ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 22-574

The Effect of Active Learning Professional Development Training on College Students' Academic Outcomes

Elizabeth S. Park
Westat

Di Xu
University of California, Irvine

Growing literature documents the promise of active learning instruction in engaging students in college classrooms. Accordingly, faculty professional development (PD) programs on active learning have become increasingly popular in postsecondary institutions; yet, quantitative evidence on the effectiveness of these programs is limited. Using administrative data and an individual fixed effects approach, we estimate the effect of an active learning PD program on student performance and persistence at a large public institution. Findings indicate that the training improved subsequent persistence in the same field. Using a subset of instructors whose instruction was observed by independent observers, we identify a positive association between training and implementation of active learning teaching practices. These findings provide suggestive evidence that active learning PD has the potential to improve student outcomes.

VERSION: May 2022

Suggested citation: Park, Elizabeth S., and Di Xu. (2022). The Effect of Active Learning Professional Development Training on College Students' Academic Outcomes . (EdWorkingPaper: 22-574). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/153n-cr12

The Effect of Active Learning Professional Development Training on College Students' Academic Outcomes

Elizabeth S. Park*

Westat

Di Xu

University of California, Irvine

Author Note

Acknowledgements

We thank Brian Sato, Andrea Aebersold and Mayank Verma for helpful conversations and feedback. Ashley Harlow, and Sabrina Solanki have also provided helpful suggestions on the manuscript. All mistakes are our own.

Funding

This work was supported by the National Science Foundation under grant number 1612258, 1750386, and 1565073.

* Corresponding author: Elizabeth S. Park, 1600 Research Blvd. Rockville, MD, 20850, lizpark@westat.com, @elizabethpark.

The Effect of Active Learning Professional Development Training on College Students' Academic Outcomes

Abstract

Growing literature documents the promise of active learning instruction in engaging students in college classrooms. Accordingly, faculty professional development (PD) programs on active learning have become increasingly popular in postsecondary institutions; yet, quantitative evidence on the effectiveness of these programs is limited. Using administrative data and an individual fixed effects approach, we estimate the effect of an active learning PD program on student performance and persistence at a large public institution. Findings indicate that the training improved subsequent persistence in the same field. Using a subset of instructors whose instruction was observed by independent observers, we identify a positive association between training and implementation of active learning teaching practices. These findings provide suggestive evidence that active learning PD has the potential to improve student outcomes.

Keywords: active learning; student outcomes; classroom observation

Between fall 2010 and fall 2018, undergraduate enrollment increased by approximately 3.4 million at four-year institutions, representing a 26% growth (National Center for Education Statistics, 2019). However, during the same period, the national retention rates of all full-time, first-year students attending four-year public institutions -- measured by the percentage of students who return to the same institution for their second year -- has been stagnant between 76% and 79% (National Student Clearinghouse, 2020). In response to the sizable number of students who withdraw from college within their first year, extensive literature has focused on classroom pedagogy. Specifically, researchers and policymakers raise the concern that the traditional lecture-intensive instruction that dominates college classrooms emphasizes memorization over conceptual learning and is thus "disengaging" for students (e.g., Braxton, Milem, & Sullivan, 2000; Deslauriers et al., 2011; Gasiewski et al., 2012; Pike, Smart, & Ethington, 2012; President's Council of Advisors on Science and Technology, 2012).

The ongoing conversation about the pitfalls of lecturing in teaching undergraduate-level courses has led to growing enthusiasm surrounding active learning instruction as a way to better engage students in their learning process (Deslauriers et al., 2011; Gasiewski et al., 2012; Prince, 2004; Wiggins et al., 2017). In contrast to lectures where students passively receive information, active learning emphasizes students' active participation through intentionally designed activities such as discussions, questions asked by the instructor, targeted in-class instructor feedback, in-class clicker questions, and small group active learning tasks or activities (Braxton, Milem, & Sullivan, 2000; Deslauriers et al., 2011).

Indeed, extant studies on active learning identified a positive association between active learning instruction and student engagement. For example, based on student surveys administered across 15 four-year institutions, Gasiewski et al. (2012) found that students reported greater engagement when taught with active learning instruction than lecture-intensive instruction. Perhaps as a result of increased engagement, students taught in active learning classrooms also display better class attendance, retain course materials longer, and perform better on exams compared to students taught in lecture-intensive classrooms across a number of disciplines (Cherney, 2008; Desauriers et al., 2011; Knight & Wood, 2005).

Given empirical evidence that active learning approaches may improve college student outcomes, faculty professional development (PD) programs on active learning instruction and practices have become increasingly popular at the postsecondary sector (Pfund et al., 2009). At the national level, the National Institute on Scientific Teaching offers multi-day workshops and various professional development opportunities for faculty to incorporate evidence-based teaching practices with support and guidance from expert facilitators. Moreover, institutions across the nation have teaching and learning centers that provide opportunities for faculty to participate in teaching institutes focused on active learning instruction strategies (i.e., University of Southern California's Faculty Teaching Institute, University of Georgia's Active Learning Summer Institute). Despite the rapid increase in training on active learning and the high hopes around it, there is a striking lack of empirical evidence on the effects of these PD programs on classroom instructional practices and on student academic outcomes.

This study addresses this research gap by examining the causal effect of an Active Learning Professional Development (ALPD) implemented during 2018-2020 at a large public institution. Specifically, we link ALPD instructor participation data with detailed student transcript data from all courses offered between fall 2016 and winter 2020 and estimate the effect of ALPD participation on students' current course performance as well as subsequent persistence and performance in the same field of study. To address instructor self-selection into ALPD and possible baseline differences between ALPD participants and nonparticipants, we leverage rich panel data and use an individual fixed effects approach that compares average student outcomes before and after the participants received the training while using non-participants to control for general contextual changes over time that may affect student outcomes. We further combine this approach with course fixed effects, thus ruling out any between-course variations in course difficulty and student outcomes. In addition, to construct a sample of non-participants who resemble the ALPD participants, we also conduct a robustness check where we first match ALPD participants with observationally similar non-participants and then estimate the individual fixed effects model with inverse probability weights based on the post-matched sample. To our knowledge, this is the first study that employs a quasi-experimental design to provide quantitative evidence on the benefits of active learning PD programs on student outcomes in the postsecondary setting.

Our analysis indicates that the ALPD improved the likelihood that a student persists into the next course within the same field by three percentage points, or a 5% increase. We do not observe any difference in student performance in the next course and marginal improvements in students' current course grades. We complement these estimates with in-depth classroom observation data of a subsample of 392 classrooms to shed light into the relationship between ALPD and instructional practices. Our results suggest that the ALPD participation is associated with an increased likelihood of using active learning approaches instead of lecture-intensive instruction. We couch these findings given the current wave of institutions seeking to scale up active learning instruction through institutionalized structures such as professional development.

Background

Studies conducted in a variety of disciplines document the benefits of active learning instruction on student engagement and performance relative to lecture-intensive instruction. Earlier studies that focus on student perceptions indicate that students in active learning classrooms perceive greater support from peers and faculty (Loes et al., 2017; Prince, 2004; Johnson, Johnson, & Smith, 1998) and feel empowered to take ownership of their learning relative to those in lecture-intensive instruction (Gasiewiski et al., 2012). As students engage with peers and faculty in active learning classrooms, students learn to cooperate with one another and improve communication skills (Johnson & Johnson, 2009). These studies contend that learning to collaborate with others and perceiving greater support are likely to lead to better learning outcomes.

A more recent and growing literature has directly examined the association between active learning and student performance outcomes, and has generally identified positive effects of active learning instruction relative to lecture-intensive instruction (e.g., Deslauriers et al., 2011; Deslauriers et al., 2019; Freeman et al., 2014; Ruiz-Primo et al., 2011; Theobald et al., 2020). For example, in their metaanalysis of 225 studies on the efficacy of active learning in STEM classrooms, Freeman et al. (2014) concluded that active learning increases course exam performance by approximately half of a letter grade (i.e., moving from a B to a B+) compared to lecture-intensive instruction. In addition, active learning has been associated with reduced equity gaps. For example, Theobald et al. (2020) conducted a metaanalysis of 15 studies across 51 STEM classrooms and found a 33% reduction in racial achievement gaps in student exam scores in active learning classrooms compared to lecture-intensive classrooms.

Despite the growing evidence for the promise of active learning instruction in engaging students, its implementation has yet to occur on a large scale in higher education (Stains et al., 2018). One of the reasons for instructors' suboptimal engagement in these practices is a lack of systematic pedagogical training that would enable faculty to apply these practices effectively to their own teaching (Brownell & Tanner, 2012; Mazur, 2009). In response, institutions and teaching institutes have begun to offer various training programs that provide space for cross-disciplinary faculty (and staff) to work together to discuss best practices and create a community of support within structured PD programs (Cox, 2004).

