

Article

Breaking the Cycle: Future Faculty Begin Teaching with Learner-Centered Strategies after Professional Development

Diane Ebert-May,* Terry L. Derting,[†] Timothy P. Henkel,[‡] Jessica Middlemis Maher,[§] Jennifer L. Momsen,^{||} Bryan Arnold,[¶] and Heather A. Passmore[†]

*Department of Plant Biology, Michigan State University, East Lansing, MI 48824; [†]Department of Biological Sciences, Murray State University, Murray, KY 42071; [‡]Department of Biology, Valdosta State University, Valdosta, GA 31698; [§]Delta Program, University of Wisconsin–Madison, Madison, WI 53706; ^{||}Department of Biological Sciences, North Dakota State University, Fargo, ND 58108; [¶]Department of Biology, Illinois College, Jacksonville, IL 62650

Submitted December 3, 2014; Revised March 11, 2015; Accepted March 23, 2015
Monitoring Editor: Deborah Allen

The availability of reliable evidence for teaching practices after professional development is limited across science, technology, engineering, and mathematics disciplines, making the identification of professional development “best practices” and effective models for change difficult. We aimed to determine the extent to which postdoctoral fellows (i.e., future biology faculty) believed in and implemented evidence-based pedagogies after completion of a 2-yr professional development program, Faculty Institutes for Reforming Science Teaching (FIRST IV). Postdocs (PDs) attended a 2-yr training program during which they completed self-report assessments of their beliefs about teaching and gains in pedagogical knowledge and experience, and they provided copies of class assessments and video recordings of their teaching. The PDs reported greater use of learner-centered compared with teacher-centered strategies. These data were consistent with the results of expert reviews of teaching videos. The majority of PDs (86%) received video ratings that documented active engagement of students and implementation of learner-centered classrooms. Despite practice of higher-level cognition in class sessions, the items used by the PDs on their assessments of learning focused on lower-level cognitive skills. We attributed the high success of the FIRST IV program to our focus on inexperienced teachers, an iterative process of teaching practice and reflection, and development of and teaching a full course.

INTRODUCTION

Despite the need to transform teaching and learning in the sciences (e.g., American Association for the Advancement

of Science, 2011; Anderson *et al.*, 2011; President’s Council of Advisors on Science and Technology, 2012; Association of American Universities, 2014) and the emerging body of research on how to do so (Amundsen and Wilson, 2012; Singer *et al.*, 2012), adoption of these findings by faculty who teach undergraduate science courses has been slow, at best (Brownell and Tanner, 2012; Smith and Valentine, 2012). The transformation of undergraduate science, technology, engineering, and mathematics (STEM) classroom experiences requires a fundamental shift in how instructors approach teaching and learning, moving from an information-transfer, teacher-centered model to one that is concept-focused, learner-centered, and collaborative (Weimer, 2002).

The way instructors approach their teaching is influenced profoundly by their beliefs and conceptions about teaching (Ho *et al.*, 2001; Lindblom-Ylänne *et al.*, 2006;

CBE Life Sci Educ June 1, 2015 14:ar22

DOI:10.1187/cbe.14-12-0222

Address correspondence to: Diane Ebert-May (ebertmay@msu.edu).

© 2015 D. Ebert-May *et al.* CBE—Life Sciences Education © 2015 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Non-commercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

Postareff *et al.*, 2007). Some authors claim that change in conceptions about teaching is a necessary prerequisite to changing instruction (Ho *et al.*, 2001), while others claim the opposite; that is, change in teaching practices occurs before change in beliefs (Guskey, 2000). In either case, conceptions about teaching take time to change (Postareff *et al.*, 2007) and require instructors to reflect on their own teaching practices and evaluate the relative consistency between their beliefs and actions in the classroom (Wlodarsky, 2005). Instructors' teaching conceptions are further influenced by the discipline and teaching context (Lindblom-Ylance *et al.*, 2006). All of these variables contribute to the complexity of teaching and learning in the higher education system. Institutions have acknowledged this reality and in response have advocated for decades the need to provide faculty opportunities to learn about effective teaching through professional development workshops (Connolly and Millar, 2006).

Even with the continued availability of and interest in teaching development opportunities, there is little evidence of resulting widespread impact on teaching practices and even less about the impact on student learning (Garet *et al.*, 2001; Gibbs and Coffey, 2004; Henderson *et al.*, 2011, 2012). For example, faculty participation in a 3-yr program of professional development that focused on transforming teaching did not result in implementation of learner-centered teaching by most participants. In surveys, the faculty reported incorporating many learner-centered activities in class, but observational data did not confirm those assertions (Ebert-May *et al.*, 2011). Other notable examples of programs that target professional development of current and future STEM faculty include the Center for the Integration of Research, Teaching, and Learning (Austin *et al.*, 2008; Pfund *et al.*, 2012), the Summer Institutes (SI) through the Center for Scientific Teaching at Yale (Handelsman *et al.*, 2004, 2006), On the Cutting Edge Workshops and Resources for Early Career Geoscience faculty (Manduca *et al.*, 2010), and the Workshop for New Physics and Astronomy Faculty (Henderson, 2008). Collectively, these programs impacted several thousands of faculty (Hilborn, 2012). Survey data from participants in these programs indicated that a large percentage of respondents made specific changes to their teaching practices, primarily shifting toward active-learning techniques. Only one of these programs reported data from direct observation, however, which indicated that 45% of the faculty who participated in workshops were transitioning toward learner-centered teaching and only 25% actually implemented learner-centered instruction (Teasdale *et al.*, 2011; Manduca *et al.*, 2014). Overall, the availability of reliable evidence for transformed teaching after professional development is limited across STEM disciplines, making the identification of professional development best practices and effective models for change difficult (Henderson *et al.*, 2011; Amundsen and Wilson, 2012).

Effective professional development requires that instructors reconceive the learning and teaching experience (Emerson and Mosteller, 2000; Henderson *et al.*, 2011), a process that can be productively viewed through the lens of conceptual change theory (Posner *et al.*, 1982; Pintrich *et al.*, 1993; Feldman, 2000). Workshop strategies that align with the theoretical framework of conceptual change are predicted to successfully help teachers transform their

conceptualization of the learning process and thus change their teaching practices. Researchers suggest that successful strategies for effecting change are collegial and community based, focus on content knowledge, and utilize concrete and coherent active-learning opportunities (Emerson and Mosteller, 2000; Garet *et al.*, 2001). Furthermore, conceptual change is more likely to occur when participants engage over an extended period of time (more than one semester; Shields *et al.*, 1998; Weiss *et al.*, 1998; Emerson and Mosteller, 2000; Henderson *et al.*, 2011). Finally, mentoring of and reflection by participants is a component of professional development that is critical for conceptual change (Hubball *et al.*, 2005; Brownell and Tanner, 2012). Collectively, these change strategies are predicted to work, because they address individual beliefs and experiences as well as situational factors that support or impede changes in teaching (Henderson and Dancy, 2007).

The aim of our research was to determine the extent to which postdoctoral fellows (i.e., future biology faculty) believed in and implemented evidence-based pedagogies after completion of a 2-yr professional development program, Faculty Institutes for Reforming Science Teaching IV (FIRST IV). If the program was effective at transforming teaching, then we predicted that 1) postdocs (PDs) would demonstrate belief in learner-centered approaches to teaching, 2) implement learner-centered teaching practices in the classroom, and 3) design assessments that are aligned with beliefs and practices of learner-centered teaching.

In developing FIRST IV, we selected implementation strategies based on scientific teaching (i.e., teaching science using evidence-based practices that include active learning and diversity; Handelsman *et al.*, 2004), research, findings from the conceptual change literature, and results from previous professional development programs. The FIRST IV program adapted theoretically based strategies in a professional development model built on a mentored, team-based approach to learning, in which participants engaged in an iterative process of curriculum development and teaching practicum, followed by reflection, revision, and a second teaching experience. Our approach was consistent with reviews of faculty professional development in which positive and/or lasting effect on teaching was associated with the use of repeated active and experiential interventions over time and collaboration (e.g., Gibbs and Coffey, 2004; Steinert *et al.*, 2006) and in which participants were actively engaged in all dimensions of learner-centered pedagogy (Henderson, 2008).

FIRST IV also built on the lessons learned from the earlier iteration of the project (referred to here as FIRST II; Ebert-May *et al.*, 2011) by 1) selecting participants who were less-experienced instructors (i.e., PDs), 2) focusing on actual teaching practices rather than focusing only on teaching tools to use in the classroom, and 3) helping participants build an entire course rather than develop a single unit of instruction. Furthermore, in response to the need for rigorous evaluation of the FIRST model, we incorporated methods that used direct observation and analysis of the participants' teaching and self-reported data. In doing so, we demonstrated that FIRST IV instructors implemented learner-centered teaching and did so to a greater degree than several comparison groups of faculty.

METHODS

The FIRST IV Program

PDs were the subjects of our research. The PDs were recruited nationally in 2009 and 2011, forming two separate, 2-yr cohorts of 99 and 102 PDs, respectively. The PDs accepted to FIRST IV were assigned to one of five regional teams located around the United States, each of which was based at a biological research field station and led by a team of two or three regional team leaders (RTLs) who were experts in biological science and pedagogy. Thirteen RTLs participated in a training workshop in the Spring of each of the first 2 yr of the project to prepare to implement the workshops (see Supplemental Material, section 1).

The PDs engaged in a 2-yr program of professional development. At the beginning of each year of participation, they completed a summer workshop. The broad objectives of the first workshop (4 d) for each cohort were for participants to 1) gain knowledge about evidence-based methods that support learner-centered STEM teaching; 2) begin to develop a learner-centered course, in which objectives, assessments, and instruction are aligned; and 3) make useful and sustainable connections with other PDs and RTLs for continuing project-related work after the workshop. The objectives for the second workshop (3 d) were for participants to 1) reflect on their teaching experience(s) from the prior year; 2) gain further practice with learner-centered, evidence-based teaching methods; 3) gain access to additional teaching and assessment tools and resources; 4) receive feedback about their teaching and job-seeking experiences; and 5) complete action plans for revision of their course and teaching in year 2.

The activities of the PDs in the academic year between workshops focused on three elements. First, each PD continued to develop a learner-centered introductory biology course with a team of PDs established during their first workshop. The second element was interaction of the PDs with their assigned RTL mentors and PD team as a means of receiving feedback about teaching, development of courses and teaching materials, and job applications. RTL mentors and their PDs established a meeting schedule and other interactions as needed. Third, the PDs completed an authentic teaching experience. Ideally, the experience was teaching one or more entire course(s); for many PDs, however, opportunities were only available to teach one unit or a few lectures of a course. In cohorts 1 and 2, 53% (74 of 140) and 67% (103 of 154) of the teaching experiences that occurred were full courses, respectively.

Assessing Teaching

To determine the effectiveness of FIRST IV for training learner-centered teachers, we used a mixed-methods approach (Creswell and Clark, 2007) that incorporated: 1) PDs' perceptions of their teaching, 2) rating of PD teaching based on independent observations of teaching videos, 3) rating of teaching videos obtained from non-FIRST faculty, and 4) the contents of assessments used by the PDs when teaching.

Participants' Perceptions about Teaching

We characterized the PDs' beliefs about their own teaching using the Approaches to Teaching Inventory 22 (ATI; Trigwell and Prosser, 2004; Trigwell *et al.*, 2005) and

surveys that we designed to document the PDs' knowledge and experience with active-learning pedagogy and teaching strategies (Supplemental Material, section 2). The ATI was developed to measure qualitative variation in two key dimensions of teaching; specifically, conceptual change/student focused (CCSF) and information transmission/teacher focused (ITTF). Results from the ATI are scored on a CCSF and an ITTF scale. Instructors who use a CCSF approach aim to change students' thinking about the material studied, with a focus on ways to challenge students' current ideas so that students construct their own knowledge. Instructors using an ITTF approach see their role as mainly to transmit information to students and to focus on development of skills that improve competency in information transfer. The two scales are independent rather than ends of a continuum (Prosser and Trigwell, 1997). Use of the ATI is context specific; thus, each PD completed the ATI at the end of each course that he or she taught. Project-created surveys about teaching knowledge and experience with education reform and active-learning approaches to teaching were completed by the PDs at the beginning and end of their participation in FIRST IV (see Supplemental Material, section 2).

Participants' Teaching Practice

Classroom teaching practices were assessed using an external review process. Each PD submitted videos for at least two complete class sessions for each full course that he or she taught during his or her FIRST IV participation. The specific class sessions that were recorded were determined by each PD. The PDs were asked to focus their videos on what they and their students did during the class. For example, we asked them to capture interactions with and among the students, to include any visual materials used (e.g., PowerPoint slides), and to accurately record audio of both their instruction and the students' conversations with each other.

Three considerations were key to our selection of an instrument for evaluating teaching practices of the PDs. First, the instrument needed to focus on the nature of student learning and in-class interactions rather than provide an ethogram of teaching behaviors. Classroom dynamics are inherent to the teaching approach used (e.g., teacher centered vs. learner centered) regardless of the topics studied during a given class period. Second, we considered the availability of comparative data from other professional development projects as a means of increasing the rigor of our project evaluation process (Hill *et al.*, 2013). Third, the instrument had to fit within the design of the FIRST IV project (e.g., time efficient given the large number of videos). Accordingly, we chose the Reformed Teaching Observation Protocol (RTOP). The RTOP is a validated observational instrument designed to measure the degree to which classroom instruction uses "reformed teaching" as defined by Sawada *et al.* (2002). The RTOP focuses on the nature of student learning and student-student and student-faculty interactions and is aligned with the theoretical underpinnings of constructivist literature about teaching and learning (Piburn *et al.*, 2000; MacIsaac and Falconer, 2002; Sawada *et al.*, 2002; Marshall *et al.*, 2011). It is a highly reliable instrument in terms of item reliability and interrater reliability across institutions and instructors (Marshall *et al.*, 2011; Amrein-Beardsley and Osborn Popp, 2012) with strong predictive validity for

Table 1. Scoring categories of the RTOP^a

RTOP category	Typical RTOP score	Type of teaching
I	0–30	Straight lecture
II	31–45	Lecture with some demonstration and minor student participation
III	46–60	Significant student engagement with some minds-on as well as hands-on involvement
IV	61–75	Active student participation in the critique as well as the carrying out of experiments
V	76+	Active student involvement in open-ended inquiry resulting in alternative hypotheses, several explanations, and critical reflection

^aAdapted from Sawada (2003).

student achievement (Falconer *et al.*, 2001; Lawson *et al.*, 2002; Bowling *et al.*, 2008). The RTOP has been used to assess the effectiveness of a variety of professional development programs (Adamson *et al.*, 2003; Addy and Blanchard, 2010; Ebert-May *et al.*, 2011). Also, comparative RTOP data were available from two previous faculty professional development programs.

Total score on the RTOP indicates the degree of learner-centered instruction and student involvement in a class session (Sawada *et al.*, 2002). The total score is obtained by summing subscores for each of five subcategories. The total score is classified into one of five categories in which categories I and II represent teacher-centered classrooms and categories III–V represent classrooms that are learner-centered to varying degrees (Sawada, 2003; Table 1). Details of the RTOP subcategories and score interpretations are explained fully in Budd *et al.* (2013).

We trained and calibrated biology education experts in the use of RTOP. During the initial calibration, all potential reviewers ($n = 18$) watched a set of eight to 14 videos, followed by discussion of their RTOP scoring. On completion of the initial calibration, reviewers who had an intraclass correlation coefficient (ICC; Gwet, 2010) of at least 0.7 and the time to commit to the video-review process, were selected as the final pool of reviewers ($n = 13$). Four of the experts were associated with a former FIRST project, two were project PDs, and the remainder were external to the FIRST IV program.

Review of videos was not initiated until year 3 of the project, so a pool that included videos from both PD cohorts was available. Each reviewer was assigned four to eight randomly selected videos to review each month. Each video ($n = 489$) was reviewed by two experts who did not know the PD, the cohort, or when the video was recorded. If the two reviewer scores for a video did not fall within the same RTOP category or lay at opposite ends of a single category (Table 1), then additional reviews were conducted by new reviewers until a majority of reviews yielded similar scores (< 8 points apart within one category). Outlier scores were not included in the final average RTOP score for that video. One randomly selected video was assigned to all reviewers each month for calibration purposes. The identity of the

calibration video was withheld from the reviewers to ensure that the review was conducted similarly to any other video. Interrater reliability of the reviewer group was measured by calculating the cumulative monthly ICC each month. If a reviewer had a video score that was an outlier for two calibration video recordings, then the reviewer was asked to reexamine the recordings and their ratings in light of scores provided by the other reviewers of that video. The average ICC for the total review period was 0.71 (range: 0.46–0.85). There was no significant change in ICC score over the 17-mo video-review period (linear regression, $r^2 = 0.017$, $p > 0.05$). Scores on the two videos submitted by a PD were averaged to obtain a final total RTOP score and subscores for each PD. There was no significant difference in total RTOP scores for the two videos submitted by the PDs (paired t -test, $p > 0.05$).

For comparison purposes, we also obtained teaching videos from 20 biology faculty who were not associated with a specific professional development program, each at a different institution, during 2011–2013. These faculty were recruited by FIRST IV PDs who were now in faculty positions and are hereafter referred to as “comparison faculty” (CF). When recruiting CF, the PDs sought faculty with teaching experience similar to their own; specifically, junior faculty who taught an introductory-level biology course. The teaching approach used by the CF was not a criterion for selection. Background information was obtained from all CF at the beginning of their semesters of participation. All but three of the faculty had less than 6 yr of teaching experience. Most of the CF (65%) reported no participation in faculty professional development programs in the prior 2 yr. Those who did engage in professional development reported activities such as attending education conferences and workshops, participating in a faculty learning community, working with a teaching mentor, and participating in a national training program. Each of the CF, the majority (78%) of whom taught an introductory-level biology course, submitted video recordings for at least two class sessions. The videos were reviewed as part of the pool of recordings described earlier.

Participants’ Assessments of Learning

An important component of the FIRST IV training was learning to design assessments that aligned with the types of learning students practiced during a course, for example, higher-order cognitive thinking and constructivist learning through cooperative work and active engagement. We evaluated the progress of PDs in their design of assessments by determining the level of cognitive skills targeted in their high-stakes assessments (i.e., exams and quizzes) when teaching an entire course. We used Bloom’s taxonomy (1956) to classify the cognitive skills assessed by each quiz/exam question used by PDs when teaching an entire course. Cognitive skill categories consist of six levels that represent a continuum from simple to complex cognitive tasks: 1) knowledge, 2) comprehension, 3) application, 4) analysis, 5) synthesis, and 6) evaluation. The first two categories can be considered to describe lower-order cognitive skills and the latter four categories to describe higher-order cognitive skills (Anderson and Krathwohl, 2001). Assessment items were assigned cognitive skill levels by two independent raters who had achieved a Cohen’s kappa of 0.87 ($n = 188$ assessments). We determined the percent of points on each quiz or exam

that was assigned in each Bloom's category (e.g., [25 points/80 points total] \times 100) and averaged the values within each Bloom's category for all assessments used in each course. If a PD ($n = 57$) taught the same course more than once, then the scores for all assessments for that course were averaged (seven of 57 PDs taught the same course two or more times).

Statistical Analyses

The surveys used to assess faculty approaches to their teaching were based on a five-point Likert scale. The data were treated as ordinal, and the statistical analyses were conducted using nonparametric tests (Roberson *et al.*, 1995). To characterize the PDs' approaches to teaching, we tested for differences between the subscales (CCSF and ITTF) using Wilcoxon signed-rank tests. To determine whether there was significant change in the PDs' approaches to teaching over the course of their training, we used mixed linear analyses, with instructor as a random effect, and paired t -tests of pre- and postscores on the ATI, with prescores obtained during the initial workshop as the PDs developed the courses they would teach the following year. Significant gains in the PDs' knowledge and firsthand experience with active-learning pedagogy and teaching strategies were determined by subtracting the prerating from the postrating and testing the resulting difference using Wilcoxon signed-rank tests.

Using a chi-square test with Yates correction, we compared the frequency of scores in the five RTOP categories (Table 1) for PDs who taught entire courses with those who taught part of a course. Differences in mean RTOP scores for PDs who taught an entire course and those who taught part of a course were analyzed using a t -test after testing the data for normality. We also tested for an effect of course level and enrollment using regression analysis. Course level was converted to a dummy variable with two categories, lower level (100–200 level) and upper level (≥ 300 level courses).

To analyze the distribution of points assigned to each Bloom's category on the tests and quizzes used by the PDs, we compared the distribution for lower-level courses (100–200 level) with that of higher-level courses (300–400) using the Kolmogorov-Smirnov two-sample test. The alignment between teaching practice, as measured by RTOP score, and the cognitive skills assessed by PDs, as determined by mean Bloom's score, was analyzed using Spearman's correlation coefficient. We used SAS version 9.3, release TS1M2 (SAS Institute, Cary, NC) for all statistical analyses. Data are presented as arithmetic means \pm 1 SE. Statistical significance was determined as $p \leq 0.05$. All protocols used in the FIRST project were approved by the Michigan State University Institutional Review Board (IRB X08-550 exempt, category 2).

RESULTS

Participants and Their Beliefs about Teaching

Participants in the two FIRST IV cohorts had similar demographics and teaching backgrounds, except that more of the cohort 2 PDs had experience as an adjunct/lecturer/instructor and more were from a non-research intensive institution compared with cohort 1 PDs (Table 2).

Results from self-reported data indicated that the PDs placed a greater emphasis on concept-focused, learner-

Table 2. Demographic and background characteristics of the two cohorts of FIRST IV postdoctoral scholars

Demographic/background variable	Cohort 1 ($n = 93$)	Cohort 2 ($n = 97$)
Gender ratio (F:M)	1.7:1	2.1:1
Home institution type (research: non-research) ^a	2.7:1	1.7:1
Prior TA experience	84%	83%
Prior instructor experience	37%	56%
Prior professional development activity (discussion group, workshop, longer-term program)	25%	29%

^aBased on Carnegie classifications.

centered teaching, compared with traditional information-transfer, teacher-centered instruction during the FIRST IV project. On average, the PDs reported significantly higher ratings on the CCSF scale (mean = 3.87 ± 0.04) of the ATI, compared with the ITTF scale (mean = 3.28 ± 0.04) when teaching a full course ($n = 190$ courses; Wilcoxon signed-rank test, $p < 0.0001$). On the ATI, participants reported the extent to which each survey item was true in their specific course on a five-point Likert scale ranging from "only rarely" to "almost always." A large majority of participants (91%) who taught a complete course had a mean score above 3 for the CCSF survey items, compared with 67% for the ITTF items. Also, 74% of the participants scored higher on the CCSF compared with the ITTF scale.

We tested for change in ATI score during participation in FIRST IV. There was no significant main effect of time on mean CCSF or ITTF score for participants who taught one or more complete courses ($n = 190$) from when they first entered the project to their last teaching experience (mixed linear analysis, $p > 0.05$). Because results of the ATI are specific to the course being taught, we also examined ATI scores for PDs ($n = 28$) who completed the ATI three times for the same course; that is, before teaching the course, at the end of their first time teaching the course, and at the end of their second experience teaching that same course. Before teaching the course, the PDs gave nearly identical ratings for their support and use of CCSF and ITTF approaches. At the end of the second teaching experience, support for the two approaches differed significantly (paired t -test, $p = 0.002$), with support for ITTF approaches decreasing over time by 0.28 ± 0.11 (paired t -test, $p = 0.019$; effect size = 0.49), and support for CCSF approaches increasing by 0.27 ± 0.15 , although not significantly (paired t -test, $p = 0.081$; effect size = 0.38).

Participants' Teaching Practice

The PDs also reported significant gains in their knowledge and firsthand experience with active-learning pedagogy and teaching strategies (Figure 1). Using data from PDs who completed both the pre- and postsurveys ($n = 130$), we found significant gains in firsthand experience in each of the five areas of pedagogy (Wilcoxon one-sample signed-rank test, $p < 0.0001$). Greatest gains in experience occurred in the areas of course/curriculum development and assessment. The gains for knowledge in the five areas of pedagogy paralleled the gains in firsthand experience. Significant gains in

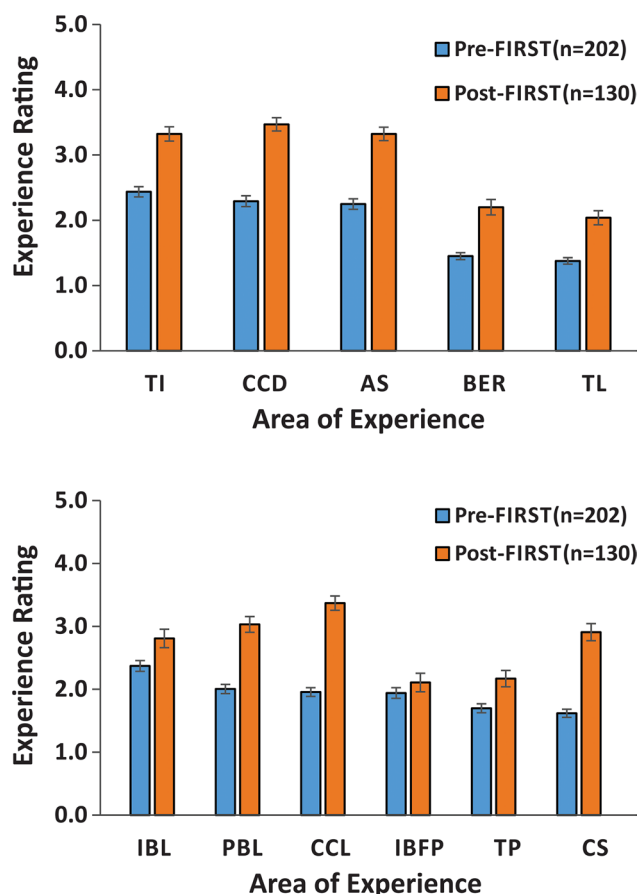


Figure 1. FIRST IV participants reported gains in firsthand experience with different dimensions of active-learning pedagogy (top: TI = technology instruction, CCD = course/curriculum development, AS = assessment, BER = biology education reform, TL = theories of learning) and strategies (bottom: IBL = inquiry-based laboratories, PBL = problem-based learning, CCL = cooperative/collaborative learning, IBFP = inquiry-based field projects, TP = teaching portfolios, CS = case studies). All responses were based on a five-point Likert-type scale with 5 as the highest rating and 1 the lowest rating. Error bars represent the SEs.

experience with five active-learning teaching strategies also occurred (Wilcoxon one-sample signed-rank test, $p < 0.01$; Figure 1). The greatest gains were in experience with cooperative/collaborative learning and case studies.

Expert reviews of the PDs' teaching videos provided independent evidence that the majority of PDs used transformed teaching practices. Almost all (86%) of the PDs who taught an entire course had total scores in RTOP categories III–V (mean = 54.1 ± 0.6 ; Figure 2), exhibiting significant student engagement and reformed teaching (MacIsaac and Falconer, 2002). In addition, the RTOP scores were aligned with the PDs' self-reported beliefs about their teaching in those same courses. The PDs' self-ratings on the ATI for the CCSF subscale were positively correlated with total RTOP score (Spearman $r = 0.43$, $p < 0.0001$; $n = 99$), while self-ratings for the ITTF subscale were negatively correlated (Spearman $r = -0.39$, $p < 0.0001$; $n = 99$). Teaching only part of a course had an adverse effect on RTOP scores. The PDs who taught a partial course had scores in a teacher-

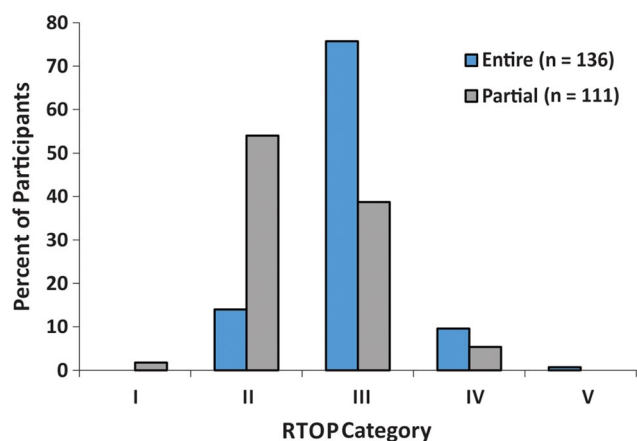


Figure 2. The frequency distribution of RTOP scores for PDs who taught an entire course and those who taught a partial course during their participation in FIRST IV.

centered RTOP category more frequently compared with those teaching an entire course ($df = 4$, $X^2 = 44.6$, $p < 0.0001$). The PDs who taught part of a course rather than an entire course also had significantly lower RTOP scores on average (mean = 43.7 ± 1.0 and 49.6 ± 0.7 , respectively; t -test, $p < 0.0001$; effect size = 0.31; Figure 2). Course level and enrollment had no significant effect on RTOP score ($F_{2, 146} = 1.51$, $p > 0.05$).

The RTOP scores of PDs differed from those of a sample of faculty who had not participated in FIRST IV (mixed analysis of variance, $p < 0.0001$). Less than 30% of the CF had scores in RTOP categories that indicated use of learner-centered teaching (Figure 3). The lower total RTOP scores of CF resulted from significantly lower scores for all subscales except propositional knowledge, compared with FIRST IV PDs (Table 3; mixed analyses of variance, $p < 0.0001$).

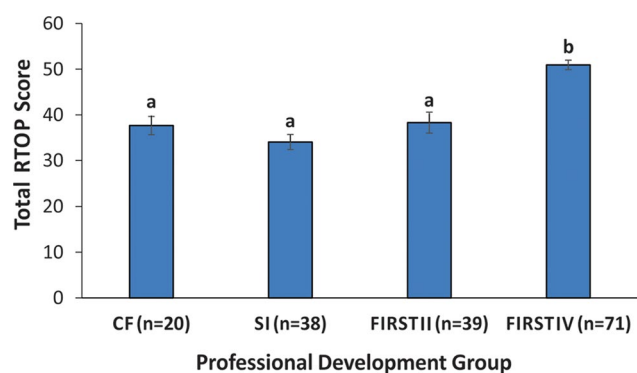


Figure 3. Mean total RTOP score (± 1 SE) for participants of three different professional development programs and a group of CF who did not participate in any of the three programs. Video recordings for SI and FIRST II were made after completion of professional development by the participants. Those for FIRST IV were from PDs who taught an entire course during the final year of professional development. Histograms bars with the same letter show means that are not statistically different from each other (mixed analysis of variance, $p < 0.0001$).

Table 3. Mean (± 1 SE) scores on the five subscales of the RTOP for teaching videos of participants in three faculty professional development programs and comparison faculty (CF)

Subscale	CF (<i>n</i> = 20)	SI faculty ^a (<i>n</i> = 37)	FIRST II faculty ^a (<i>n</i> = 37)	FIRST IV PDs (<i>n</i> = 145)
Lesson design	5.92 \pm 0.59	5.40 \pm 0.42	6.31 \pm 0.56	9.01 \pm 0.20
Propositional knowledge	13.94 \pm 0.36	13.11 \pm 0.25	12.59 \pm 0.46	13.87 \pm 0.12
Procedural knowledge	4.10 \pm 0.48	4.20 \pm 0.34	4.94 \pm 0.50	7.26 \pm 0.18
Communicative interactions	6.62 \pm 0.40	5.57 \pm 0.42	6.93 \pm 0.54	9.49 \pm 0.18
Student–teacher interaction	6.79 \pm 0.52	6.09 \pm 0.45	7.56 \pm 0.69	9.99 \pm 0.18

^aData were obtained as described in Ebert-May *et al.* (2011).

Participants' Assessments of Learning

In addition to analyzing teaching practices, we also determined the Bloom's level of assessments (*n* = 188) used by the PDs. The primary foci of the quiz and test questions used by the PDs who taught an entire course (*n* = 57) were knowledge and comprehension, which represented lower-order cognitive skills (Bloom, 1956). Twelve percent of the quiz and exam points, on average, were allocated to assessing higher-order thinking (Figure 4). There were no significant differences in the distribution of points assigned within any Bloom's level between upper- and lower-level courses (Kolmogorov-Smirnov two-sample test, $p > 0.05$). In addition, there was no significant relationship between the teaching practices of PDs based on RTOP score and the corresponding mean Bloom's score for the assessments used (Spearman correlation coefficient, $r = 0.22$, $p > 0.05$).

DISCUSSION

We documented the teaching practices of postdoctoral fellows who participated in the FIRST IV professional development program. The results supported our predictions about participant beliefs and teaching practices but not design of assessments. We use our results, from both direct observation and surveys, to identify best practices for the design and implementation of effective professional development programs, models of which are missing from the literature (Henderson *et al.*, 2011; Hill *et al.*, 2013; Wilson, 2013).

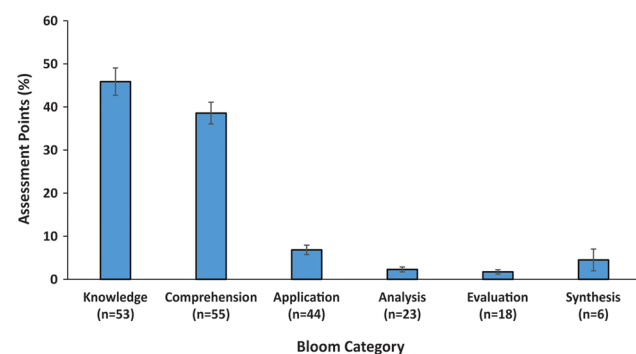


Figure 4. Mean percentage of assessment points per course (*n* = 57) categorized into each Bloom's category for PDs who taught an entire course.

Evidence of Effectiveness

We predicted that, if the FIRST IV program was effective, then the PDs would implement learner-centered teaching practices in the classroom and teach in ways that were different from peers who had not completed the FIRST IV program. Our data supported this prediction.

Comparison of actual teaching practices by the PDs with that of participants in other professional development programs is difficult, because approaches to program assessment vary and published results typically rely on self-reported data rather than independent evaluation by external experts (Hill *et al.*, 2013). We used the RTOP to evaluate teaching by the PDs, in part because we could compare the results with RTOP scores from the CF and faculty from two prior faculty professional development programs, SI (Pfund *et al.*, 2009) and FIRST II. We originally reported data from SI and FIRST II faculty in Ebert-May *et al.* (2011).

Baseline data about the CF and participants in the professional development programs and the courses they taught are presented in Table 4. The emphasis of FIRST II and SI on development of experienced faculty is reflected in their participants' significantly greater number of years of teaching experience compared with those in the FIRST IV and CF groups. The other notable difference among groups was the larger class size, on average, for SI faculty. The frequency of males and females and courses taught at the introductory level did not differ significantly among the groups (χ^2 , $p > 0.05$). There were no differences in the self-reported knowledge of active-learning pedagogies or firsthand experience with active-learning (mixed analysis of variance, $p > 0.05$). Smith *et al.* (2014) suggested that, when participants in a professional development program are asked to provide self-reported information about their teaching, they may feel pressured to provide positive responses as feedback to the program leaders. Such a response could contribute to the lack of agreement between self-reported data and data from external reviewers about the teaching practices of faculty (e.g., Ebert-May *et al.*, 2011). If pressure to meet program expectations was a significant factor, then we expected that the self-reported data from faculty in the professional development programs would be high and data from the CF would be significantly lower. Instead, we found that all faculty, whether in a professional development program or not, perceived themselves as having equally high levels of experience with active-learning teaching strategies (Table 4).

The procedure used for obtaining and reviewing recordings of class sessions for all groups was the same.

Table 4. Characteristics of the participants in three professional development programs (SI = Summer Institutes, FIRST = Faculty Institutes for Reforming Science Teaching) and a comparison faculty (CF) group^a

Group	<i>n</i>	Participants	Teaching experience ^b		Active-learning knowledge		Active-learning experience		Course level	Class size
		Female (%)	Mean (years)	SE	Mean	SE	Mean	SE	% Introductory	% Large (> 75 students)
SI	39	47	14.5 [†]	1.6	39.7	1.7	35.6	1.5	91	80 [†]
FIRST II	38	56	11.6 [†]	1.2	38.5	1.5	35.5	1.2	79	37*
FIRST IV	71	66	1.8*	0.14	42.1	1.2	40.2	1.2	67	17*
CF	20	47	3.9*	1.0	34.6	2.6	34.9	2.3	64	32*

^aKnowledge of and experience with active learning was determined from survey data. Introductory courses were 100- and 200-level courses.

^bValues in a column that share the same symbol (*, †) and values in columns without symbols are not significantly different from one another (generalized linear model, $p > 0.05$).

Specifically, participants chose at least two class sessions to record when teaching a complete course. The SI and FIRST II faculty provided recordings of courses taught after completion of their professional development program. The recordings were deidentified and reviewed by at least two experts who were trained and calibrated in the use of the RTOP and who did not know the instructor in the recordings. Any biases by participants in providing recordings, such as selecting class sessions that were closely aligned with pedagogy learned in a professional development program, would likely be similar among the groups. As for the videos from the FIRST IV PDs, there was no statistically significant difference in RTOP scores for the first compared with the second video submitted by participants in the SI or FIRST II programs or the CF (paired *t*-test, $p > 0.05$). Consistency in RTOP scores from class to class and independence of RTOP score from the topic taught in a class session were also reported by Budd *et al.* (2013) for most instructors. Our comparison here focuses solely on teaching practices after completion of a professional development program, because baseline data documenting the teaching practices of participants before professional development were not available for any of these programs.

The effectiveness of the FIRST IV program was striking when compared with other professional development programs and faculty with no FIRST IV training. The average RTOP score was significantly greater for FIRST IV participants compared with those of the FIRST II, SI, and CF groups (Figure 3; mixed analysis of variance, $p < 0.0001$). Three-fourths (74%; Figure 5) of the FIRST IV PDs who taught an entire course implemented learner-centered teaching (i.e., RTOP categories III–V) during their second year of professional development, based on RTOP scores assigned by expert raters. In contrast, less than one-third of the video-recorded faculty participants in the FIRST II or SI programs had a total RTOP score within a learner-centered RTOP category (i.e., categories III–V). The same was true of CF, of whom < 30% had RTOP scores that indicated use of learner-centered teaching (Figure 5). The scores of FIRST IV PDs on four of the five subscales of the RTOP were 37–77% greater than those of the FIRST II, SI, and CF groups (Table 3). Higher scores by the FIRST IV PDs on the four subscales resulted from their greater engagement of students in the learning process (e.g., making and testing predictions, creating and using models, and communicating their ideas to others; Budd *et al.*, 2013).

Situational differences among the groups could have marked influences on RTOP scores. Therefore, we tested for effects of the variables in Table 4 on total RTOP scores, excluding “knowledge of active-learning pedagogy” because of its high correlation with “experience with active learning.” The most influential variable on RTOP score by far was “group” (i.e., professional development program or lack thereof; general linear model, partial $\eta^2 = 0.14$), with SI and CF group having significantly lower RTOP scores on average compared with FIRST II and IV faculty (general linear model, $p = 0.002$). The only other variable of statistical significance was class size (general linear model, $p = 0.033$; partial $\eta^2 = 0.05$). Faculty teaching larger courses had lower RTOP scores in general compared with faculty teaching smaller courses. Thus, the relatively low RTOP scores of SI faculty could be attributable, in part, to the preponderance of large courses they taught (Table 4). When we controlled for variation in class size in the analysis of differences among the groups, however, the results were the same as without the covariate (Figure 3). Mean RTOP score for FIRST IV participants was still significantly greater than for the other three groups (analysis of covariance, $p < 0.0001$), and scores among the non-FIRST IV groups did not differ significantly from one another. It is worth noting that large class size does not prevent

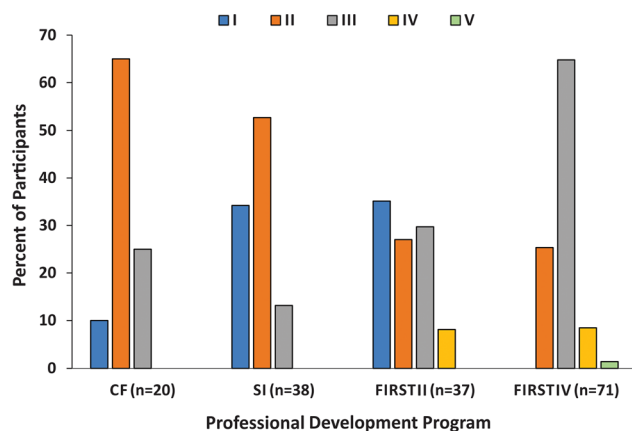


Figure 5. Frequency of RTOP scores among the five RTOP categories (Table 1) for participants of three different professional development programs and a group of CF who did not participate in any of the three programs.

implementation of learner-centered courses (Kober, 2015). As pointed out by Budd *et al.* (2013), some faculty in their study who taught large courses had high RTOP scores. The same was true in our comparisons in which seven faculty with courses of 100 to more than 200 students had total RTOP scores that were in a learner-centered category (score > 50). Gender of the professor, years and perception of teaching experience, and course level had no main effect on RTOP score.

The professional context of the participants' teaching in the four comparison groups might also have influenced their teaching practices. The PDs were employed in full-time research positions and also had the support of their principal investigators for teaching the course. The PDs, especially the 60% who taught an entire course, were in effect balancing research and teaching within their profession. However, since they were not in faculty positions for which teaching was a formal responsibility and part of annual evaluations, their interest and willingness to implement transformed teaching was perhaps less constrained by traditional departmental expectations. Faculty in the other groups, especially those who were untenured (SI = 45%, FIRST II = 66%, CF = 90% untenured), might have felt more limited in their teaching options. Data from the background survey completed by the CF provided some insight into possible impacts of tenure. Specifically, the CF were asked to rate "the extent to which tenure-related issues pose a challenge as you implement an active-learning course." On average, the responses were between "not challenging" and "somewhat challenging" (mean = 2.5 ± 0.3 , where 2 = not challenging and 3 = somewhat challenging). These results suggested that tenure, or the lack thereof, did not have a major influence on RTOP scores.

We also predicted that, if FIRST IV was effective, then the PDs would demonstrate belief in learner-centered approaches to teaching. The PDs believed that they increased their knowledge and experience with learner-centered pedagogy and teaching strategies (Figure 1), based on self-reported data. These gains were typical of self-reported data from faculty following professional development (e.g., Light *et al.*, 2009; Pfund *et al.*, 2009; Ebert-May *et al.*, 2011). Further evidence for the PDs' belief in learner-centered teaching was obtained from the ATI. When all courses were studied, the PDs reported greater use of CCSF compared with ITTF approaches to teaching in their classes. Similar outcomes from the ATI were reported for junior (Light *et al.*, 2009) and international (Gibbs and Coffey, 2004) faculty after professional development, but not for their untrained control groups.

We expected that the PDs' responses on the ATI would change over time, with an increased emphasis on CCSF approaches. When we examined only courses that were taught at least twice by the same PD, the PDs reported equal support for the CCSF and ITTF scales the first time they taught. The mean score for these PDs on the CCSF scale (3.60 ± 0.14 , $n = 28$) was similar to that reported for faculty teaching in other hard sciences, while the score for the ITTF scale (3.50 ± 0.12 , $n = 28$) was somewhat higher than for other faculty in the hard sciences (Lindblom-Ylance *et al.*, 2006). Our results indicated an initial level of dissonance in the PDs' beliefs about how they could help students learn. A high score on both scales may result from a belief that conceptual change in students' thinking could be accomplished through information transfer only, an approach that is not supported by science education

research (Prosser *et al.*, 2003). The subsequent decrease in PD scores on the ITTF scale indicated that, with experience and professional development, the PDs realized that information accumulation did not lead to conceptual change in students' understanding. Prosser *et al.* (2003) suggested that such shifts in teaching approach require understanding of the students' experience of learning (i.e., learner-centered approach) rather than the instructor's experience of teaching.

Our final prediction was that, if FIRST IV was effective, the PDs would design assessments aligned with beliefs and practices of learner-centered teaching. Thus, the types of learning practiced by students during the course on exams and quizzes would include an emphasis on assessing students' higher-order cognitive skills. During the FIRST IV workshops, the PDs learned to design assessments that revealed student proficiency with the types of thinking and skills students used in the classroom, using Bloom's taxonomy (Bloom, 1956) as a framework for levels of cognitive thinking. For example, the PDs practiced designing exam questions that incorporated conceptual model development and evaluation, argumentation, and data analysis (Crowe *et al.*, 2008). Yet the majority of the assessments used by the PDs during their teaching experiences focused on questions and tasks that required lower-order cognition, despite engaging students in a variety of activities in class that required higher-level thinking. The PD's use of lower-level assessment questions agreed with findings by Momsen *et al.* (2010). These questions emphasize primarily factual and concrete concepts, excluding the integration of concepts and skills that align with a learner-centered classroom experience. One explanation for the use of lower-level questions could be the practical difficulties associated with grading extended-response questions that focus on higher-level cognitive skills (e.g., synthesis, evaluation) in large courses. Courses taught by the PDs ranged in size from ten to 205 students (mean = 48), yet there was only a marginally significant negative correlation between class size and mean Bloom's score on assessments (Pearson's correlation, $r = 0.26$, $p = 0.05$). Momsen *et al.* (2010) found no relationship between course size and the cognitive level of assessments. The alignment between the training the PDs received and the assessments they used could potentially be improved by adding more peer and expert mentoring around assessment development, and analysis of assessment items during workshops.

Contributions to Future Professional Development Programs

What components of the FIRST IV program led to a large majority of participants who implemented learner-centered teaching practices? This question is difficult to answer without controlled studies in which specific program components are investigated. We offer here our best insights based on knowledge of previous professional development programs and their outcomes and feedback from the FIRST IV participants.

An obvious and unique component of FIRST IV was its focus on postdoctoral scholars rather than faculty with experience in teaching. Based on outcomes from the FIRST II project, less-experienced instructors were expected to more readily learn and adopt nontraditional teaching practices

(Gibbs and Coffey, 2004; Ebert-May *et al.*, 2011). In fact, that expectation was consistent with our results (Figures 3 and 5). The PDs chose to participate in FIRST IV, suggesting that they considered teaching as a key component of their professional identities and balance (Rybarczyk *et al.*, 2011; Brownell and Tanner, 2012).

The second key component was reflection by the PDs on their understanding of transformed teaching. Reflection was enabled during the iterative process of learning new pedagogical strategies during workshops, implementing the new strategies in courses, and reflecting on teaching with mentors and peers that occurred over 2 yr. Given the effectiveness of visual review of teaching practices (e.g., Baecher *et al.*, 2013; Osborne *et al.*, 2013), the reflection activities specifically included mentored review of video exemplars, videos of other PDs, and the participants' own teaching. The PDs' interactions with their mentors included examination of formative feedback from teaching videos, discussions about course design, students' feedback, and formal self-reflection before the second annual workshop. Critical reflection and dialogue with mentors may have enabled the PDs to develop congruence between their beliefs about teaching and subsequent classroom practices (Guskey, 2000; McAlpine and Weston, 2000; Sandretto *et al.*, 2002; Wlodarsky, 2005; Hatzipanagos and Lygo-Baker, 2006).

A third difference from most professional development events was our focus on the PDs developing an entire course rather than a smaller piece of instruction or individual teaching tools (e.g., case study, clickers). The basis for PDs' development of a course was to gain experience in creating course goals and objectives, designing assessments, and selecting course activities. By doing this, the PDs worked within a learner-centered course framework that enabled them to further develop, modify, and add lessons over time. In FIRST II and SI, the faculty developed pieces of instruction rather than a framework for an entire course and the RTOP scores were significantly lower than those of the FIRST IV PDs (Figure 3). We assert that this holistic approach to course design was more effective than developing a single lesson of learner-centered instruction to insert into a framework of a traditional course.

As schools, colleges, universities, and funding agencies continue to address the need for transformation of the STEM classroom experience through current and future faculty professional development, the consideration of professional development program design features is critical to improving the returns on time and funds invested (Hill *et al.*, 2013; Wilson, 2013). Results from FIRST IV are consistent with prior research on professional development (e.g., Osborne *et al.*, 2013), suggesting that duration (e.g., Desimone *et al.*, 2002) of professional development and practice is key to successful future programs. This means that participants in programs with education objectives must teach a full course, or at minimum the majority of a course, and teach more than once. Furthermore, the participants' teaching experiences must be paired with expert feedback, ideally from a mentor who is constructivist minded. Both the mentor and mentee need to make judicious use of teaching observations and review of course materials combined with an iterative process of reflection by the mentee. For all practical purposes, professional development is all about implementation and feedback, and we need to construct professional "learning"

opportunities in workshops that genuinely model what we do in the STEM classroom itself.

ACKNOWLEDGMENTS

The research was funded by the National Science Foundation under Division of Undergraduate Education Award 08172224 to D.E.-M. and T.L.D. We are appreciative of the postdoctoral scholars who participated in the FIRST IV program. We are deeply indebted to the regional team leaders (Stephanie Aamodt, Janet Batzli, Marguarite Brickman, Elizabeth Derryberry, Clarissa Dirks, Christopher Finelli, Janet Hodder, Jenny Knight, Debora Linton, Tammy Long, Marcy Osgood, Emily Rauschert, Courtney Richmond, Alison Roark, Christopher Tubbs, Kathy Williams, and Michelle Withers) for implementing the training workshops and mentoring the postdocs and to the experts who provided many hours of work reviewing video recordings. We extend our thanks to Sarah Jardeleza, Rachel Nye, Matt Berry, Dan Totzkay, Alec Aiello, and Gregory Moyerbrailean for their assistance in implementing the project and to Chris Mecklin for his statistical advice.

REFERENCES

- Adamson SL, Debra Banks, Burtch M, Cox F III, Judson E, Turley JB, Benford R, Lawson AE (2003). Reforming undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *J Res Sci Teach* 40, 939–957.
- Addy TM, Blanchard MR (2010). The problem with reform from the bottom up: instructional practices and teacher beliefs of graduate teaching assistants following a reform-minded university teacher certificate programme. *Int J Sci Educ* 32, 1045–1071.
- American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Final Report, