SOCIAL SCIENCES

Early-career faculty practice learner-centered teaching up to 9 years after postdoctoral professional development

Nathan C. Emery[1]*, Jessica Middlemis Maher[2], Diane Ebert-May[1]

The goal of STEM professional development for teaching is that participants continue to practice what they learn in the long term. However, we do not know if the outcomes are achieved and ultimately persist. We tracked postdoctoral participants from the Faculty Institutes for Reforming Science Teaching (FIRST) IV program into their current positions as early-career biology faculty. We assessed their teaching approaches, practices, and student perceptions of the learning environment over 6 to 9 years after finishing the program. Simultaneously, we evaluated paired faculty in the same departments. We found that professional development outcomes persisted over time and across a career transition. FIRST IV faculty maintained their learner-centered practices and were more learner-centered than their peers. Last, we found that teaching approaches were correlated with teaching practices in all faculty participants. These results provide evidence for the success of the FIRST IV program and the long-term persistence of professional development outcomes.

INTRODUCTION

Teaching professional development (PD) programs have proliferated during the last few decades, with the goal of facilitating measurable change in STEM (science, technology, engineering, and mathematics) higher education. However, we know little about the long-term outcomes of faculty PD efforts (1–4). Given the substantial investment in time and resources that these PD programs represent, it is critical that we assess their longitudinal impact on instructional practices and additional program outcomes.

Many teaching PD programs in STEM focus on convincing instructors to change the way they teach (5–8), by emphasizing evidence-based practices that support student learning through collaboration and inquiry. Often, achieving this goal requires shifting faculty away from teacher-centered approaches and toward learner-centered teaching, whereby students engage in constructing knowledge and conceptual learning (9). The expected outcomes of teaching PD are transformed learner-centered courses (10) and changes in faculty attitudes, approaches, and practices in the classroom (3, 4, 11). Approaches to teaching revolve around the instructor's intentions and strategies for teaching a given course (12), while teaching practices are how these strategies manifest in classroom settings (13).

Evaluating these outcomes is essential to determining how teaching PD affects the larger landscape of STEM higher education (14). Several large-scale PD programs reported changes in faculty soon after program completion (3, 4, 11, 15–17). For example, participants in the Summer Institutes program reported using scientific teaching approaches more frequently after completion of the program (11). In addition, former participants of the On the Cutting Edge program reported using more active learning exercises, and teaching observations revealed greater student engagement in classrooms, suggesting a shift to more learner-centered teaching practices (3).

The Faculty Institutes for Reforming Science Teaching (FIRST) IV program, another national-level PD effort, was implemented from 2009 to 2013 and aimed to guide biology postdoctoral scholars (postdocs) in implementing learner-centered courses and teaching practices. Details about the program structure are described in previous studies (10, 15). Over a 2-year period, two cohorts of postdocs (201 in total) engaged in an iterative process of curriculum development and mentored teaching practicums. The program used learning theory and evidenced-based active-learning strategies to guide and mentor teams of postdocs to shape their teaching practice around the principle of scientific teaching (18). Accordingly, FIRST IV postdocs learned instructional strategies to engage students in using scientific practices to learn the core concepts and content of biology during class meetings. Use of scientific practices, such as creating and testing models, building scientific arguments, and analyzing data, was intended to enable students to build and organize knowledge around core ideas in biology. Concurrently, the postdoctoral participants learned to develop assessments that integrated scientific practices with core ideas to measure not only what students know but also how they use their knowledge. Shortly after completion of FIRST IV, participants reported using more learner-centered teaching approaches and were directly observed using more learner-centered classroom practices compared to their peers (15).

A key design feature of FIRST IV was the focus on the development of postdocs as future biology faculty. If this PD program was to have a lasting impact on undergraduate teaching and learning, then FIRST IV participants would need to sustain their learner-centered focus over time and transfer the approaches and practices that they established as postdocs into their teaching as faculty. Transfer of training—as measured by participants' continued use of what they learned during PD—is complicated in the case of FIRST IV participants, who have moved on to different positions as faculty following their PD experience as postdocs. Thus, we use the concept of transfer of training to understand whether and how PD persists across both time and career transitions (19). Measuring long-lasting change, which is critical to the assessment of program impacts, requires a longitudinal approach. Longitudinal studies compare instructors

[1]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA. [2]Delta Program in Research, Teaching, and Learning, University of Wisconsin-Madison, Madison, WI 53706, USA.
*Corresponding author. Email: emeryna1@msu.edu

to themselves over time and are valuable to education research because they reveal maintenance of behavior and long-term impacts on teaching and track changes in instructors' behavior (20).

In this study, we define longitudinal as comparing data from instructors following PD, as postdocs to current data from instructors as faculty. We are not defining this as a set time period but rather as a shared set of events over time (postdoc PD experience followed by the career transition to faculty). Time between the completion of the FIRST IV program and the beginning of our study ranged from 3 to 8 years. Our data collection was timed to capture the whole range of early-career (pretenure) faculty years. Few studies have examined faculty instructors in the "novice" or early-career stage (4, 21) despite the importance of developing teaching practices during this career stage (22).

Most longitudinal studies of faculty teaching rely on qualitative data and self-reported surveys (23–25) that have inherent drawbacks (26, 27). In our study, we evaluated a combination of faculty self-reported approaches to teaching, observed teaching practices, student assessments, and student perceptions of the classroom environment among both FIRST IV faculty and a set of comparison faculty within the same academic departments (14). The questions we sought to answer were threefold: (i) How are PD outcomes maintained over time, as measured by teaching approaches and practices used by FIRST IV faculty? (ii) How do teaching approaches, practices, student assessments, and student perceptions compare between FIRST IV faculty and paired colleagues in their respective departments? (iii) How well do self-reported teaching approaches reflect observed practices in the classroom and student perceptions of the classroom environment, and do the relationships differ between groups of faculty? We hypothesized that FIRST IV faculty would continue to use the same degree of learner-centered teaching practices over time, to a greater extent than their peers. Specifically, we predicted that former FIRST IV participants would report a more learner-centered approach to teaching and demonstrate higher student engagement in the classroom and that these differences would be reflected in student assessments and student perceptions of the classroom environment.

RESULTS
Description of participants and courses
Initially, we verified that FIRST IV and comparison faculty were similar in relation to demographics, time allocation, and teaching experiences by administering a background survey. The majority of both groups of faculty identified as women (60% FIRST IV; 70% comparison). The percent of pretenure faculty in both groups was the same (86%) and both had approximately 6 years of teaching experience. The greatest amount of time invested in a single PD activity, including the FIRST IV program, was significantly higher for FIRST IV faculty than for comparison faculty, with more than twice the number of hours reported (Table 1; Student's t test, $P < 0.001$, Cohen's $d = 1.04$).

Course characteristics were also similar across the two groups of faculty (Table 2). Course enrollment ranged from 3 to 624 students with a mean and SD for FIRST IV faculty of 72 ± 106 and for comparison faculty of 61 ± 62. Most courses in our study were designed for undergraduate science majors, and the distribution of courses for majors versus non-majors was similar between the FIRST IV faculty and comparison faculty.

Table 1. Instructor identity, time allocation, and teaching experience at the beginning of the study. Asterisk (*) indicates statistically significant differences between faculty groups for $P < 0.05$.

Instructor descriptors	FIRST IV faculty (N = 40)	Comparison faculty (N = 40)
Identify as a woman/man	24/16	28/12
Tenured/tenure track/ nontenure track	5/32/3	5/31/4
Percent time allocation Teaching (means ± SD)	53.2 ± 19.6	53.1 ± 20.4
Percent time allocation Research (means ± SD)	34.9 ± 20.4	31.6 ± 19
Years as faculty before study (means ± SD)	3.4 ± 1.6	2.9 ± 3.1
Years of teaching experience before study (means ± SD)	6.5 ± 3.4	6.4 ± 6.1
Highest engagement PD activity in hours (means ± SD)	90.44 ± 58.8*	39.45 ± 37.4

Table 2. Course and student information for 2016–2019.

Course category	FIRST IV (N = 111)	Comparison (N = 105)
Large lecture (>60 students)	18	16
Large lecture with lab	11	13
Small lecture (<60 students)	28	25
Small lecture with lab	49	44
Small discussion course	1	4
Other course type	4	3
Course enrollment (mean number of students ± SD)	70.8 ± 105	61.6 ± 61.3
Type of students		
Majors	70	70
Non-majors	12	7
Both majors and non-majors	29	28

Teaching approach [Approaches to Teaching Inventory (ATI); (12)] and teaching practice [Reformed Teaching Observation Protocol (RTOP); (13)] did not change significantly over the 3 years of our study for either the FIRST IV faculty or the comparison faculty (2016–2019). For teaching approach, there are two independent subscale scores derived from survey responses: conceptual change/student-focused (CCSF) and information transfer/teacher-focused (ITTF). ATI subscale scores did not significantly change over 3 years for CCSF [repeated measures analysis of variance (ANOVA), $F = 1.667$,

$P = 0.198$] or ITTF (repeated measures ANOVA, $F = 0.759$, $P = 0.385$). Within-instructor variance was slightly higher (0.395) than between-instructor variance (0.087) for the CCSF subscale. The opposite was true for the ITTF subscale, as within-instructor variance was slightly lower (0.338) than between-instructor variance (0.392). RTOP scores did not significantly shift over the course of our study (repeated measures ANOVA, $F = 0.026$, $P = 0.873$). Within-instructor variance was lower (136.8) than between-instructor variance (525.7) for RTOP scores. There were no significant interactions for faculty group (FIRST IV or comparison) by year of study for either of the ATI subscales or RTOP. In addition, there was no significant relationship between RTOP scores and years as a faculty member (linear regression; $r^2 = 0.002$, $P = 0.68$) or years of teaching experience (linear regression; $r^2 = 0.001$, $P = 0.756$).

The ATI instrument consists of 22 items and is viewed as contextually dependent on the population surveyed (28). A confirmatory factor analysis ($N = 314$) found that the ATI model was a moderate fit for our data. A Kaiser-Meyer-Olkin test for sampling adequacy resulted in a "meritorious" designation (0.84). According to thresholds described by Harshman and Stains (28), our model fell within the Best Fit designation for "standardized root mean square residual" (SRMR; 0.078) and Better Fit designation for x^2/df (2.69) and "root mean square error of approximation" (RMSEA; 0.073). For the fit statistics "comparative fit index" (CFI; 0.809) and "Tucker-Lewis index" (TLI; 0.788), our ATI model fell within the Poorer Fit designation. Further discussion of the standard regression coefficients for items can be found in section S1.

Persistence of teaching approaches and practices

FIRST IV faculty reported similar teaching approaches (ATI) in courses at the end of the FIRST IV PD program (2011 or 2013) and courses in this study (2016–2019; Fig. 1A). No significant differences in FIRST IV faculty's support for the use of CCSF or ITTF teaching strategies in their courses were noted between past and present values of the ATI subscales: CCSF (paired Wilcoxon signed-rank test, $P = 0.51$, Cohen's $d = 0.097$) or ITTF (paired Wilcoxon signed-rank test, $P = 0.103$, Cohen's $d = -0.217$). This lack of change in ATI subscales is evidence of persistence in faculty teaching approach. For both past and present ATI scores, the CCSF subscale score was significantly higher than the ITTF score (past: paired Wilcoxon signed-rank test, $P < 0.001$, Cohen's $d = 1.102$; present: paired Wilcoxon signed-rank test, $P < 0.001$, Cohen's $d = 1.216$). There was no difference between instructor genders with respect to teaching approach over time. There were also no significant differences in ATI subscales over time whether a faculty participant was at a master's college and university or a doctoral university.

The mean RTOP score of videos for FIRST IV faculty over the 3 years of our study (48.03 ± 11.7) was within RTOP category 3, which is characterized by substantial student engagement with some minds-on and active involvement nearly half of the class period. RTOP scores were not significantly different between the present study and past results (paired Student's t test, $t = -0.99$, $P = 0.326$, Cohen's $d = 0.252$). Faculty demonstrated similar teaching practices in the present as they had in the past (Fig. 1B), evidence of persistence in observed teaching practices post-PD. There was no difference between instructor genders with respect to teaching practice over time. There were also no significant differences in RTOP scores over time for faculty at a master's college and university versus a doctoral university.
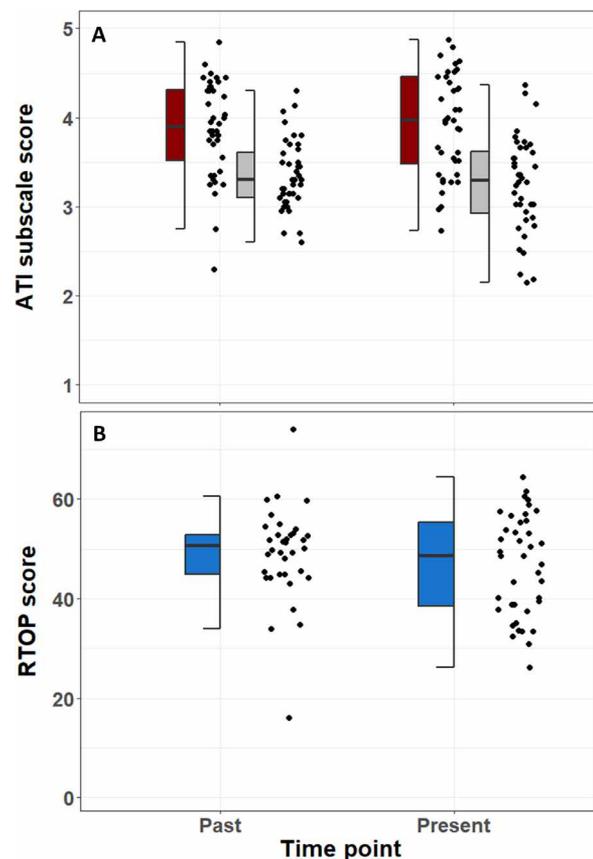
Fig. 1. Past and present teaching approach and teaching practice. (**A**) Past ATI scores were not significantly different from present scores for FIRST IV study participants (paired Wilcoxon sign-test; CCSF, $P = 0.51$, Cohen's $d = 0.097$; ITTF, $P = 0.103$, Cohen's $d = -0.217$). Each point represents the mean ATI score for CCSF (red) and ITTF (gray) subscales per participant. The boxes represent the interquartile range. (**B**) Past RTOP scores were not significantly different from present scores for FIRST IV study participants (paired Student's t test, $t = -0.99$, $P = 0.326$, Cohen's $d = 0.252$). Each point represents the mean RTOP score per participant. The boxes represent the interquartile range.