Yet, PD may not necessarily lead to alteration in instructional practices and improved student outcomes if the training is insufficient or if there is inadequate support and incentives for faculty to apply what they have learned to their own teaching. Indeed, the broad literature of teacher PD has increasingly emphasized the complex links between the design and implementation of PD and its effectiveness (Darling-Hammond et al., 2017; Elliot et al., 2017; Pelletreau et al., 2018 Penuel et al., 2007). For instance, the duration of the training, the quality of learning materials and activities, and the presence of collective participation and community support among PD participants can all influence the effect of a PD program (Desimone, 2009; Elliot et al., 2017). Accordingly, researchers have reached consensus that in order to fully understand the value of any particular PD training, it is critical to build an empirical knowledge base that documents how specific programs are designed and implemented, and assess their effectiveness in terms of both teaching practices and concrete student outcomes (Darling-Hammond et al., 2017; Kutaka et al, 2017).

Although PD on active learning pedagogies is increasingly popular in higher education, there is limited documentation of how these programs are designed, and even less is known about their impacts on instructional practices and student achievement. Among the handful of studies that describe such programs, there are substantial variations in the duration and community support across programs (e.g., Ebert-May et al., 2011; Pfund et al., 2009). In this study, we address this research gap about active learning PD training by documenting and rigorously evaluating an active learning PD program at a large, four-year public research university. We describe in detail how the program was designed and implemented, and also empirically assess the impact of the training on students' current course performance and downstream outcomes. Finally, based on detailed class observation data from a subgroup of courses in our sample, we also document the association between the PD training and instructional practices to illuminate possible mechanisms through which the training may influence student outcomes.

Research Design and Data

Program Description: Active Learning Professional Development Training

The active learning professional development (ALPD) under study was implemented at a large public research institution and is open to faculty across all disciplines and ranks. The program introduces faculty to active learning instructional methods and tools in a systematic way by involving participants in hands-on activities to analyze active learning pedagogical strategies and apply evidence-based practices to the participants' own lesson designs. The ALPD was officially launched in fall 2018. At the beginning of each term, several campus-wide emails were sent out to solicit faculty's participation. Because there was limited space available for each session, the program was offered on a first-come, first-serve basis, and the program was typically filled within one day. Since its inception until winter 2020, a total of 105 faculty have gone through the training.[1]

[1] As of Spring 2021, there were 278 instructors listed as either "in progress" or "completed." When we limit the sample to instructors with training dates and certified information the sample dropped to 105.

ALPD includes eight 90 minutes weekly sessions, through which faculty worked to revamp their own instructional materials to incorporate more active learning under the guidance of an expert facilitator from the Teaching and Learning Center on campus and in a supportive collegial atmosphere. Each session included a short lecture, assignments, and several readings related to topics in that session. Some of the key topics covered include the role of assessment, different forms of feedback, ways to increase inclusivity, linking course goals with assignments and activities, and leveraging technology. For each topic, faculty were first introduced to the general and discipline-specific literature underlying evidence-based practices in active learning, coupled with active learning strategies, instructional tools, and concrete examples that were specific to their discipline. The participants were then guided to apply evidence-based practices and implementation to their own course materials and lesson designs.

Community building and support was an important part of ALPD and several activities have been intentionally designed to facilitate community development. First, to ensure collaborative participation, all applicants to ALPD were required to commit to attending at least six sessions of the eight-session program. In addition, at each session, participants were assigned into small groups of four. The facilitator intentionally assigned participants into different groups for the first few weeks until participants had met everyone and could start selecting their own groups. Each session started with small group discussion, where participants shared their own personal teaching experiences. Finally, some of the assignments were intentionally designed to facilitate collaborative participation. For example, one assignment required participants to redesign a class period and present their work to the group, where all group members would provide feedback to the presenter.

At the end of the eight-week training, faculty received a certificate of completion if they were observed by independent classroom observers using the Classroom Observation Protocol of Undergraduate STEM (COPUS) (Smith et al., 2013).[2] Using the COPUS protocol, observers recorded both the instructor's and students' behaviors in each two-minute period of a class session. Specifically, trained observers identified what the instructor did using 12 instructor behavioral codes (i.e., lecturing, answering student questions) and what the students did using 13 student behavioral codes (i.e., listening, asking questions) in each two-minute interval.[3] ALPD trained instructors were awarded a certificate of completion if the observer confirmed that the instructor lectured less than 50% of the class period and incorporated instructor-student interactions as well as student-student interactions. Some of the incentives for receiving the certificate include priority scheduling at the technology-enabled active learning classrooms, using the certificate as a second piece of evidence of instruction quality for tenure/promotion review, and the opportunity to help facilitate future ALPD. 42% of the 105 trained faculty in our data were certified whereas the remaining faculty completed just the training.

Data

We leverage three data sources to examine the effect of active learning professional development training on current and subsequent student outcomes. The first source of information comes from detailed administrative data. We first identified all courses offered between fall 2016 through winter 2020 and then pulled information on all students who took any of these courses taught by either ALPD participants or non-participants (N=1,022 courses). Among these observations, 36% of the student enrollments (n=54,130 unique students) were in

[2] Although the COPUS protocol was initially developed to observe STEM classroom instructions, it has also been used in observing non-STEM classrooms (Denaro et al., 2021).

[3] For example if the observer tallies 13 times that the instructor lectured during a 50-minute course, we would say that the instructor lectured 52% of class time (13/25).

courses that have within-course variations in instructors by ALPD participation (i.e. courses with some sections taught by ALPD participants and some sections by non-participants).[4] We exclude summer terms partly because fewer courses are offered during the summer and partly because summer courses follow different lengths and possibly different structure than courses offered in the fall, winter, and spring terms. In addition, given the goal of this study, we exclude courses that are not instruction focused, such as independent study, undergraduate research courses, and lab sessions. The data include the name of the course, term-year in which it was offered, class size, the class location, and the primary instructor of record, as well as student demographic characteristics and prior academic achievement profiles. We next merge the administrative data with ALPD participation data which include a list of instructors who participated in the training and the term when each participant completed the training as well as their certification date if the instructor pursued a certification.

The third data source is classroom observation data. Starting in the 2018-2019 academic year, all undergraduate courses (i.e., excluding discussions, seminars, or labs) that enrolled at least 60 students at this institution were solicited to be observed using the COPUS protocol. A total of 392 classes across 289 unique instructors were observed by winter 2020. Among these instructors, 71 went through the ALPD training by the time of the observation while 218 instructors did not.

Student Outcomes

In assessing the effectiveness of ALPD, we consider both current and subsequent student academic outcomes in the same field of study. Specifically, we begin our inquiry with student

[4] Because of the way the data was obtained, we have a relatively large proportion of courses for which ALPD instructors did not teach the course. We have conducted analyses by limiting the sample to courses that were taught by both ALPD participants and non-participants and found that our results were similar regardless of this restriction.

contemporaneous course performance, as measured by course grade on a 1 to 4 grading scale. Yet, current course grades alone may not be sufficient in fully capturing the impact of ALPD for two reasons. First, current grades may not be a reliable measure of actual learning due to instructor grading leniency. Indeed, existing studies on teacher effectiveness indicate that students tend to receive lower grades in introductory coursework from instructors who are most effective in preparing students for subsequent advanced courses (Carrell & West, 2010). In the context of the current study, if ALPD also influences instructors' grading practices, a change in average student grades may not necessarily reflect actual improvement (or deterioration) of teaching quality.

In addition, the effects of ALPD may unfold in different ways and some of them may not show in immediate course performance. In particular, existing studies have advocated for the promise of active learning in promoting student interest in a subject area (e.g. Gasiewski et al., 2012), which arguably can be better captured through subsequent individual choices such as enrolling in another course in the same field of study than immediate performance outcomes.

Therefore, we build on the existing literature on teaching effectiveness (e.g., Carrell & West, 2010; Figlio et al., 2015; Xu, 2019; Xu & Solanki, 2020), and further include downstream outcomes to provide a more comprehensive understanding of the impact of ALPD. Specifically, we use subsequent field persistence — whether a student took another course in the same field of study in the immediate next term — to measure student interests in a subject. In addition, we also examine students' performance in the next course to capture possible lasting impacts of ALPD on learning and engagement in the same field of study.[5]

[5] To construct subsequent course achievement measures, we first looked at the entire course-taking records of each student and identified the next course within the same field for every course taken between fall 2016 to spring 2020 excluding summer terms. Repeat courses were excluded from next course persistence and performance.

