. With time, as they become more adept at implementing large scale active learning classes, this will probably shift to improved student achievement. Further, there are delayed effects of shifting pedagogical practices on student achievement. In fact, when faculty switch from teacher-centered to more student-centered practices, there is often a decrease in student achievement for a short period of time, as both faculty and students adjust to the new classroom environment.

Another possible explanation for the lack of change in student achievement are the measures utilized in this study. The only measure of student achievement we examined was final grade in a class. While this is a common measure of student achievement, it is not without problems. First, final grades are often composed largely of performance on final exams, which are not the best measure of student learning. Lastly, faculty might have preconceived ideas about student grade distributions. This professional development program did not address topics of grade distributions, so if faculty had fixed thoughts about grade distributions, student achievement measured through a final grade would likely not shift. Therefore, even if student learning actually improved in each course, this might not have been captured through final grade data. Assigning final student grades is a complex process influenced by many factors (Walton et al., 2008; Love & Kitchen, 2010). Broadening scope of study to include other measures of achievement, such as comprehension activities or other grade-related data, might yield different results.

Further, the only measure of active learning included in the study was RTOP scores. While RTOP is a reliable, validated instrument, this might not best measure the implementation of active learning in each classroom. Further, RTOP data comes from point in time analysis, with only two observations each. It is likely that some of the classes observed were either more lecture heavy or more student-centered than a typical class throughout the semester. As such, future research should incorporate more regular observations to better gauge implementation of active learning in each classroom.

Each semester a course is taught brings different variables that can impact results on student achievement. Active learning, fortunately, can be adapted to meet those changes, but this ability to shift practices requires a higher level of comfort in working with a new instructional practice. Similarly, the type of active learning strategies, such as formative feedback, incorporated into the class can have more substantial impacts on student achievement than simply segmenting or shortening pre-existing lectures.

## CONCLUSION

While some of the results of this paper didn't fully align with results from substantiated research, limitations and trends discussed here offer some explanation for the contradictory evidence. For example, using the RTOP to measure active learning is highly effective, but only provides a snapshot into the instructor's overall practice. If an instructor is observed during a heavy-lecture class, even when most of the other courses are infused with active learning strategies, the score will still show a lecture-based or teacher-centered classroom. This is a limitation of the RTOP: the lack of instruments designed to measure the overall impact of active learning over the course of a semester restricts the ability to fully measure the impact of active learning in a class overall.

Ultimately, evidence of a positive GPA shift was present in courses with fewer than 100 students, indicating that, along with active learning strategies, other factors can impact student outcomes. One of those factors that cannot be ignored is class size. Implementing engagement strategies into large classes has definitive physical and organizational limitations. Oftentimes, for example, large classes are conducted in lecture halls with stadium seating; this physical space is not conducive to collaboration among students. Additionally, instructors see their time dwindle if they try to use their physical presence to engage students

because it ultimately takes them longer to move about the room and address questions than it would in a class with fewer than 100 students.

With ongoing support and implementation over an extended period of time, the results explored in this paper indicate that student achievement would be positively impacted by active learning strategies consistently present in the engineering classroom. However, we feel that an even more valuable lesson contained in this paper is that measuring student achievement in a large-scale study such as this is no easy feat. We have found that aligning schedules and classes can be difficult in order to gain the data needed to measure the impact of active learning on achievement. Furthermore, we acknowledge that significant change takes time; previous studies conducted by the team have indicated up to a four-year time period required for a significant impact on student achievement. In the future, projects that devote several years of active data collection paired with RTOP evaluation and ongoing support should be able to present more optimistic findings in relation to the degree of positive impact that active learning strategies have on student achievement in the engineering classroom.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of this work by the National Science Foundation under Grant No. 1524527.

## REFERENCES

- Felder, R.M., & Brent, R. (2016). *Teaching & learning STEM: A practical guide*. San Francisco, CA: Jossey-Bass.
- Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., & Wenderoth, M.P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS*, *11*(23), 8410-8415.
- Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*, 64.
- Hattie, J., Biggs, & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta analysis. *Review of Educational Research*, *66*(2), 99-136. Retrieved from <http://www.jstor.org/stable/1170605>
- Jungst, S., Liklider, L.L., & Wiersema, J. (2003). Providing support for faculty who wish to shift to a learning-centered paradigm in their higher education classrooms. *The Journal of Scholarship of Teaching and Learning*, *3*(3), 69-81.
- Krause, S., Baker, D., Alford, T., Ankeny, C., Carberry, A., Koretsky, M., . . . Gibbons, B. (2015). Effect of implementation of JTF engagement and feedback pedagogy on change of faculty beliefs and practice and on student performance. *2015 ASEE Annual Conference Proceedings*.
- Love, D.A., & Kotchen, M.J. (2010). Grades, course evaluations, and academic incentives. *Eastern Economic Journal*, *36*, 151-163.
- Piburn, M., Sawada, D., Falconer, K., Turley, J. Benford, R., & Bloom, I. (2000). *Reformed Teaching Observation Protocol (RTOP)*. Retrieved from [http://physiced.buffalostate.edu/AZTEC/rtop/RTOP\\_full/PDF/](http://physiced.buffalostate.edu/AZTEC/rtop/RTOP_full/PDF/)
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223-231.
- Ross, L., Glassmeyer, K., Honeycutt, C.F., Judson, E., Krause, S.J., Middleton, J.A., . . . Culbertson, R.J. (2019, June). *Examining Effects of an Evidence-Based Professional Development Program on Student Achievement*. Paper presented at the annual meeting of the American Society for Engineering Education (ASEE). Tampa, FL.
- Rutz, C., Ciner, E., & Marfleet, G. (2010 February). Faculty Development and Assessment: Gateway to Measuring Student Gains. *AAC&U Conference on General Education and Assessment*. Seattle.

Stephens, J., Battle, D., Gormally, C., & Brickman, P. (2017). Show me the way: Future faculty prefer directive feedback when trying active learning approaches. *Journal of College Science Teaching*, 42, 57-65.

Walton, I., Cousin, A., Davison, D., Echeverri, A., Koller, B., .... Takahashi, J. (2008). *Promoting thoughtful faculty conversations about grade distributions*. Report for the Academic Senate for California Community Colleges, 2007-2008 Educational Policies Committee.

Webster-Wright, A. (2010). *Authentic professional learning: Making a difference through learning at work*. Springer Netherlands.

Yoon, K.S., Duncan, T., Lee, S.W-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>

## APPENDIX A RTOP EVALUATION FRAMEWORK

<b>Reformed Teaching Observational Protocol (RTOP)</b>		
<b>LESSON DESIGN AND IMPLEMENTATION</b>		
1	The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.	0 1 2 3 4
2	The lesson was designed to engage students as members of a learning community.	0 1 2 3 4
3	In this lesson, student exploration preceded formal presentation.	0 1 2 3 4
4	This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	0 1 2 3 4
5	The focus and direction of the lesson was often determined by ideas originating with students.	0 1 2 3 4
<b>CONTENT— Propositional knowledge</b>		
6	The lesson involved fundamental concepts of the subject.	0 1 2 3 4
7	The lesson promoted strongly coherent conceptual understanding.	0 1 2 3 4
8	The teacher had a solid grasp of the subject matter content inherent in the lesson.	0 1 2 3 4
9	Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.	0 1 2 3 4
10	Connections with other content disciplines and/or real world phenomena were explored and valued.	0 1 2 3 4
<b>CONTENT—Procedural Knowledge</b>		
11	Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	0 1 2 3 4
12	Students made predictions, estimations and/or hypotheses and devised means for testing them.	0 1 2 3 4
13	Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	0 1 2 3 4
14	Students were reflective about their learning.	0 1 2 3 4
15	Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	0 1 2 3 4

CLASSROOM CULTURE—Communicative Interactions		
16	Students were involved in the communication of their ideas to others using a variety of means and media.	0 1 2 3 4
17	The teacher's questions triggered divergent modes of thinking.	0 1 2 3 4
18	There was a high proportion of student talk and a significant amount of it occurred between and among students.	0 1 2 3 4
19	Student questions and comments often determined the focus and direction of classroom discourse.	0 1 2 3 4
20	There was a climate of respect for what others had to say.	0 1 2 3 4
CLASSROOM CULTURE— Student/Teacher Relationships		
21	Active participation of students was encouraged and valued.	0 1 2 3 4
22	Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	0 1 2 3 4
23	In general the teacher was patient with students.	0 1 2 3 4
24	The teacher acted as a resource person, working to support and enhance student investigations.	0 1 2 3 4
25	The metaphor "teacher as listener" was very characteristic of this classroom.	0 1 2 3 4

## APPENDIX B

### AVG GPA CHANGE FROM PRE TO POST PROFESSIONAL DEVELOPMENT PROGRAM

Course Sample 1-36	pre_GP A_avg	post_G PA_avg	GPA_change	Course Sample 37-72	pre_GP A_avg	post_G PA_avg	GPA_change	Course Sample 73-108	pre_GP A_avg	post_G PA_avg	GPA_change
1	2.43	3.80	1.37	37	3.11	3.27	0.16	73	3.36	3.23	-0.14
2	2.73	3.67	0.95	38	3.29	3.43	0.15	74	2.69	2.55	-0.14
3	2.79	3.62	0.83	39	3.69	3.83	0.14	75	3.00	2.86	-0.14
4	2.91	3.61	0.70	40	3.51	3.63	0.13	76	2.64	2.49	-0.15
5	2.90	3.45	0.55	41	2.77	2.87	0.10	77	3.10	2.95	-0.15
6	3.40	3.92	0.53	42	2.39	2.47	0.09	78	3.14	2.96	-0.18
7	3.23	3.75	0.52	43	3.43	3.52	0.08	79	2.45	2.27	-0.18
8	3.43	3.90	0.47	44	2.87	2.93	0.06	80	3.58	3.39	-0.19
9	3.35	3.79	0.44	45	2.41	2.47	0.06	81	2.75	2.55	-0.20
10	3.50	3.93	0.43	46	2.97	3.02	0.05	82	3.55	3.32	-0.23
11	3.16	3.57	0.41	47	2.57	2.62	0.05	83	2.51	2.27	-0.24
12	3.21	3.62	0.40	48	3.33	3.38	0.05	84	2.82	2.57	-0.24
13	2.71	3.08	0.37	49	3.33	3.37	0.04	85	2.90	2.65	-0.24
14	3.26	3.63	0.37	50	3.52	3.56	0.04	86	2.70	2.44	-0.25
15	3.19	3.52	0.33	51	2.67	2.70	0.03	87	3.27	3.00	-0.27
16	3.51	3.84	0.33	52	2.65	2.68	0.03	88	2.77	2.49	-0.28
17	3.04	3.33	0.29	53	3.16	3.18	0.02	89	3.16	2.85	-0.31
18	3.35	3.64	0.28	54	2.99	3.00	0.01	90	2.82	2.50	-0.31
19	3.54	3.82	0.28	55	3.56	3.56	0.01	91	3.23	2.90	-0.33
20	2.49	2.77	0.28	56	2.74	2.73	0.00	92	2.94	2.60	-0.34
21	3.61	3.87	0.27	57	3.34	3.33	0.00	93	3.35	3.00	-0.35
22	3.21	3.47	0.26	58	3.13	3.13	-0.01	94	3.21	2.86	-0.36

23	3.45	3.70	0.25	59	3.45	3.43	-0.01	95	3.29	2.92	-0.37
24	3.61	3.86	0.25	60	3.50	3.49	-0.02	96	3.57	3.18	-0.39
25	3.51	3.76	0.25	61	3.45	3.42	-0.03	97	3.27	2.87	-0.40
26	2.41	2.65	0.23	62	2.57	2.54	-0.03	98	2.92	2.50	-0.42
27	3.33	3.56	0.23	63	3.57	3.53	-0.04	99	3.37	2.94	-0.43
28	3.30	3.53	0.23	64	3.63	3.59	-0.04	100	3.45	3.00	-0.45
29	3.29	3.52	0.23	65	2.87	2.80	-0.07	101	3.72	3.27	-0.45
30	3.62	3.83	0.21	66	2.36	2.29	-0.07	102	2.68	2.23	-0.45
31	3.14	3.35	0.21	67	3.28	3.21	-0.07	103	3.53	3.07	-0.46
32	2.78	2.99	0.21	68	3.48	3.40	-0.07	104	2.98	2.51	-0.47
33	3.16	3.36	0.20	69	3.32	3.23	-0.09	105	3.05	2.50	-0.55
34	3.00	3.19	0.19	70	3.28	3.17	-0.11	106	2.62	1.98	-0.64
35	3.49	3.68	0.19	71	2.77	2.66	-0.11	107	3.19	2.48	-0.71
36	3.62	3.81	0.18	72	3.39	3.26	-0.13	108	2.73	1.81	-0.92