Despite teaching at different institutions with different student populations, there were no significant differences between past and present student perceptions of the teaching and learning environment, as measured by the Experience in Teaching and Learning Questionnaire [ETLQ; (29)] for FIRST IV faculty participants (paired Wilcoxon signed-rank tests). All differences in ETLQ subscales were non-significant, including deep approach to learning and studying ($P = 0.107$, Cohen's $d = -0.329$), surface approach to learning and studying ($P = 1$, Cohen's $d = -0.096$), alignment of teaching and learning ($P = 0.066$, Cohen's $d = 0.391$), choice about learning ($P = 0.75$, Cohen's $d = 0.056$), encouraging high-quality learning ($P = 0.058$, Cohen's $d = 0.398$), organization, structure, and content ($P = 0.063$, Cohen's $d = 0.356$), and support from other students ($P = 0.468$, Cohen's $d = 0.060$). The only instructor gender difference observed in ETLQ subscales was in encouraging high-quality learning. Students of women instructors reported significantly higher subscale scores in the present study than in the past ($P = 0.038$, Cohen's $d = 0.566$), whereas there was no difference in students of men instructors with regards to the encouraging high-quality learning subscale ($P = 0.97$, Cohen's $d = 0.003$). Our confirmatory factor

analysis of the seven ETLQ subscales used in this study ($N = 6{,}870$) found that our ETLQ model fell within the Best Fit designation for the fit statistics RMSEA (0.048) and SRMR (0.046). For the fit statistics CFI (0.938) and TLI (0.927), our ETLQ model fell within the Better Fit designation. Our ETLQ model fell into the Poorer Fit designation for x^2/df (16.68). A Kaiser-Meyer-Olkin test for sampling adequacy resulted in a "marvelous" designation (0.93). Further discussion of the standard regression coefficients for items can be found in section S2.

Faculty comparisons

FIRST IV faculty had a greater student-focused approach in their courses and a lesser teacher-focused approach than the comparison faculty (Table 3). FIRST IV faculty had significantly higher CCSF scores than comparison faculty (paired Wilcoxon signed-rank test, $P = 0.039$, Cohen's $d = 0.356$) and lower ITTF scores (paired Wilcoxon signed-rank test, $P = 0.016$, Cohen's $d = -0.455$). For comparison faculty, there was no significant difference between the CCSF and ITTF subscales (Wilcoxon signed-rank test, $P = 0.07$, Cohen's $d = 0.434$).

FIRST IV faculty had significantly higher RTOP scores than comparison faculty (Table 3 and fig. S1; paired Student's t test, $t = 5.35$, $P < 0.001$, Cohen's $d = 0.93$). Total RTOP scores for FIRST IV faculty ranged from 20.75 to 73.5 with an average and SD of 48.03 ± 11.67. This corresponds to RTOP level 3 that describes the classroom as substantial student engagement with some minds-on and hands-on involvement. Total RTOP scores for comparison faculty ranged from 19.25 to 64 with an average and SD of 38.62 ± 9.67 (Table 3). This corresponds to RTOP level 2 that describes the classroom as primarily lecture with some demonstration and minor student participation.

Faculty in the two groups assessed their students' learning differently. FIRST IV faculty assigned more points to assessment items that measured how students used knowledge. We used the three-dimensional learning assessment protocol [3D-LAP; (30)] to analyze assessment items because it is a tool that facilitates characterizing whether assessment items include three dimensions (core idea, scientific practice, and cross-cutting concept). We coded 242 total exams from 72 courses in year one of the study on two dimensions of the 3D-LAP (core ideas and scientific practices), and 97% of the items that assessed a scientific practice were also assessing a core idea.

The average weighted percent of items with scientific practices on exams used by FIRST IV faculty was $25.03 \pm 18.1\%$ (means $\pm$ SD) and those used by comparison faculty was $21.25 \pm 20.4\%$ (paired Student's t test, $t = 1.84$, $P = 0.078$, Cohen's $d = 0.197$). The difference between faculty groups was most evident in the constructed response items where the difference was marginally significant (paired Student's t test, $t = 2.04$, $P = 0.052$, Cohen's $d = 0.245$). There were no measured differences in assessments used by faculty at a master's college and university versus a doctoral university and no significant relationships with either Carnegie institution size or course size.

Student perceptions of the classroom did not differ significantly between the two groups of faculty for any of the ETLQ subscale scores. Despite no differences in the ETLQ, students did report a greater frequency of scientific teaching in the classroom for FIRST IV faculty than comparison faculty according to the Measurement Instrument for Scientific Teaching [MIST; (31)] composite score (Student's t test, $t = 2.82$, $P = 0.008$, Cohen's $d = 0.552$).

Relationship between reported teaching approaches and observed teaching practices
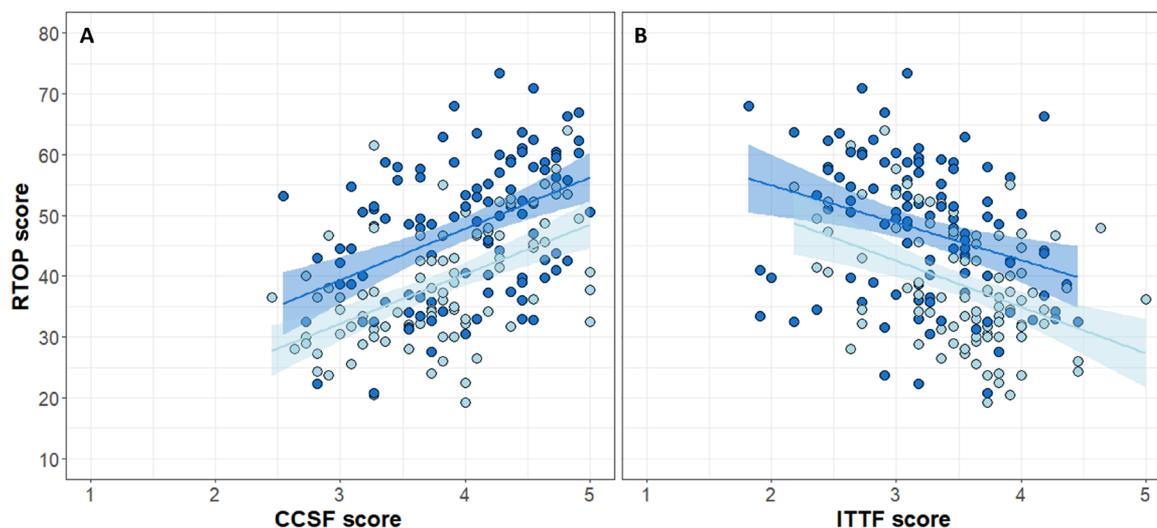
For both FIRST IV and comparison faculty, CCSF subscale scores were positively correlated with RTOP scores (FIRST IV: linear regression, $r^2 = 0.186$, $P < 0.001$; comparison: linear regression, $r^2 = 0.262$, $P < 0.001$; Fig. 2). In addition, ITTF scores were negatively correlated with RTOP scores for both FIRST IV (linear regression, $r^2 = 0.086$, $P = 0.0014$; Fig. 2) and comparison faculty (linear regression, $r^2 = 0.176$, $P < 0.001$). The relationships between teaching approach and teaching practice were the same for both women and men faculty in the study.

Observed teaching practices, student assessments, and student perceptions of the classroom

Student assessments appeared to be aligned with observations of learner-centered teaching in the classroom. There was a positive relationship between RTOP scores and the weighted percent of exam items that asked students to use scientific practices, but this relationship was only statistically significant for FIRST IV faculty (linear regression; FIRST IV, $r^2 = 0.26$, $P = 0.001$; comparison, $r^2 = 0.08$, $P = 0.068$). Across both groups of faculty, there were few relationships between observed teaching practice (measured by RTOP) and student perceptions of the classroom environment as measured by the ETLQ. A negative correlation existed between the RTOP score and the ETLQ subscale surface approach for comparison faculty (linear regression, $r^2 = 0.082$, $P = 0.0056$), and a positive correlation existed between the RTOP score and the ETLQ subscale choice for FIRST IV faculty (linear regression, $r^2 = 0.043$, $P = 0.022$). There was a significant positive relationship between student perceptions of scientific teaching practices as measured by the MIST and observed teaching practices measured by RTOP for FIRST IV (linear regression, $r^2 = 0.084$, $P = 0.012$; Fig. 3) and comparison faculty (linear regression, $r^2 = 0.228$, $P < 0.001$; Fig. 3). The relationships between teaching practice and student perceptions of the classroom were consistent across instructor genders. There was no significant instructor gender difference in student survey responses for any of the subscales except for "MIST: Course and Self-Reflection" that had higher scores for men instructors (Student's t test, $t = -2.12$, $P = 0.038$, Cohen's $d = -0.433$) and "ETLQ: Surface approach" that had higher scores for men instructors (Wilcoxon signed-rank test, $P = 0.012$, Cohen's $d = -0.426$).

Table 3. Means and SD of scores from different assessment types for FIRST IV faculty and comparison faculty. For the ATI, we used a paired Wilcoxon signed-rank test and for the RTOP and MIST, a paired Student's t test. Asterisk (*) indicates statistically significant differences between faculty groups for $P < 0.05$.

Assessment type	FIRST IV faculty	Comparison faculty
ATI: CCSF (means $\pm$ SD); range = 1 to 5	3.98 ± 0.6*	3.76 ± 0.64
ATI: ITTF (means $\pm$ SD); range = 1 to 5	3.21 ± 0.56*	3.51 ± 0.54
RTOP (means $\pm$ SD); range = 1 to 100	48.03 ± 11.67*	38.62 ± 9.67
MIST: Composite score (means $\pm$ SD); range = 1 to 100	58.4 ± 8*	53.2 ± 9.4

Fig. 2. Relationship between teaching approach (ATI subscales) and teaching practice (RTOP). (**A**) There was a significant correlation between the CCSF subscale and RTOP scores for each course assessed in the study (FIRST IV: linear regression, $r^2 = 0.186$, $P < 0.001$; comparison: linear regression, $r^2 = 0.262$, $P < 0.001$). (**B**) There was a significant correlation between the ITTF subscale and RTOP scores (FIRST IV: linear regression, $r^2 = 0.086$, $P = 0.0014$; comparison: linear regression, $r^2 = 0.176$, $P < 0.001$). Each point represents data from a single course for FIRST IV faculty (dark blue) and comparison faculty (light blue). The lines represent linear regressions with confidence intervals.
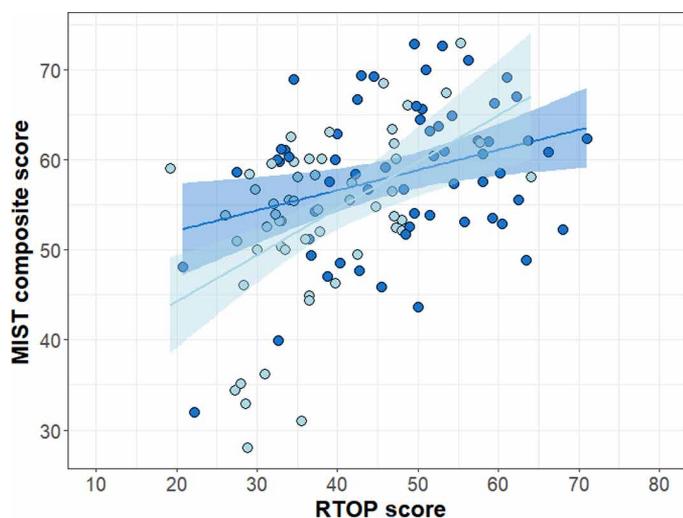
Fig. 3. Relationship between observed teaching practices (RTOP) and student perceptions of scientific practices (MIST composite) for both faculty groups. There was a significant correlation between teaching practice and student perceptions for FIRST IV (dark blue; linear regression, $r^2 = 0.084$, $P = 0.012$) and comparison faculty (light blue; linear regression, $r^2 = 0.228$, $P < 0.001$). Each data point represents a single course with a greater than 30% student response rate for the MIST survey instrument. The lines represent a linear regression of the relationship with confidence intervals.

DISCUSSION

A shared goal across all science teaching PD programs is to change the way that science is taught at the undergraduate level, which requires that participants continue to implement what they learn long after they have left the program. From 2009 to 2013, FIRST IV postdocs in the biological sciences participated in a rigorous teaching development program aimed at cultivating active-learning instruction that emphasized student engagement with scientific practices to learn concepts (10). Therefore, both instructional activities and assessment tasks evaluated students' ability to use knowledge with practices.

Soon after, many FIRST IV participants were found to have significantly greater learner-centered classrooms than their peers (15). Results from the present study extend this finding substantially and suggest that the outcomes of teaching PD can persist and transfer across a career transition and well into a faculty's career in higher education (Fig. 1, A and B). In addition, the former PD participants studied here demonstrated significantly greater learner-centered approaches and practices in the classroom than their peers (Table 3). Our study provides powerful evidence of the impact of the FIRST IV model on future faculty teaching development. Results from our study supported our hypothesis that FIRST IV faculty teaching did not change significantly between the PD program and the present day and that FIRST IV faculty have more learner-centered classrooms than matched comparison faculty.

It is possible that faculty teaching practices become more learner-centered with more experience; however, our results did not suggest that this was the case for either the FIRST IV or comparison faculty. Faculty with more teaching experience were not using more learner-centered approaches or implementing more student-centered practices in the classroom. Previous work found a negative correlation between years of teaching experience and learner-centered teaching practices (32), suggesting that while senior faculty may benefit from PD (33), it may be more valuable to focus PD efforts on early-career academics (32). Our results suggest that by targeting change in early-career instructors, higher education institutions can reap the benefits in the long-term. The most effective way to assess these long-term outcomes is with more longitudinal studies in higher education (1, 2, 20, 34). By tracking PD participants over time, we gain a greater understanding of program outcomes and effectiveness of the program itself.

Persistence of teaching development outcomes

Opportunities for repeated practice and reflection following PD may be critical for the long-term transfer of training (19). During FIRST IV, postdoctoral participants engaged in iterative cycles of teaching

practice, feedback, and reflection, and learner-centered instruction was an immediate outcome of the program (10, 15). In this study, we further sought evidence of long-term transfer of training from the context of a postdoctoral PD program to faculty teaching in undergraduate classrooms 6 or more years later. Over this time period, FIRST IV participants showed no significant change in their approach to teaching or use of teaching practices (Fig. 1, A and B). In addition, persistence was consistent across instructor genders and types of institutions.

Implementation of PD outcomes can change for many reasons, including lack of repetition (34), unsupportive communities of practice (6, 35), changes to incentives and pressures (36), and continued PD and growth as an instructor (1). Early-career faculty, in particular, face many new pressures that take time and resources (37), potentially prompting faculty to reallocate their focus away from learner-centered teaching practices. Therefore, we consider it noteworthy that FIRST IV faculty have maintained their learner-centered practice across the career transition from postdoc to faculty. Previous longitudinal studies have examined participant outcomes with surveys and interviews and found some evidence of persistent PD outcomes (21, 24, 25). Our results reinforce past work and confirm the long-term impacts of PD programs with validated survey instruments and observations of teaching.

Trained faculty are more student focused

A core objective of STEM teaching PD programs is to change instructor attitudes, approaches, and teaching practices. By comparing FIRST IV faculty to paired faculty from the same departments, we were able to detect differences in teaching between FIRST IV faculty and their peers and reduce the variability related to department. The differences observed between FIRST IV and comparison faculty provide strong evidence of an "effect" of the FIRST IV PD program. FIRST IV alumni approached their courses from a more learner-centered perspective and implemented more learner-centered practices in the classroom (Table 3 and table S3). This finding corroborates recent work demonstrating that PD has a positive impact on the learner-centered nature of the classroom (38). While our results might be expected for an effective PD program, they are notable given both the length of time since the FIRST IV program ended and the major career transition from postdoc to faculty at a different academic institution.

Teaching approach translates into teaching practice

The relationship between an instructor's self-reported approach and actual teaching practice may vary (32, 39, 40). The connection between attitude/beliefs and behavior is not limited to teaching and has been explored thoroughly, suggesting that an instructor's self-described teaching approach could influence their teaching practice (41). Our study results provided clear evidence of this relationship (Fig. 2). Both FIRST IV and comparison faculty showed a relationship between their self-reported teaching approach [as measured by the ATI; (12)] and their teaching practices in the classroom as assessed by independent observations [RTOP; (13)]. The higher the learner-centered approach, the higher the RTOP score. Conversely, the higher the teacher-focused approach, the lower the RTOP score. This suggests that self-reported teaching approaches were reflected in the teaching practices of both faculty groups, despite differences in teaching PD. This relationship informs future efforts to assess changes in teaching approaches and possibly the effectiveness of teaching PD.

Teaching practices, assessments, and student perceptions of the learning environment

Student assessment not only promotes learning but also is critical to understanding the effectiveness of teaching practices (42). Use of integrated assessments and related instructional materials that ask students to use scientific practices in the context of core ideas has the potential to elicit evidence of learning. In our study, we observed a positive relationship between learner-centered teaching practices (RTOP scores) and the proportion of exam items that included scientific practices (3D-LAP scores). FIRST IV faculty's assessments tended to emphasize students' use of practices more than the assessments of their peers. FIRST IV workshops focused on active-learning instruction that engaged students in using scientific practices. Therefore, both instructional activities and assessment tasks evaluated students' ability to use knowledge with practices. Our results suggest that PD may have not only affected classroom teaching practices but also promoted the use of assessments that elicit evidence of student learning through asking students to use the practices of science to engage with course concepts.

Students' perceptions of learner-centered teaching practices can vary (43, 44). The MIST reflected results from the ATI and RTOP, that FIRST IV faculty are teaching more learner-centered courses than comparison faculty (Table 3). Instructor observation scores (RTOP) were also positively correlated with MIST composite scores for both groups of faculty (Fig. 3). While FIRST IV faculty had higher RTOP and MIST scores overall, the positive relationship between instructor observations and student perceptions of scientific teaching was consistent across all faculty. These results suggest that differences in learner-centered teaching are ultimately perceived by students over the entire course. A recent study also found high agreement between instructor and student perceptions of scientific teaching (44). Our findings suggest that MIST may be a useful proxy for gauging the learner-centered nature of a course.

The ETLQ was designed to assess student perceptions of the teaching-learning environment and how learning occurs in the classroom (29). It is realistic to predict that these perceptions might change over time with different student populations and differ between groups of faculty. However, there were no significant changes for FIRST IV faculty in any of the ETLQ subscales despite the completely new teaching-learning environment from their previous institution during FIRST IV. This lack of change may be due to conserved faculty teaching approaches and practices (Fig. 1). However, there were no differences in ETLQ subscale scores between the two faculty groups. Thus, it is not clear how student perceptions of faculty teaching, as measured by the ETLQ, may or may not have changed over time. It is possible that the ETLQ subscales do not match well with the ATI or RTOP instruments, as previous studies have shown disconnects between faculty and student perceptions of the learning environment (45, 46). Overall, the ETLQ instrument did not inform our research questions on persistence or faculty differences in the degree to which the classroom was learner centered.

Limitations and future steps

While it is likely that the FIRST IV program significantly shifted participant beliefs, attitudes, and practices, it is still unknown how learner-centered the participants were before PD. Past data were collected during and toward the end of the FIRST IV program and thus reflect "post" PD outcomes. There is also potentially a self-selection bias from participants who chose to enter the FIRST IV

program. The postdocs may have been more receptive to evidence-based teaching and/or predisposed to learner-centered approaches to teaching before the start of FIRST IV. Thus, the finding of no change may be indicative of previously held teaching beliefs or teaching practices established before the FIRST IV program. Likewise, the differences observed between FIRST IV and comparison faculty may be due, in part, to previously held beliefs and practices.

Longitudinal studies of PD programs have great potential for revealing long-term outcomes for PD efforts. However, it is important to recognize that tracking the effects of a single PD program longitudinally is contingent upon observed change after the initial PD program. If there was little to no shift in teaching approach, attitudes, or practices after the program was completed, then any long-term observed change could not be definitively attributed to the outcomes of the PD program of study. Faculty teaching development can take many forms, and teasing apart the relative influence of different PD activities is difficult. Tracking participants in the long term who had not demonstrated change initially will likely not provide additional information about target PD program outcomes.

Conclusion

Persistent, long-term impacts are critical to the success of STEM teaching PD programs. Training of FIRST IV postdocs resulted in learner-centered teaching approaches and practices (*10*, *15*). We now show that FIRST IV participants have transferred this training to their current faculty instructional roles. Thus, the outcomes of FIRST IV have persisted across time and career transitions, providing evidence of success of the FIRST IV model for teacher PD. These findings support the claim that there are inherent benefits to training graduate students and postdocs before they enter academic careers (*21*, *47*).

In addition to providing evidence for PD outcomes, our study uncovered important relationships between instructor and student. Learner-centered teaching was positively correlated with students' use of scientific practices to do something with knowledge and concepts in assessments. In addition, student perceptions of scientific teaching in courses also aligned with the degree of learner-centered practices used by the instructor. These emergent relationships among instructors and students are invaluable for future research that examines change in instructor attitudes, approaches, and practices in STEM higher education.

MATERIALS AND METHODS

Experimental design

For three consecutive years, this study assessed teaching approaches, teaching practices, and student perceptions of the learning environment by administering survey instruments to faculty participants, video recording classroom teaching practices, and distributing survey instruments to students in participants' courses (fig. S2) (*14*). In addition, faculty participants completed a background survey at the beginning of the study. All surveys were distributed through the Qualtrics Survey system, and video recordings of classrooms were conducted by participants and uploaded to secure cloud servers. Our study was approved by the Social Science/Behavioral/Education Institutional Review Board at Michigan State University (IRB no. x16-627e).

Participants

We solicited former participants of the FIRST IV program who, before the start of the study, were all early-career faculty in biological

disciplines at academic institutions across the country. We contacted primarily tenure-track faculty at different types of institutions (doctoral, masters, baccalaureate, and associate levels). Participants in the longitudinal study came from 17 doctoral universities, 11 master's colleges and universities, 4 baccalaureate colleges, and 2 associate colleges according to the Carnegie classification system (*48*). Institution size ranged from 800 to 50,000 enrolled students. The institutions are located in 24 states from all regions of the United States including New England, Great Lakes, South, Midwest, Mountain West, and West Coast.

Each participating FIRST IV faculty was asked to seek out a matched colleague for the study that was preferably a tenure-track faculty at a similar career stage in the same department, who had not participated in the FIRST IV program. There is inherent value in comparing program participants to similar instructors who did not experience the same PD. The goal was to establish a comparison group of faculty who were similar to the FIRST IV faculty in their career stage, teaching experience, and differed only in their PD experience. We chose to have both FIRST IV faculty and their paired comparison faculty in the same department to reduce instructor variability associated with departmental influence on teaching approaches and practices. Our study design allowed us to ask questions about how the selected group (FIRST IV faculty) was different from the comparison group (*49*). There were 40 FIRST IV faculty and 40 paired comparison faculty. Each participant who collected and submitted data over the three consecutive years of the study was given an honorarium. Some study participants did not collect data for all 3 years (31 of 80 faculty) including several participants who were recruited mid-study (6 faculty). One FIRST IV participant switched institutions mid-study, and data collection continued at the new institution with a new comparison faculty participant. All participant data were included in the analyses.

Past data collection for FIRST IV participants

Data were also available from FIRST IV participants when they originally completed the PD program (6 to 10 years ago). Survey and classroom observation data were collected for the courses taught by postdocs while participating in the FIRST IV program. We aggregated past participant and student-reported data collected using the ATI (*12*), RTOP (*13*), and the ETLQ (*29*) from a previous study (*10*). Scores for each instrument were averaged across courses for each participant. The number of courses taught per participant during FIRST IV varied from one to three courses. These data are referred to as "past" data, whereas all data from the current study are considered "present" data. These two sets of data (past and present) allow us to make longitudinal comparisons of teaching approaches, practices, and student perceptions. These data are only available for FIRST IV participants.

Background and course information

We administered a background survey through Qualtrics to each faculty participant as soon as they consented to participate in the longitudinal study (section S3). This survey collected information pertaining to the academic/training background of each participant, their current academic position, and knowledge, experience, and confidence relating to pedagogical practices and previous teaching activities. During the 3 years of the study, faculty were asked to use one course taught per year for data collection, and when possible, faculty participants used the same course each academic year. We collected course information from faculty at the beginning of each

course. This included, for example, course size, student type (major/non-major), and course type (lecture/lab; Table 2).

Faculty teaching approach

For each course included in the study, faculty participants completed the ATI survey on Qualtrics at the beginning of their course. This instrument was used for FIRST IV participants in the past and assesses the extent to which an instructor uses teacher-focused and student-focused teaching approaches. It is a self-reporting instrument and is useful for assessing instructors' perceived approach for courses in different disciplines (28, 50). The ATI produces two subscale scores, CCSF and ITTF. Instructors who teach with a primarily CCSF approach focus on engaging students with deep approaches to build and modify their conceptual understanding of the course material. Instructors who teach with an ITTF approach transfer information to their students and focus on competency (12). The two subscales are independent of one another; thus, two scores were calculated per participant per course. To assess the reliability of the ATI (22-item, two-factor instrument) in our faculty population, we conducted a confirmatory factor analysis on historical FIRST IV faculty data and 3 years of our longitudinal study data ($N = 314$) using the lavaan R package (51). A Royston test determined that the data were not normally distributed, and thus, the maximum likelihood estimator was used for the confirmatory factor analysis.

Observations of teaching practices

For each course, we asked study participants to collect video recordings of two class sessions. We advised participants to allow at least 1 week between each video recording and to focus the recorder on the instructor as they taught the class and interacted with students. Audio was particularly important to capture on the recording. In less than 5% of courses, participants were only able to record one video per course. All recordings were deidentified before review. As in the previous study of FIRST IV participants (10), the RTOP (13) was used to rate teaching videos on the extent to which the course was student centered and transformed. RTOP is a teaching observation tool designed to assess teaching reform through use of learner-centered practices (13). The RTOP developers recommend using at least two observations per participant, although recent work suggests that at least four observations are necessary for a reliable characterization of teaching (52). For our study, there were two observations per course each year for each participant, resulting in a maximum of six observations per participant over the 3-year study.

To avoid potential bias due to visual recognition of instructors, we recruited independent video raters with expertise as doctoral-trained biologists from several academic institutions. Raters were trained in the use of the RTOP instrument until they calibrated with the trainer and the rater cohort and were periodically recalibrated. Over the duration of the study, 12 raters were involved in reviewing videos, and these raters were calibrated to one another with an intraclass correlation coefficient of 0.77 over six videos (53). We randomly assigned each deidentified teaching video to two raters, and the scores were averaged. Any video scores with an SD above seven were assigned to a third rater, and the subsequent two closest scores were used in the analyses.

Student assessments

In the first year of the study, we requested quiz/exam assessments of high value (any single quiz/exam worth greater than 20% of the to-

tal grade) from participants to characterize how they were assessing their students. Exam questions were scored using a simplified version of the 3D-LAP tool for characterizing the dimensions of scientific practices in the context of core ideas (30). We calculated the weighted percent (by point assignment) of a given exam that assessed a particular scientific practice and core idea for both constructed response and selected response items. For faculty that had multiple exams per course, metrics were averaged across all exams within the course.

Student perceptions of the classroom and scientific teaching practices

For each course, faculty participants administered the ETLQ to their students near the end of the term to assess students' perceptions of the teaching and learning environment. Students completed the questionnaire on Qualtrics and submitted their responses through a faculty participant-specific link. Students entered their student ID so that faculty could receive a list of survey completion but not their responses. Some faculty chose to incentivize students with course credit or extra credit for the survey, while others chose not to assign points to the survey. We only distributed and analyzed several of the subscales found in the ETLQ that were most aligned with the teaching approaches and practices measured in this study. They included the following: deep approach; surface approach; alignment; choice; encouraging high-quality learning; organization, structure, and content; and support from other students. Previous studies have also modified the ETLQ instrument in education research (54–56). The subscales included in our study are best aligned with the research questions and reflect student perceptions of a learner-centered classroom. For example, a deep approach to teaching is perceived by students as relating ideas and using evidence, while a surface approach consists of memorizing without understanding and compiling fragmented knowledge. The subscale alignment is how well the concepts and practices that students were taught matched the goals of the course. The subscale choice refers to students' perceptions about choices in their own learning. Encouraging high-quality learning is whether students felt that they were prompted in class to self-reflect and learn about how knowledge is developed. Organization, structure, and content is the perceived clarity of course objectives, organization, and flow of the classroom. Support from other students assesses how students supported each other and felt comfortable working with others (29). To assess the reliability of this instrument in our faculty population, we conducted a confirmatory factor analysis on 3 years of student survey data for the seven subscales included in this study ($N = 6870$).

In addition to student perceptions of the classroom environment, we assessed student perceptions on the frequency of scientific teaching practices they experienced in the classroom using the MIST (31). The MIST instrument was designed to gauge the frequency of scientific teaching practices in the classroom. It is possible that as faculty teach more learner-centered courses (higher RTOP scores), there is an increase in the frequency of scientific teaching practices, and students could perceive the difference. This instrument was published after the first year of data collection and thus was distributed in courses during years 2 and 3 of this study. Faculty participants distributed the MIST to students along with the ETLQ at the end of each course.

All student survey data were filtered to remove incomplete responses, duplicate responses, and poor responses from students (e.g., entering the same response for every item). The response rate

ranged from 7.6 to 100% with an average response rate of 67.6% across all 3 years of the study. We examined the variance in student responses to choose a percent response rate threshold. For five courses with large enrollments and high percent response rate (>90%), subscale scores were resampled at 10, 20, 30, 40, and 50% response rates. The subsequent SDs in student responses were plotted by resampling, and it was determined that a 30% response rate was the lowest threshold indistinguishable from 50%, an accepted threshold of response in the literature (57). Thus, ETLQ and MIST data were only analyzed from courses with >30% response rate. This resulted in the removal of 26 courses (of 212) from the present ETLQ and MIST data analysis.

Statistical analysis

All past (from FIRST IV) and present (all current courses) scores for ATI, RTOP, and ETLQ scores were averaged per participant to encompass their teaching approach, practices, and student perceptions. We tested all past and present variables for normality before statistical analysis. Only the ITTF subscale of the ATI had a non-normal distribution. For the ATI and ETLQ subscales, which are nominal, we ran a nonparametric paired Wilcoxon signed-rank test between the past mean scores and the present mean scores for FIRST IV faculty. For RTOP scores, we ran a paired Student's t test between the past scores and the present scores. To examine differences in persistence for faculty at master's colleges and universities ($N = 14$) or doctoral universities ($N = 20$), we conducted a separate paired Wilcoxon signed-ranked tests for ATI subscales and a paired Student's t test for the RTOP instrument.

To determine the difference in ATI and ETLQ subscales between the FIRST IV faculty and the comparison faculty, we ran paired Wilcoxon signed-rank tests using the within-department pairings of FIRST IV faculty and comparison faculty. We similarly used a paired Student's t test to compare RTOP and 3D-LAP scores between the two groups.

It is possible that ATI and RTOP scores have shifted over the 3 years of data collection (2016 to 2019). To determine whether scores changed significantly during the study, we took instrument subscale scores from each course and ran a repeated measures ANOVA with faculty group and year of study as explanatory variables. In addition, it is possible that time as a faculty instructor or years of teaching experience may influence their approach and teaching practice. We analyzed the relationship for self-reported "years as faculty" and "years of teaching experience" with ATI subscales and RTOP results. Linear regressions estimated how "years as faculty" or "years of teaching experience" were related to teaching approach and teaching practice.

Instructor approaches and teaching practices in the classroom may be related to one another. We calculated the mean score per participant (across all years of participation) and used a simple linear regression of both ATI subscales (CCSF and ITTF) and RTOP to determine whether and how approaches may be reflected in teacher practices in the classroom. In addition, we regressed RTOP scores with 3D-LAP exam scores, ETLQ, and MIST subscale scores to determine whether teacher practices were detectable in assessments and self-reported student perceptions of the classroom. These analyses were conducted separately for FIRST IV faculty and comparison faculty, as the relationships might differ depending on the faculty population.

All analyses were also examined for differences in responses according to self-reported instructor gender identity. For significance testing, we adhered to an α level of 0.05 and determined effect sizes for analyses based on prescriptions from past research (58). All analyses were performed using R version 3.6.1 (59).

REFERENCES AND NOTES

1. A. Stes, M. Min-Leliveld, D. Gijbels, P. Van Petegem, The impact of instructional development in higher education: The state-of-the-art of the research. *Educ. Res. Rev.* **5**, 25–49 (2009).
2. Y. Steinert, K. Mann, B. Anderson, B. M. Barnett, A. Centeno, L. Naismith, D. Prideaux, J. Spencer, E. Tullo, T. Viggiano, H. Ward, D. Dolmans, A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. *Med. Teach.* **38**, 769–786 (2016).
3. C. A. Manduca, E. R. Iverson, M. Luxenberg, R. H. Macdonald, D. A. McConnell, D. W. Mogk, B. J. Tewksbury, Improving undergraduate STEM education: The efficacy of discipline-based professional development. *Sci. Adv.* **3**, e1600193 (2017).
4. M. R. Connolly, Y. G. Lee, J. N. Savoy, The effects of doctoral teaching development on early-career STEM scholars' college teaching self-efficacy. *CBE Life Sci. Educ.* **17**, ar14 (2018).
5. R. C. Hilborn, The role of scientific societies in STEM faculty workshops: Meeting overview. College Park, MD: American Association of Physics Teachers (2013).
6. S. E. Brownell, K. D. Tanner, Barriers to faculty pedagogical change: Lack of training, time, incentives, and tensions with professional identity? *CBE Life Sci. Educ.* **11**, 339–346 (2012).
7. M. A. Pelch, D. A. McConnell, Challenging instructors to change: A mixed methods investigation on the effects of material development on the pedagogical beliefs of geoscience instructors. *Int. J. STEM Educ.* **3**, 5 (2016).
8. O. R. Aragón, S. L. Eddy, M. J. Graham, Faculty beliefs about intelligence are related to the adoption of active-learning practices. *CBE Life Sci. Educ.* **17**, ar47 (2018).
9. M. M. Cooper, M. D. Marcos, D. Ebert-May, C. L. Fata-Hartley, S. E. Jardeleza, J. S. Krajcik, J. T. Laverty, R. L. Matz, L. A. Posey, S. M. Underwood, Challenge faculty to transform STEM learning. *Science* **350**, 281–282 (2015).
10. D. Ebert-May, T. L. Derting, T. P. Henkel, J. M. Maher, J. L. Momsen, B. Arnold, H. A. Passmore, Breaking the cycle: Future faculty begin teaching with learner-centered strategies after professional development. *CBE Life Sci. Educ.* **14**, ar22 (2015).
11. C. Pfund, S. Miller, K. Brenner, P. Bruns, A. Chang, D. Ebert-May, A. P. Fagen, J. Gentile, S. Gossens, I. M. Khan, J. Labov, Summer institute to improve university science teaching. *Science* **324**, 470–471 (2009).
12. K. Trigwell, M. Prosser, Development and use of the approaches to teaching inventory. *Educ. Psychol. Rev.* **16**, 409–424 (2004).
13. D. Sawada, M. D. Piburn, E. Judson, J. Turley, K. Falconer, R. Benford, I. Bloom, Measuring reform practices in Science and Mathematics classrooms: The reformed teaching observation protocol. *School Sci. Math.* **102**, 245–253 (2002).
14. N. Emery, J. M. Maher, D. Ebert-May, Studying professional development as part of the complex ecosystem of STEM higher education. *Innov. High. Educ.* **44**, 469–479 (2019).
15. T. L. Derting, D. Ebert-May, T. P. Henkel, J. M. Maher, B. Arnold, H. A. Passmore, Assessing faculty professional development in STEM higher education: Sustainability of outcomes. *Sci. Adv.* **2**, e1501422 (2016).
16. K. N. Pelletreau, J. K. Knight, P. P. Lemons, J. S. McCourt, J. E. Merrill, R. H. Nehm, L. B. Prevost, M. Urban-Lurain, M. K. Smith, A faculty professional development model that improves student learning, encourages active-learning instructional practices, and works for faculty at multiple institutions. *CBE Life Sci. Educ.* **17**, es5 (2018).
17. M. T. Owens, G. Trujillo, S. B. Seidel, C. D. Harrison, K. M. Farrar, H. P. Benton, J. R. Blair, K. E. Boyer, J. L. Breckler, L. W. Burrus, D. T. Byrd, Collectively improving our teaching: Attempting biology department–wide professional development in scientific teaching. *CBE Life Sci. Educ.* **17**, ar2 (2018).
18. J. Handelsman, D. Ebert-May, R. Beichner, P. Bruns, A. Chang, R. DeHaan, J. Gentile, S. Lauffer, J. Stewart, S. M. Tilghman, W. B. Wood, Scientific teaching. *Science* **304**, 521–522 (2004).
19. S. L. Yelon, J. K. Ford, S. Golden, Transfer over time: Stories about transfer years after training. *PIQ* **25**, 43–66 (2013).
20. R. T. White, H. J. Arzi, Longitudinal studies: Designs, validity, practicality, and value. *Res. Sci. Educ.* **35**, 137–149 (2005).
21. A. Stes, M. Clement, P. Van Petegem, The effectiveness of a faculty training programme: Long-term and institutional impact. *Int. J. Acad. Dev.* **12**, 99–109 (2007).

22. A. E. Austin, Promoting evidence-based change in undergraduate science education, in *Fourth committee meeting on status, contributions, and future directions of discipline-based education research* (2011).

23. A. M. Knight, K. A. Cole, D. E. Kern, L. R. Barker, K. Kolodner, S. M. Wright, Long-term follow-up of a longitudinal faculty development program in teaching skills. *J. Gen. Intern. Med.* **20**, 721–725 (2005).

24. M. M. Tennill, M. W. Cohen, 9: Assessing the long-term impact of a professional development program. *To Improve the Academy* **32**, 145–159 (2013).

25. M. Stewart, Making sense of a teaching programme for university academics: Exploring the longer-term effects. *Teach. Teach. Educ.* **38**, 89–98 (2014).

26. G. M. Bodner, Changing how data are collected can change what we learn from discipline-based educational research. *PERC Proc.* **765**, 3–10 (2016).

27. S. R. Porter, M. E. Whitcomb, W. H. Weitzer, Multiple surveys of students and survey fatigue. *New Dir. Inst. Res.* **121**, 63–73 (2004).

28. J. Harshman, M. Stains, A review and evaluation of the internal structure and consistency of the Approaches to Teaching Inventory. *Int. J. Sci. Educ.* **39**, 918–936 (2017).

29. N. Entwistle, V. McCune, J. Hounsell, Approaches to study and perceptions of university teaching–learning environments: Concepts, measures and preliminary findings (Edinburgh, Scotland: Enhancing Teaching-Learning Environments in Undergraduate Courses Project, University of Edinburgh, Coventry University, and Durham University, 2002).

30. J. T. Laverty, S. M. Underwood, R. L. Matz, L. A. Posey, J. H. Carmel, M. D. Caballero, C. L. Fata-Hartley, D. Ebert-May, S. E. Jardeleza, M. M. Cooper, Characterizing college science assessments: The three-dimensional learning assessment protocol. *PLOS ONE* **11**, e0162333 (2016).

31. M. F. Durham, J. K. Knight, B. A. Couch, Measurement Instrument for Scientific Teaching (MIST): A tool to measure the frequencies of research-based teaching practices in undergraduate science courses. *CBE Life Sci. Educ.* **16**, ar67 (2017).

32. D. Ebert-May, T. L. Derting, J. Hodder, J. L. Momsen, T. M. Long, S. E. Jardeleza, What we say is not what we do: Effective evaluation of faculty professional development programs. *Bioscience* **61**, 550–558 (2011).

33. T. Huston, C. L. Weaver, Peer coaching: Professional development for experienced faculty. *Innov. High. Educ.* **33**, 5–20 (2008).

34. C. Henderson, A. Beach, N. Finkelstein, Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *J. Res. Sci. Teach.* **48**, 952–984 (2011).

35. J. A. Middleton, S. Krause, K. Beeley, E. Judson, J. Ernzen, R. Culbertson, Examining the relationship between faculty teaching practice and interconnectivity in a social network, paper presented at the 2015 ASEE/IEEE Frontiers in education conference, El Paso, TX, 2015.

36. A. E. Austin, Faculty cultures, faculty values. *New Dir. Inst. Res.* **68**, 61–74 (1990).

37. A. E. Austin, Preparing the next generation of faculty: Graduate school as socialization to the academic career. *J. High. Educ.* **73**, 94–122 (2002).

38. K. Viskupic, K. Ryker, R. Teasdale, C. Manduca, E. Iverson, D. Farthing, M. Z. Bruckner, R. McFadden, Classroom observations indicate the positive impacts of discipline-based professional development. *J. STEM Educ. Res.* **2**, 201–228 (2019).

39. L. Fung, L. P. Y. Chow, Congruence of student teachers' pedagogical images and actual classroom practices. *Educ. Res.* **44**, 313–321 (2010).

40. T. M. Addy, P. Simmons, G. E. Gardner, J. Albert, A new "class" of undergraduate professors: Examining teaching beliefs and practices of science faculty with education specialties. *J. College Sci. Teach.* **44**, 91–99 (2015).

41. I. Ajzen, The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* **50**, 179–211 (1991).

42. G. Joughin, Assessment, learning and judgement in higher education: A critical review, in *Assessment, Learning and Judgement in Higher Education* (Springer, 2009), pp. 1–15.

43. P. L. Machemer, P. Crawford, Student perceptions of active learning in a large cross-disciplinary classroom. *Active Learn. Higher Educ.* **8**, 9–30 (2016).

44. M. F. Durham, J. K. Knight, E. K. Bremers, J. D. DeFreece, A. R. Paine, B. A. Couch, Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). *Int. J. STEM Educ.* **5**, 31 (2018).

45. C. W. Beck, L. S. Blumer, Alternative realities: Faculty and student perceptions of instructional practices in laboratory courses. *CBE Life Sci. Educ.* **15**, ar52 (2016).

46. A. B. Heim, E. A. Holt, Comparing student, instructor, and expert perceptions of learner-centeredness in post-secondary biology classrooms. *PLOS ONE* **13**, 10.1371/journal.pone.0200524 (2018).

47. D. A. Greer, A. Cathcart, L. Neale, Helping doctoral students teach: Transitioning to early career academia through cognitive apprenticeship. *Higher Educ. Res. Dev.* **35**, 712–726 (2016).

48. Carnegie Foundation for the Advancement of Teaching, The Carnegie Classification of Institutions of Higher Education, 2010 edition, Menlo Park, CA: Author (2011).

49. J. R. Fraenkel, N. E. Wallen, H. H. Hyun, *How to Design and Evaluate Research in Education* (McGraw-Hill Humanities/Social Sciences/Languages, New York, 2011).

50. C. T. Williams, E. M. Walter, C. Henderson, A. L. Beach, Describing undergraduate STEM teaching practices: A comparison of instructor self-report instruments. *Int. J. STEM Educ.* **2**, 18 (2015).

51. Y. Rosseel, Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *J. Stat. Softw.* **48**, 1–36 (2012).

52. M. Stains, J. Harshman, M. K. Barker, S. V. Chasteen, R. Cole, S. E. DeChenne-Peters, M. K. Eagan, J. M. Esson, J. K. Knight, F. A. Laski, M. Levis-Fitzgerald, Anatomy of STEM teaching in North American universities. *Science* **359**, 1468–1470 (2018).

53. K. L. Gwet, *How to Compute Intraclass Correlation With MS EXCEL: A Practical Guide to Inter- Rater Reliability Assessment for Quantitative Data* (Advanced Analytics LLC, Gaithersburg, MD, 2010).

54. D. Mogashana, J. M. Case, D. Marshall, What do student learning inventories really measure? A critical analysis of students' responses to the approaches to learning and studying inventory. *Stud. Higher Educ.* **37**, 783–792 (2012).

55. A. Stes, L. Coertjens, P. Van Petegem, Instructional development in higher education: Impact on teachers' teaching behaviour as perceived by students. *Instruct. Sci.* **41**, 1103–1126 (2013).

56. J. Nigh, K. E. Pytash, R. E. Ferdig, W. Merchant, Investigating the potential of MOOCs in K-12 teaching and learning environments. *J. Online Learn. Res.* **1**, 85–106 (2015).

57. K. Fosnacht, S. Sarraf, E. Howe, L. K. Peck, How important are high response rates for college surveys? *Rev. Higher Educ.* **40**, 245–265 (2017).

58. J. M. Maher, J. C. Markey, D. Ebert-May, The other half of the story: effect size analysis in quantitative research. *CBE Life Sci. Educ.* **12**, 345–351 (2013).

59. R Foundation for Statistical Computing, 161 Vienna, Austria, www.R-project.org/.

60. L. MacNell, A. Driscoll, A. N. Hunt, What's in a name: Exposing gender bias in student ratings of teaching. *Innov. Higher Educ.* **40**, 291–303 (2015).

Citation: N. C. Emery, J. M. Maher, D. Ebert-May, Early-career faculty practice learner-centered teaching up to 9 years after postdoctoral professional development. *Sci. Adv.* **6**, eaba2091 (2020).

Science Advances

Early-career faculty practice learner-centered teaching up to 9 years after postdoctoral professional development

Nathan C. Emery, Jessica Middlemis Maher and Diane Ebert-May

ARTICLE TOOLS	http://advances.sciencemag.org/content/6/25/eaba2091
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2020/06/15/6.25.eaba2091.DC1
REFERENCES	This article cites 50 articles, 6 of which you can access for free http://advances.sciencemag.org/content/6/25/eaba2091#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service