Sample Description

Table 1 shows the summary statistics of the average outcome measures (panel A), characteristics of students (panel B), and characteristics of course-sections (panel C) taught by three groups of instructors: 1) instructors who never participated in the ALPD ("ALPD Non-participants" in column 1-3); 2) ALPD participants during the terms prior to the training ("ALPD Participants: Pre-Training" in column 4-6); and 3) ALPD participants during the terms after the training ("ALPD Participants: Post-Training" in column 7-9). Results presented in panel A reveal several baseline differences between ALPD participants and non-participants prior to the training, where the ALPD participants seem to be associated with consistently better student outcomes. This highlights the importance of accounting for the baseline differences between the ALPD participants and nonparticipants in estimating the impact of ALPD on student outcomes. Raw comparisons between the pre- and post-training periods among ALPD participants suggests that students' average grades in subsequent courses are higher in courses taught by ALPD participants during the post-training terms. Yet, these differences may be partly due to different courses taught by each group and the characteristics of students enrolled in those courses.

< Insert Table 1 >

Indeed, results presented in panels B and C revealed a number of differences in the type of courses taught and the characteristics of students between the ALPD participants and nonparticipants prior to the training, as well as between the pre-training and post-training terms among the ALPD participants. Specifically, compared with the nonparticipants, ALPD participants during the pre-training terms taught courses with a larger proportion of transfer students and students with lower high school GPA and SAT scores. The participants were also more likely to teach large classes (enrollment size>=60) than nonparticipants. Looking at ALPD participants' pre-training and post-training terms, ALPD participants during the post-training terms taught students with, on average, higher high school GPA and SAT scores than ALPD participants during the pre-training terms. The differences in student composition and class size may be partly driven by different fields of study between the participants and nonparticipants, where the participants seem to be more heavily concentrated in non-STEM disciplines.

In a similar vein, there are also noticeable differences in student characteristics between pre-training and post-training periods among the ALPD participants, where courses taught in the post-training terms had students with better pre-college academic performance. In addition, courses taught in the post-training periods were more heavily concentrated in STEM fields than in the pre-training periods. These differences may be partly driven by general changes in student composition as well as course offering over time at this institution, which highlight the importance of accounting for between-course and over-time variations in student outcomes in estimating the impact of the training on student outcomes.

To obtain an understanding of how our sample instructors compare to the university as a whole, Appendix Table 1 shows the comparison between ALPD participants and the population of instructors who taught during the study timeframe. Among the participants, there are fewer engineering instructors than the population as a whole. In terms of the teaching load and characteristics of courses taught, ALPD participants assumed a heavier teaching load and were more likely to teach undergraduate courses than graduate courses.[6] Finally, there are more assistant or associate professors and teaching-focused professors among the participant group relative to all instructors at this university.

[6] On average, the ALPD participants taught four credits per term whereas all instructors at this institution taught about two credits on average per term. In addition, ALPD participants taught fewer graduate courses and were more likely to teach undergraduate courses compared to the population of instructors.

Identification Strategy

To account for instructor self-selection into the PD training, we compare average student outcomes of ALPD participants after they completed the training to average outcomes of the same instructor before the training. This approach has been used widely in the education literature to address any time-invariant factors at the individual level such as ability in estimating the causal impact of educational investment (e.g., Cellini & Chaudhary, 2014; Jacobson, Lalonde, & Sullivan, 2005; Jepson, Troske, & Coomes, 2014; Xu & Trimble, 2016). In addition to individual fixed effects, we also include course fixed effects to compare average academic performance of students in different sections of the same course, as well as term fixed effects to account for general changes in student composition and outcomes over time. We estimate the following equation:

$$Y_{ijsct} = \alpha_j + \beta_1(ALPD_{jt}) + X_{ijsct}\beta + \theta_{sct}\pi + \gamma_c + \phi_t + \varepsilon_{ijsct}$$

where Y_{ijsct} is the outcome for student *i* taught by instructor *j* in section *s* of course *c* offered during term-year *t*. α_j refers to instructor fixed effects that control for all observed and unobserved instructor individual-level characteristics that are constant over time. $ALPD_{jt}$ captures whether an instructor has already received the ALPD training in a given term, which is identified as "0" during the terms leading up to when the instructor received the training and switches to "1" during the term after the instructor received the training and each term thereafter. X_{ijsct} includes student-level covariates such as students' race/ethnicity and high school GPA and θ_{sct} refers to section-level attributes such as enrollment size of a section. The equation also includes course fixed effects γ_c that control for any between-course variations in student composition and performance, and quarter-year fixed effects φ_t that help address overall fluctuations in student composition and outcomes over time due to other contextual factors. Accordingly, β_1 can be interpreted as additional changes in student performance as a result of ALPD training aside from other changes that would have occurred in the absence of the ALPD training.

When estimating students' subsequent course grade in the same field of study, we draw on prior literature and further include next class fixed effects (e.g., Figlio, Schapiro, & Soter, 2015; Ran & Xu, 2019). By doing so, we are able to compare grades of students in the same next class with variations in taking a prior course with an instructor who had received ALPD versus an instructor who had not. This is to address the concern that learning experiences in a course may influence a student's subsequent course choice. For example, if a student had particularly inspiring experiences with an ALPD instructor, the student may intentionally opt into another course taught by the same instructor. In a similar vein, prior experiences may also influence the difficulty of the next class a student selects. By including next section fixed effects, we are able to account for selection biases that arise from students shopping across different next classes within a field by comparing student performance in exactly the same section. All of our estimates are clustered at the instructor level to account for correlation in student outcomes within an instructor.

As a robustness check, we re-estimated all of the main results after matching ALPD participants with observationally similar non-participants through propensity score matching method. We first obtained data that include instructor-level pre-treatment characteristics and estimated the propensity score using a probit function. Then, we conducted nearest-neighbor matching with replacement to match control instructors with treated instructors based on observable characteristics. With this matched data, we estimated the treatment effect using the individual fixed effects model and inverse probability weights.

Results

Impact of ALPD Training on Student Outcomes

Table 2 presents the estimated effect of participating in ALPD on three student outcome measures: current course grade (column 1), whether or not a student took another class in the same field in the immediate next term (column 2), and the grade received in that next class (column 3). The results indicate that students who took a course with an instructor who had received ALPD on average had higher course grades by 0.006 grade points on a 0-4 grading scale, although this effect is marginally significant at the 0.1 level.

The effect of the ALPD becomes more pronounced when we examine students' subsequent persistence in the same field. Specifically, students who took a course with an ALPD instructor in post-training terms were three percentage points more likely to persist within the field compared to students who took the same course with the same instructor in pre-training terms. Considering that the average next course persistence rate in our sample is 68%, a three percentage point increase would represent a 5% improvement. Finally, column 3 presents results on subsequent course performance conditional on enrolling in another course in the same field of study. The estimated impact of ALPD is small and not significantly different from zero. Taken together, our results suggest that ALPD is associated with marginal improvements in current course performance and modest boost in field persistence.[7] [8]

<Insert Table 2>

Relationship between Training and Instructional Practices

In view of the positive effects of the ALPD training on student persistence, we then explore whether such benefit is partly driven by altering instructors' teaching practices.[9] We look at a subset of instructors whose classroom was observed using the COPUS protocol. When using the observation data to code instructional approaches, we follow the criteria used for ALPD certification in defining active learning classes -- lecturing less than 50 percent of the class period. Out of the 392 classes observed, 34 percent are classified as active learning.

< Insert Figure 1>

We see that the instructors in courses that are categorized as active learning were more likely to display varied activities during class (i.e., pose questions, move through class). Figure 1

---

[7] We also conducted analyses to see whether the effect of the training differs depending on the infrastructure (i.e., whether the course is offered in an active learning classroom), and the size of the class. Classroom layouts, for example, can allow for easier adoption of active learning techniques such as in-class group activities (Beicher & Saul, 2002; Dori & Belcher, 2005). As such, if the class is offered in a classroom designed to facilitate active learning, instructors may be more effective in raising student performance. In a similar vein, class size is an important consideration that determines whether active learning is adopted, with smaller class sizes being more conducive to implementing active learning (Carbone & Greenberg, 1998; Freeman et al., 2014; Heim & Holt, 2018). In both instances, we did not find that the impact of ALPD training is moderated by either the classroom infrastructure or class size. Yet, our sample size is small and we recognize that our findings may be due to a lack of power to deliver a precise estimate of any interaction effects.

[8] Appendix Table 2 shows a breakdown of the next course that students in our sample took. About half of the courses are lower division courses and the other half are upper division courses. In addition, 56% of the courses students took in the next term are non-STEM while 44% of the courses are STEM courses. 84% of the courses that students took as their next course were relatively small (under 100 student seats per class), indicating that persistence effects are concentrated in small classes.

[9] Starting in the 2018-2019 academic year, all undergraduate courses (i.e., excluding discussions, seminars, or labs) that enrolled at least 60 students at this institution were solicited to be observed using the COPUS protocol. A total of 250 classes between fall 2018 and winter 2020 were observed twice within the same term by independent observers affiliated with the Teaching and Learning Center and an additional 142 classes were observed once during this timeframe for a total of 392 classes. For classes that were observed twice, we averaged the classroom observation records.

shows the distribution of instructor activities performed during class time. The figures provide a visual contrast in instructor behaviors in lecture-intensive classes (figure on the left) versus active learning classes (figure on the right). Most notably, instructors in lecture-intensive classes spent 80 percent of two-minute intervals of observed class time lecturing whereas the corresponding percent is 25% of observed class time. In addition, instructors in active learning classes on average spent more than a quarter of the class time posing questions and moving through class compared to only 5 percent in lecture-intensive classes. This behavioral breakdown aligns with the prior literature on the characteristics of lecture-intensive versus active learning instruction (Braxton, Milem, & Sullivan, 2000; Deslauriers et al., 2011; Stains et al., 2013). Appendix Figure A.1 provides a snapshot of student behaviors in active learning instruction versus lecture-intensive instruction classrooms by showing the distribution of students' activities performed within each two-minutes interval of the observed class time.

Table 3 presents the estimated correlation between ALPD training and the likelihood of using active learning instruction. Due to the small sample size of instructors with class observation data, we conduct a cross-section comparison between instructors who had received ALPD by the time of the classroom observation and instructors who had not on their likelihood to implement active learning. Column 1 presents the raw comparison between the two groups while column 2 further includes available class-level covariations, such as student composition, enrollment size, field of study, and term-year fixed effects.

< insert Table 3 >

Our results indicate that ALPD trained instructors were 17 percentage points more likely to implement active learning instruction than non-ALPD trained instructors ($p < 0.001$) (column 1). This relationship remains significant at the 0.1 level after we further control for all available covariates (p = 0.053) (column 2).[10] The positive estimates across models provide suggestive evidence that the ALPD training increases the likelihood of implementing active learning instruction.

Is There Additional Value of Receiving a Certificate of Completion?

Our results thus far suggest that receiving the ALPD training is associated with small improvement in course performance and modest boost in subsequent field persistence. As mentioned in the background section, at the end of the training, all ALPD participants were offered the opportunity to receive a certification of completion upon successful classroom observation. The classroom observation and feedback associated with the certification process may create additional space for instructors to reflect on their practices and apply what they have learned to their own teaching. In addition, the certificate of active learning may also serve as a label which, in turn, may influence instructors' self-identity and behavior (Nelson et al., 2008; Hayes et al., 2020).

< Insert Table 4 >

To estimate the effect of receiving the certificate in addition to the ALPD training, we use the same empirical strategy but restrict our sample to the ALPD participants. Specifically, we flag a "0" for courses offered during the terms before an instructor receives the certificate and a "1" after receiving the certificate. Instructors who went through ALPD but never received the certificate would not contribute to the estimator directly, but could help take into account possible evolution of the training effects on student outcomes over time. The estimates presented

[10] We may be concerned that ALPD participants and non-participants differ in their instructional approaches even in the absence of the training. Accordingly, we further conduct a pre- versus post-training comparison among ALPD participants only and control for all available covariates. The estimated coefficient shown in Appendix Table 3 is positive (coefficient=0.13, p = 0.187) and fairly comparable to the estimate shown in Table 3 column 2 that is based on the cross-sectional comparison. However, since only 71 ALPD participants have pre and post classroom observation data, the sample size is too small to yield a precise estimate.

in Table 4 are consistently small and nonsignificant, suggesting that there is no additional boost induced by receiving the certification on students' outcomes. In other words, the positive impact of ALPD on current and downstream outcomes presented in Table 2 is primarily driven by participating in the training rather than going through any additional certification process.

Robustness Check

Our main empirical strategy compares the average student outcomes of the same instructor before and after the PD, while netting out time trends of non-participants using a fixed effects model. One potential threat to this model is possible presence of time-varying factors that are not captured in our data, such as the proclivity to adopt active-learning approaches overtime even in the absence of the ALPD training. To address this concern, we obtained additional instructor-level data and conducted a robustness check in which we used propensity score matching techniques to match non-participants with participants based on instructor-level pre-treatment characteristics to construct a control group of non-participants who resemble the ALPD participants along all of the observable characteristics. In estimating the propensity score, we included instructor department affiliation, instructor title and rank, and a number of pre-treatment teaching characteristics (i.e., prior to fall 2018) such as the number of upper division courses taught, the number of pre-requisite courses taught, the number of large versus small courses taught, the number of independent studies offered, and the number of graduate courses taught. We conducted k-nearest neighbor matching where k refers to the number of similar neighbors to which treated units will be matched, and in our case, we run the algorithm multiple times with $\{k = 1,2,3,4,5\}$. During the matching process, ALPD participants who had no near match from the group of non-participants (using a caliper of width equal to 0.1 standard deviation of the propensity score) were dropped from the analysis. Through each iteration of the matching process, we then checked whether we had succeeded in balancing the covariates and concluded that $k = 5$ resulted in the smallest standardized difference in means across the variables. Thus, the final matched sample includes 100 ALPD participants and 358 non-participants who are matched to the participants.

In Appendix Table 4 through 6, we show the pre-match and post-match standardized differences in means. Prior to matching instructors, we note several differences across teaching characteristics and instructor rank. For instance, ALPD instructors are more likely to be assistant or associate professors than the non-participant group. The matching algorithm achieved satisfactory overlap between the ALPD participants and non-participants, and improved the balance between the two groups along observable characteristics

Appendix Tables 7 and 8 show the estimated results using the individual fixed effects and inverse probability weights based on the post-matched sample. The estimated effects are similar in magnitude and direction as those from our main analyses. We see improvements in next course persistence among students who took a course with an ALPD participant than observationally similar non-participant. Moreover, these results also indicate no additional boost to receiving the certificate above and beyond receiving the training.

Limitation

While our study provides a first step in rigorously documenting the effect of an active learning PD on college student outcomes, our small sample size limits the conclusion we can draw on the positive effect of the PD on student outcomes. A follow-up study with a larger sample size will bolster the finding that the PD resulted in improved student outcomes.

Also, additional research is needed that explores the conditions for effective implementation of PD programs on active learning instruction at scale. For instance, an effective training program that is tailored to the specific needs of the program participants while also cognizant of the unique context of the institution may lose its effectiveness when taken at scale. Indeed, a number of studies conducted in K-12 settings have documented the challenges associated with scaling up effective small teacher PD programs (e.g. Kraft et al., 2018; Cabell et al., 2011). In addition, studies should examine different structures of active learning training (i.e., summer workshops offered by a national institute versus campus-level training) and whether certain training formats are more effective at improving student outcomes. Future studies that are able to document and relate elements of the large-scale training with changes to instructor practice and student outcomes will further complement the findings of the study.

Relatedly, designing and implementing a resource-intensive program such as the ALPD may be associated with high personnel costs of staffing experienced program facilitators. Accordingly, the effects of such programs need to be considered relative to program costs. The field would benefit from additional analyses benchmarking the program cost and the efficacy of the training. If intensive face-to-face guidance and interactions over a sustained amount of time is found to be at the core of effective models for active learning PD programs, then this approach in improving college classroom instruction is likely to require substantial financial investment. Although such costs should not be prohibitively expensive, it will be helpful to conduct cost-effectiveness analysis for systematic comparisons and evidence-based choice between different approaches to reforming college instruction.

Finally, it is important to note that our study only focuses on short-term academic outcomes while theories of active learning have underscored several non-academic and long-term benefits that are not fully captured in the current study, such as students' stress-level, test anxiety, development of social skills, long-term college and field persistence, and graduation rates (Ballen et al., 2017; Johnson, Johnson, & Smith, 2014; Loes et al., 2017). These possible benefits are important considerations and warrant attention in future studies to fully understand the effects of active learning training on student outcomes.

Discussion and Conclusion

The growing evidence on the promise of active learning instruction in engaging students has spurred increasing interests in promoting active learning approaches in the college classroom (Freeman et al., 2014; McKeachie et al., 1990; Prince, 2004; Pfund et al., 2009; Ruiz-Primo et al., 2011; Theobald et al., 2020). Despite the expansion of professional development efforts on active learning and the high hopes surrounding them, there is limited knowledge about the impacts of these programs on teaching practices and student achievement outcomes (Ebert-May et al., 2011). To address this gap, we leverage detailed college administrative data and program participation data and use a quasi-experimental design to estimate the impact of participating in an Active Learning Professional Development (ALPD) on students' contemporaneous and downstream outcomes.

Consistent with the existing literature on active learning, we find that ALPD is associated with an increase in marginal improvement in concurrent course grade; the estimated effect is small in magnitude and is only marginally significant at the 0.1 level. Yet, we also find that ALPD is associated with a more pronounced increase in subsequent persistence into another course in the same field-- a five percent improvement from the baseline persistence rate of 68%. Our subsequent exploratory analyses using classroom observation data reveal a positive association between ALPD and active learning teaching practices, providing suggestive evidence that the impact on students' outcomes may be driven by instructors' implementation of active learning approaches. Indeed, the magnitude of the effect on field persistence estimated in the present study corresponds to other studies that examined the relationship between active learning opportunities and downstream persistence outcomes. For example, Loes et al. (2017) found a five percentage point increase in second year college persistence when students are provided with more collaborative learning opportunities in the classroom. In a similar vein, Braxton et al. (2000) also identified a five percentage points increase in students' intent to re-enroll in the following term when comparing classrooms with high in-class discussions with low in-class discussions. Our results extend previous findings that professional development on active learning, by promoting the use of active learning approaches in the classroom, may increase students' persistence in the field. Accordingly, our results also highlight the importance of taking student subsequent outcomes into account when evaluating the effectiveness of any active learning PD training programs.

Our study is related to the broad literature on teacher professional development that underscores the complex relationship between specific program design features and the effectiveness of a program. These discussions have led to growing efforts in documenting implementation details of a PD program and assessing its impacts on student achievement outcomes across K-12 settings. Our study contributes to this literature by providing detailed description on how a successful active learning PD program was administered and implemented at the college setting, as well as assessing its effectiveness on student outcomes. From a theoretical perspective, there has been a growing consensus on several conditions under which PD programs might produce more favorable outcomes, including sustained duration, coherence, collective participation, active learning, and local support (Darling-Hammond et al., 2017). The

ALPD program in the current study combines several of these key features. For instance, the program involved highly committed and experienced program facilitators with a coherent training agenda, and a requirement of continuous and active participation from all participants during an eight-week time span. In addition, a number of activities were also intentionally designed to facilitate building a supportive and collaborative professional development community, aligning with existing literature on features of effective PD (Cox, 2001; Elliot et al., 2017). Therefore, our study complements the current literature that is primarily conducted at K-12 settings by lending support for incorporating these features in designing effective teacher PD programs on college instruction.

The findings from our study indicate that the professional development on active learning instruction may lead to increased persistence in the field through instructional improvement in the college classrooms. We encourage additional research to be conducted to bolster these results and to illuminate specific conditions under which such PD programs may produce favorable outcomes. Taken as a whole, our study lends suggestive support to PD programs on active learning as a promising way to innovate college instruction and improve student outcomes. As such, professional development may be a way to institutionalize the use of active learning in higher education.

References

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education, 16*(4), 1-6. https://doi.org/10.1187/cbe.16-12-0344

Beichner, R.B. & Saul, J.M. (2003). *Introduction to the SCALE-UP (student-centered activities for large enrollment undergraduate programs) project.* Proceedings of the International School of Physics, Varenna, Italy. https://projects.ncsu.edu/per//Articles/Varenna_SCALEUP_Paper.pdf

Braxton, J. M., Milem, J. F., & Sullivan, A. S. (2000). The influence of active learning on the college student departure process: Toward a revision of Tinto's theory. *The Journal of Higher Education*, *71*(5), 569-590. https://doi.org/10.2307/2649260

Brownell, S. E., & Tanner, K. D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and… tensions with professional identity?. *CBE—Life Sciences Education*, *11*(4), 339-346.

Cabell, S. Q., Justice, L. M., Piasta, S. B., Curenton, S. M., Wiggins, A., Turnbull, K. P., & Petscher, Y. (2011). The impact of teacher responsivity education on preschoolers' language and literacy skills. *American Journal of Speech-Language Pathology, 20*(4), 315–330. https://doi.org/10.1044/1058-0360(2011/10-0104)

Carbone, E., & Greenberg, J. (1998). Teaching large classes: Unpacking the problem and responding creatively. *To Improve the Academy*, *17*(1), 311-326.

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*(3), 409-432. https://doi.org/10.1086/653808

Cellini, S. R., & Chaudhary, L. (2014). The labor market returns to a for-profit college education. *Economics of Education Review*, *43*, 125-140. https://doi.org/10.1016/j.econedurev.2014.10.001

Cherney, I. D. (2008). The effects of active learning on students' memories for course content. *Active learning in higher education*, *9*(2), 152-171. https://doi.org/10.1177/1469787408090841

Cox, M. D. (2004). Introduction to faculty learning communities. *New directions for teaching and learning*, *2004*(97), 5-23. https://doi.org/10.1002/tl.129

Denaro, K., Sato, B., Harlow, A., Aebersold, A., & Verma, M. (2021). Comparison of cluster analysis methodologies for characterization of Cassroom Observation Protocol for Undergraduate STEM (COPUS) Data. *CBE—Life Sciences Education*, *20*(1), 1-11. https://doi.org/10.1187/cbe.20-04-0077

Darling-Hammond, L., Hyler, M.E., & Gardner, M. (2017). *Effective teacher professional development.* Palo Alto, CA: Learning Policy Institute. Retrieved from: https://files.eric.ed.gov/fulltext/ED606743.pdf

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181-199. https://doi.org/10.3102/0013189X08331140

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332(6031), 862-864. https://doi.org/10.1126/science.1201783

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus perceptions of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251-19257. https://doi.org/pnas.1821936116

Dori, Y.J. & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? *The Journal of*

The Learning Sciences, *14*(2), 243-279. http://web.mit.edu/edtech/casestudies/pdf/teal1.pdf

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. *BioScience*, *61*(7), 550-558. https://doi.org/10.1525/bio.2011.61.7.9

Elliott, E. R., Reason, R. D., Coffman, C. R., Gangloff, E. J., Raker, J. R., Powell-Coffman, J. A., & Ogilvie, C. A. (2016). Improved student learning through a faculty learning community: How faculty collaboration transformed a large-enrollment course from lecture to student centered. *CBE—Life Sciences Education*, *15*(2), ar22. https://doi.org/10.1187/cbe.14-07-0112

Figlio, D. N., Schapiro, M. O., & Soter, K. B. (2015). Are tenure track professors better teachers?. *Review of Economics and Statistics*, *97*(4), 715-724.

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. *Proceedings of the National Academy of Sciences*, *111* (23), 8410-8415. https://doi.org/pnas.1319030111

Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., & Chang, M. J. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education*, *53*(2), 229-261.

Heim, A. B., & Holt, E. A. (2018). Comparing student, instructor, and expert perceptions of learner-centeredness in post-secondary biology classrooms. *PloS One, 13*(7), e0200524. https://doi.org/10.1371/journal.pone.0200524

Jacobson, L., LaLonde, R. J., & Sullivan, D. (2005). The impact of community college retraining on older displaced workers: Should we teach old dogs new tricks?. *ILR Review*, *58*(3), 398-415. https://www.jstor.org/stable/30038595

Jepsen, C., Troske, K., & Coomes, P. (2014). The labor-market returns to community college degrees, diplomas, and certificates. *Journal of Labor Economics*, *32*(1), 95-121. https://doi.org/10.1086/671809

Johnson, D.W., Johnson, R.T., & Smith K. (1998). Cooperative learning returns to college: What evidence is there that it works?, *Change*, 27-35

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365-379. https://doi.org/10.3102/0013189X09339057

Knight, Jennifer K., and William B. Wood. "Teaching more by lecturing less." *Cell Biology Education* 2005(4), 298-310. https://doi.org/10.1187/05-06-0082

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547-588. https://doi.org/10.3102/0034654318759268

Kutaka, T. S., Smith, W. M., Albano, A. D., Edwards, C. P., Ren, L., Beattie, H. L., ... & Stroup, W. W. (2017). Connecting teacher professional development and student mathematics achievement: a 4-year study of an elementary mathematics specialist program. *Journal of Teacher Education*, *68*(2), 140-154. https://doi.org/10.1177/0022487116687551

Loes, C. N., An, B. P., Saichaie, K., & Pascarella, E. T. (2017). Does collaborative learning influence persistence to the second year of college?. *The Journal of Higher Education, 88*(1), 62-84. https://doi.org/10.1080/00221546.2016.1243942

Mazur, E. (2009). Farewell, lecture. *Science*, *323*(5910), 50-51.

McKeachie, W. J. (1990). Research on college teaching: The historical background. *Journal of Educational Psychology*, *82*(2), 189–200. https://doi.org/10.1037/0022-0663.82.2.189

National Center for Education Statistics. (2019). *The condition of education.* Retrieved from: https://nces.ed.gov/programs/coe/indicator_cha.asp#:~:text=2000%20through%202029.-,See%20Digest%20of%20Education%20Statistics%202019%2C%20table%20303.70

National Student Clearinghouse. (2020). *Persistence & retention – 2019*. Retrieved from: https://nscresearchcenter.org/snapshotreport35-first-year-persistence-and-retention/

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*(4), 921-958.

Pelletreau, K. N., Knight, J. K., Lemons, P. P., McCourt, J. S., Merrill, J. E., Nehm, R. H., ... & Smith, M. K. (2018). A faculty professional development model that improves student learning, encourages active-learning instructional practices, and works for faculty at multiple institutions. *CBE—Life Sciences Education*, *17*(2), es5. https://doi.org/10.1187/cbe.17-12-0260

Pfund, C., Miller, S., Brenner, K., Bruns, P., Chang, A., Ebert-May, D., ... & Handelsman, J. (2009). Summer institute to improve university science teaching. *Science*, *324*(5926), 470-471.

Pike, G. R., Smart, J. C., & Ethington, C. A. (2012). The mediating effects of student engagement on the relationships between academic disciplines and learning outcomes: An extension of Holland's theory. *Research in Higher Education*, *53*(5), 550–575. https://doi.org/10.1007/s11162-011-9239-y

President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Retrieved from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223-231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Ran, F. X., & Xu, D. (2019). Does contractual form matter? The impact of different types of non-tenure-track faculty on college students' academic outcomes. *Journal of Human Resources, 54*(4), 1081-1120. https://doi.org/10.3368/jhr.54.4.0117.8505R

Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, *331*(6022), 1269-1270. https://doi.org/10.1126/science.1198976

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education, 12*(4), 618-627. https://doi.org/10.1187/lse.13.2.359

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., ... & Levis-Fitzgerald, M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468-1470. https://doi.org/10.1126/science.aap8892

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Grummer, J. A. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476-6483. https://doi.org/pnas.1916903117

Wiggins, B. L., Eddy, S. L., Wener-Fligner, L., Freisem, K., Grunspan, D. Z., Theobald, E. J., ... & Crowe, A. J. (2017). ASPECT: A survey to assess student perspective of engagement in an active-learning classroom. *CBE—Life Sciences Education*, *16*(2), 1-13.

Xu, D. (2019). Academic performance in community colleges: The influences of part-time and full-time instructors. American Educational Research Journal, 56(2), 368-406. https://doi.org/10.3102/0002831218796131

Xu, D., & Solanki, S. (2020). Tenure-track appointment for teaching-oriented faculty? The impact of teaching and research faculty on student outcomes. Educational Evaluation and Policy Analysis, 42(1), 66-86.https://doi.org/10.3102/0162373719882706

Xu, D., & Trimble, M. (2016). What about certificates? Evidence on the labor market returns to nondegree community college awards in two states. Educational Evaluation and Policy Analysis, 38(2), 272-292. https://doi.org/10.3102/0162373715617827

Table 1

Summary Statistics

	(1)	(2)	(4)	(5)	(7)	(8)		
	ALPD Non-participants		ALPD Participants: Pre-Training		ALPD Participants: Post-Training			
	M or %	SD	M or %	SD	M or %	SD	Pre vs. Post SMD	Pre-ALPD vs. Non-ALPD SMD
Panel A. Student Outcomes								
Current course grade	2.95	1.02	3.08	0.97	3.13	0.96	-0.05	0.13
Next course (%)	0.68	0.47	0.75	0.43	0.73	0.44	0.04	0.07
Next course grade	3.03	0.99	3.06	0.99	3.20	0.98	-0.15	0.03
Panel B. Student-Level Characteristics								
Women (%)	0.52	0.50	0.51	0.50	0.49	0.50	0.03	-0.01
Black (%)	0.03	0.17	0.03	0.18	0.04	0.19	-0.03	0.00
Latinx (%)	0.25	0.43	0.26	0.44	0.26	0.44	0.00	0.01
AAPI (%)	0.56	0.50	0.54	0.50	0.54	0.50	0.01	-0.01
White (%)	0.13	0.34	0.14	0.34	0.13	0.34	0.01	0.01
Other (%)	0.03	0.18	0.03	0.17	0.04	0.18	-0.02	-0.01
URM (%)	0.48	0.50	0.49	0.50	0.50	0.50	-0.01	0.01
Transfer student (%)	0.17	0.37	0.20	0.40	0.22	0.41	-0.04	0.05
First-generation (%)	0.49	0.50	0.51	0.50	0.49	0.50	0.04	0.02
Low-income (%)	0.32	0.47	0.34	0.47	0.32	0.47	0.05	0.02
Weighted HS GPA	3.89	0.35	3.86	0.34	3.89	0.38	-0.07	-0.08
SAT Math	633.77	95.85	624.65	95.81	631.53	100.68	-0.07	-0.10
SAT Verbal	572.39	93.15	568.69	93.46	581.54	95.10	-0.14	-0.04
Panel C. Course-Section Level Characteristics								
STEM (%)	0.52	0.50	0.43	0.50	0.49	0.50	-0.11	-0.09
Offered in an Active Learning Classroom (%)	0.08	0.27	0.04	0.21	0.24	0.42	-0.57	-0.07

Small class (Fewer than 61 seats) (%)	0.22	0.42	0.15	0.36	0.17	0.38	-0.06	-0.10
Instructors	1359		75		30			
Course-by-Term	7568		878		289			
Observations	589603		80195		27917	697715		

Note. SMD = Standardized difference in means. AAPI = Asian American and Pacific Islanders; URM = Underrepresented Racial Minorities defined as Black, Latinx, Southeast Asians/Pacific Islanders, and Native Americans. The sample was limited to courses that were offered during fall 2016 to winter 2020, excluding summer terms. Courses that have fewer than 20 students and directed research/independent study courses were excluded in all analyses. Only those who took another course in the same field are observable for next course grade. Courses that were taken as a repeat course were not considered in determining next course persistence or grades.

Table 2

Effect of the ALPD on Student Outcomes

	(1) Course Grade	(2) Next Course Persistence	(3) Next Course Grade
ALPD Training	0.006+	0.032**	-0.016
	(0.003)	(0.012)	(0.012)
Instructor FE	Yes	Yes	Yes
Next Section FE	No	No	Yes
R^2	0.150	0.160	0.060
Instructors	1464	1464	1451
Student-by-Section-Term Observations	697715	697715	478505
Average student outcomes taught by instructors without training	2.98	0.689	3.033

Note. ALPD = Active Learning Professional Development. The sample was limited to courses that were offered during fall 2016 to winter 2020, excluding summer terms. Courses that have fewer than 20 students and directed research/independent study courses were excluded in all analyses. All models include course fixed effects, entry term fixed effects, and term-year fixed effects. Next course grade analysis further includes next section fixed effects. Student-level covariates include students' race, gender, transfer status, low-income status, first-generation status, SAT math, SAT verbal, and weighted HS GPA. We also include the number of students in the course to account for class size, the proportion of transfer students, first-generation college students, low-income students, and women in the course, and the average high school GPA and course grades of students in the course. Only those who took another course in the field has next course grade. Courses that are taken as a repeat course are not considered in the calculation. Standard errors are clustered at the instructor level.

*+ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$*

Table 3

Likelihood to Implement Active Learning Approaches

	(1) No Covariates	(2) With Term FE
ALPD Instructor	0.171**	0.105+
	(0.054)	(0.054)
Average Course Grade		0.170**
		(0.049)
Average % of Low-Income Students		0.384
		(0.371)
Average % of First-Gen Students		0.265
		(0.301)
Average % of Women		0.191
		(0.145)
Average % of RM Students		-0.467+
		(0.243)
Average % of Transfer Students		-0.249
		(0.205)
Average HS Unweighted GPA		-0.029
		(0.033)
Indicator of STEM Course		0.014
		(0.055)
Class Size		-0.001**
		(0.000)
Term Fixed Effects		X
Constant	0.300**	0.017
	(0.028)	(0.249)
R^2	0.025	0.195
Instructors	289	289
Section-by-term	392	392

Note. ALPD = Active Learning Professional Development. Some instructors were observed twice. Courses were selected based on the following criteria: large classrooms (84+ seats) and lecture halls. Graduate courses and undergraduate discussions sections were excluded. Instructors were given the option to opt-out. There are 71 instructors who completed the training in this data and 218 instructors who did not complete the training.

*+ p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001*

Table 4

The Effect of Certification on Student Outcomes

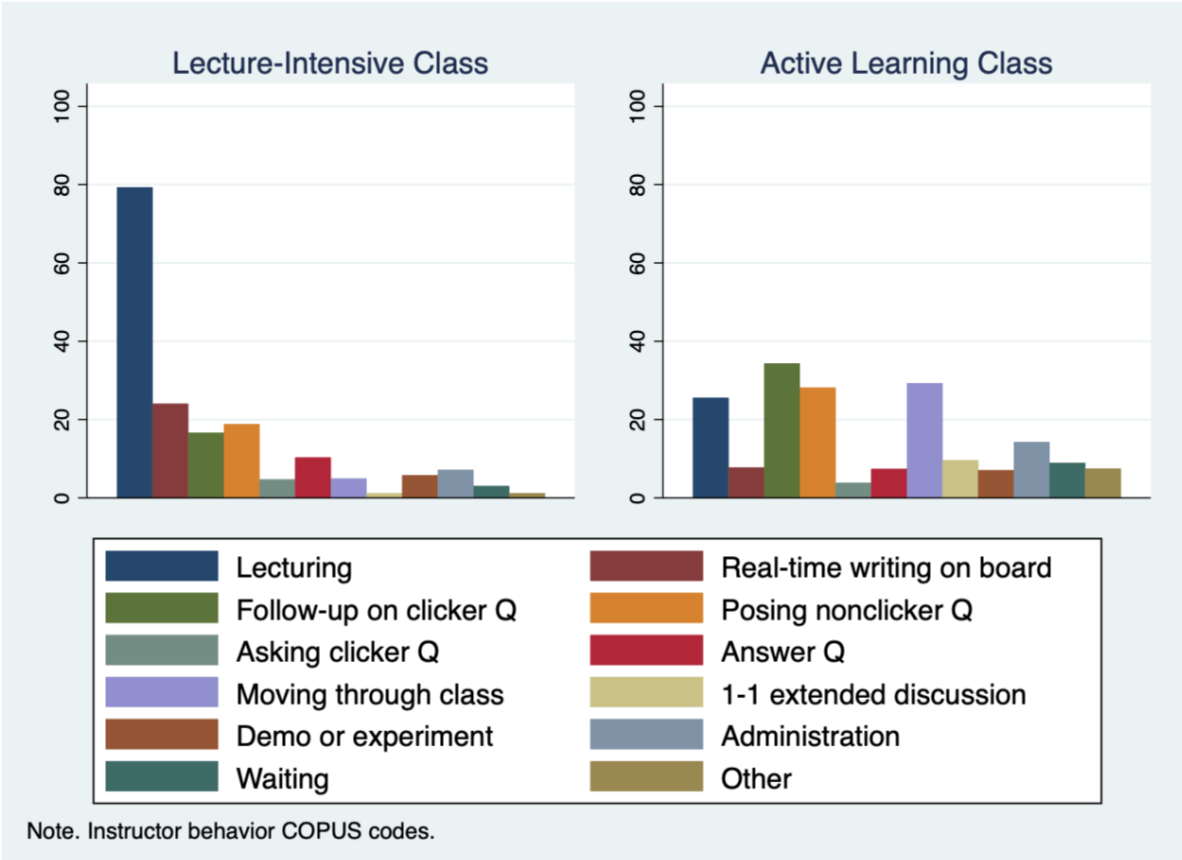
	(1) Course Grade	(2) Next Course Persistence	(3) Next Course Grade
ALPD Certified	-0.000	-0.003	0.043
	(0.009)	(0.034)	(0.034)
Instructor FE	Yes	Yes	Yes
Next Section FE	No	No	Yes
R^2	0.197	0.200	0.065
Instructor	89	89	89
Student-by-Course-Term Observations	41926	41926	30783

Note. ALPD = Active Learning Professional Development. The sample was restricted to 89 instructors who received ALPD training. The instructors who are not observable in the data for the post-trained period (i.e., did not teach after getting trained) or did not receive certification were removed from the sample. 45 of the 89 trained instructors who were deemed to have lectured less than 50% were certified. Therefore, for the 45 certified instructors, we changed their post-period to the term when they were certified. For the remaining instructors who were trained but not certified, their pre-trained periods were removed.

*+ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$*

Figure 1

Distribution of Instructor Behavior in Active Learning and Lecture-Intensive Class



Note. Active learning classes are defined as classes where the instructor was categorized as lecturing less than 50% of class time according to the COPUS codes. The graph depicts instructor behaviors only. n=289 instructors across 392 course-sections.

Supplemental Materials for:
The Effect of Active Learning Professional Development Training on College Students' Academic Outcomes

Appendix Table 1
Generalizability to all Instructors at the Institution

	All Instructors (n=6,432)		ALPD Instructors (n=105)	
	M or %	SD	M	SD
Administrative	6%		7%	
Biology	7%		7%	
Business	4%		1%	
Education	3%		4%	
Engineering	11%		8%	
Humanities	19%		20%	
Informatics and Computer Science	9%		9%	
Medicine	8%		3%	
Physical Sciences	15%		16%	
Social Ecology	1%		2%	
Social Sciences	5%		4%	
Professor	12%		16%	
Associate Professor	5%		22%	
Assistant Professor	5%		20%	
Teaching Professor	0%		1%	
Teaching Associate Professor	0%		1%	
Teaching Assistant Professor	1%		6%	
Lecturer	12%		24%	
Other	66%		10%	
Average Number of Credits Per Term	1.58	2.55	3.21	1.96
Proportion of Credits that are Upper Division Per Term	67%	0.45	67%	0.43
Proportion of Credits that are Graduate Courses Per Term	45%	0.44	34%	0.38
Proportion of Credits that are Independent Study/Research Per Term	8%	0.23	7%	0.17
Average Class Size Per Term	39.91	54.40	33.92	43.07

Note. These are teaching characteristics of all instructors who ever taught between Winter 2016 through Winter 2020. Summer terms are excluded.

Appendix Table 2

Next Section Characteristics

	University Sections (n=30,420)	Analytic Sample (n=20,909)
Biology	12%	10%
Chemistry	6%	7%
Computer Science	3%	3%
Engineering	8%	6%
Mathematics	4%	6%
Physics	6%	6%
Other STEM	5%	6%
Non-STEM	55%	56%
Small Class (<=100 students)	89%	84%
Lower Division	38%	49%
Upper Division	62%	51%

Note. Next section spans from winter 2017 through spring 2014. We used student-level transcript data to calculate next course within the same field. We excluded summer terms and graduate courses.

Appendix Table 3

Likelihood to Implement Active Learning Approaches, ALPD Instructor Subsample

	(1) No Covariates	(2) With Term FE
Post Trained Period	0.110	0.129
	(0.101)	(0.097)
Average Course Grade		0.131
		(0.100)
Average % of Low-Income Students		0.452
		(0.781)
Average % of First-Gen Students		0.373
		(0.658)
Average % of Women		-0.333
		(0.285)
Average % of RM Students		-0.496
		(0.506)
Average % of Transfer Students		1.039*
		(0.422)
Average HS Unweighted GPA		0.114+
		(0.066)
Indicator of STEM Course		-0.036
		(0.111)
Class size		-0.002**
		(0.001)
Constant	0.426**	-0.151
	(0.064)	(0.492)
R-sq	0.012	0.229
Instructors	71	71
Section-by-term	102	102

Note. Some instructors were observed twice. Courses that were observed were selected based on the following criteria: large classrooms (60+ seats) and lecture halls. Graduate courses and undergraduate discussions sections were excluded. Post-period is the period after instructors completed the training. There are 71 instructors who completed the training and 102 classes observed in the data.

Appendix Table 4

Pre-Matched Summary Statistics

	Treated (n=102)		Control (n=1268)		Pre-Match SMD
	Mean	SD	Mean	SD	
Administrative	0.059		0.035		0.114
Biology	0.059		0.052		0.030
Business	0.020		0.021		-0.012
Education	0.029		0.030		-0.003
Engineering	0.078		0.080		-0.007
Humanities	0.206		0.361		-0.350
Informatics and Computer Science	0.088		0.057		0.122
Medicine	0.049		0.027		0.116
Physical Sciences	0.167		0.160		0.018
Social Ecology	0.039		0.062		-0.105
Social Sciences	0.206		0.114		0.252
Professor	0.186		0.221		-0.086
Associate Professor	0.176		0.111		0.187
Assistant Professor	0.206		0.095		0.313
Teaching Professor	0.010		0.006		0.039
Teaching Associate Professor	0.020		0.010		0.077
Teaching Assistant Professor	0.069		0.023		0.220
Lecturer	0.225		0.256		-0.070
Other	0.108		0.278		-0.441
Number of Upper Division	3.108	3.303	2.312	3.446	0.236
Number of Prerequisites	2.706	4.418	1.477	2.406	0.345
Number of Large Classes (>100 students)	1.696	3.102	0.899	1.899	0.310
Number of Independent Study Offered	1.667	2.748	1.222	2.430	0.172
Number of Graduate Courses	0.529	1.355	0.489	1.254	0.031

Note. SMD = Standardized Difference in Means. The treatment instructor count slightly differs from the initial 105 participants in the main analysis. This is due to the discrepancy in the common identifier obtained in the instructor data and the original dataset. There were three instructors in our original data who we could match in the instructor dataset.

Appendix Table 5

Post-Matched Summary Statistics

	Treated (n=100)		Control (n=358)		Post-Match SMD
	Mean	SD	Mean	SD	
Department					
Administrative	0.05		0.062	0.241493	-0.052
Biology	0.06		0.044	0.205382	0.072
Business	0.02		0.018	0.133137	0.015
Education	0.03		0.034	0.181483	-0.023
Engineering	0.08		0.08	0.271673	0.000
Humanities	0.2		0.182	0.386385	0.046
Informatics and Computer Science	0.09		0.094	0.292237	-0.014
Medicine	0.05		0.07	0.255504	-0.084
Physical Sciences	0.17		0.186	0.389651	-0.042
Social Ecology	0.04		0.062	0.241493	-0.100
Social Sciences	0.21		0.168	0.37439	0.107
Professor	0.19		0.208	0.406445	-0.045
Associate Professor	0.18		0.192	0.394424	-0.031
Assistant Professor	0.2		0.266	0.442483	-0.157
Teaching Professor	0.01		0.012	0.109038	-0.019
Teaching Associate Professor	0.02		0.016	0.125651	0.030
Teaching Assistant Professor	0.06		0.022	0.146889	0.193
Lecturer	0.23		0.184	0.388027	0.114
Other	0.11		0.1	0.30042	0.033
Number of Upper Division	3.11	3.333015	3.188	4.676939	-0.019
Number of Prerequisites	2.32	2.970835	2.066	2.809929	0.088
Number of Large Classes (>100 students)	1.7	3.125328	1.186	2.195648	0.190
Number of Independent Study Offered	1.67	2.767251	1.69	2.742793	-0.007
Number of Graduate Courses	0.53	1.366667	0.556	1.264647	-0.020

Note. SMD = Standardized Difference in Means. Two instructors were further dropped due to the lack of common support resulting in 100 participants. K=5 and caliper is defined as 0.1 standard deviation of the propensity score.

Appendix Table 6

Summary Statistics of K-N Nearest Neighbor Matched Sample

	(1)	(2)	(4)	(5)	(7)	(8)		
	ALPD Non-participants		ALPD Participants: Pre-Training		ALPD Participants: Post-Training			
	M or %	SD	M or %	SD	M or %	SD	Pre vs. Post SMD	Pre-ALPD vs. Non-ALPD SMD
Panel A. Student Outcomes								
Current course grade	2.896	1.030	3.086	0.965	3.157	0.941	-0.075	-0.190
Next course (%)	0.707	0.455	0.758	0.428	0.754	0.431	0.009	-0.115
Next course grade	3.002	1.002	3.042	1.001	3.195	0.981	-0.155	-0.039
Panel B. Student-Level Characteristics								
Women (%)	0.529	0.499	0.510	0.500	0.495	0.500	0.031	0.038
Black (%)	0.033	0.178	0.031	0.174	0.038	0.190	-0.034	0.007
Latinx (%)	0.260	0.439	0.258	0.438	0.254	0.436	0.008	0.005
AAPI (%)	0.539	0.498	0.543	0.498	0.539	0.499	0.009	-0.008
White (%)	0.138	0.345	0.137	0.344	0.134	0.340	0.009	0.003
Other (%)	0.030	0.172	0.031	0.173	0.036	0.185	-0.027	-0.002
URM (%)	0.508	0.500	0.496	0.500	0.495	0.500	0.002	0.025
Transfer student (%)	0.166	0.372	0.209	0.406	0.232	0.422	-0.057	-0.110
First-generation (%)	0.496	0.500	0.508	0.500	0.490	0.500	0.036	-0.024
Low-income (%)	0.331	0.470	0.343	0.475	0.318	0.466	0.052	-0.026
Weighted HS GPA	3.888	0.358	3.856	0.347	3.880	0.385	-0.067	0.093
SAT Math	628.967	94.957	624.052	95.710	631.213	100.011	-0.073	0.052
SAT Verbal	574.619	92.133	568.875	93.435	581.992	95.045	-0.139	0.062
Panel C. Course-Section Level Characteristics								

STEM (%)	0.608	0.488	0.438	0.496	0.504	0.500	-0.133	0.346
Offered in an Active Learning Classroom (%)	0.065	0.247	0.046	0.210	0.229	0.420	-0.551	0.081
Small class (Fewer than 61 seats) (%)	0.232	0.422	0.138	0.345	0.150	0.357	-0.036	0.245
Instructors	358		68		32			
Course-by-Term	2821		760		250			
Observations	211048		72738		25323			

Note. SMD=Standardized difference in means. AAPI = Asian American and Pacific Islanders; URM = Underrepresented Racial Minorities defined as Black, Latinx, Southeast Asians/Pacific Islanders, and Native Americans. The sample was limited to courses that were offered during fall 2016 to winter 2020, excluding summer terms. Courses that have fewer than 20 students and directed research/independent study courses were excluded in all analyses. Only those who took another course in the same field are observable for next course grade. Courses that were taken as a repeat course were not considered in determining next course persistence or grades.

Appendix Table 7

Effect of the ALPD Training on Student Outcomes using Post-Matched Sample and Inverse Probability Weights

	Course Grade	Next Course Persistence	Next Course Grade
ALPD Trained	0.003	0.030*	-0.011
	(0.004)	(0.014)	(0.012)
Instructor FE	Yes	Yes	Yes
Next Section FE	No	No	Yes
R2	0.164	0.154	0.065
Instructors	397	397	396
Student-by-Section-Term Observations	309109	309109	221926

Note. ALPD = Active Learning Professional Development. The sample was limited to courses that were offered during fall 2016 to winter 2020, excluding summer terms. Courses that have fewer than 20 students and directed research/independent study courses were excluded in all analyses. All models include course fixed effects, entry term fixed effects, and term-year fixed effects. Next course grade analysis further includes next section fixed effects. Student-level covariates include students' race, gender, transfer status, low-income status, first-generation status, SAT math, SAT verbal, and weighted HS GPA. We also include the number of students in the course to account for class size, the proportion of transfer students, first-generation college students, low-income students, and women in the course, and the average high school GPA and course grades of students in the course. Only those who took another course in the field has next course grade. Courses that are taken as a repeat course are not considered in the calculation. Standard errors are clustered at the instructor level.

*+ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$*

Appendix Table 8

Effect of Certification on Student Outcomes using Post-Matched Sample and Inverse Probability Weights

	(1) Course Grade	(2) Next Course Persistence	(3) Next Course Grade
ALPD Certified	-0.002	-0.002	0.054
	(0.010)	(0.035)	(0.034)
Instructor FE	Yes	Yes	Yes
Next Section FE	No	No	Yes
R^2	0.189	0.187	0.066
Student-by-Course-Term Observations	39086	39086	29291

Note. ALPD = Active Learning Professional Development. The sample was restricted to 86 instructors who received ALPD training. The instructors who are not observable in the data for the post-trained period (i.e., did not teach after getting trained) or did not receive certification were removed from the sample.

*+ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$*

Appendix Table 9

Student and Instructor Fixed Effects Results using Post-Matched Sample and Inverse Probability Weights

	(1) Course Grade	(2) Next Course Persistence	(3) Next Course Grade
ALPD Trained	0.002	0.018	0.008
	(0.013)	(0.012)	(0.013)
Instructor FE	Yes	Yes	Yes
Course FE	Yes	Yes	Yes
Next Section FE	No	No	Yes
Student FE	Yes	Yes	Yes
R2	0.612	0.470	0.672
Student-by-Section-Term Observations	261046	261046	175896

Note. ALPD = Active Learning Professional Development. The sample was limited to post-matched sample.

*+ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$*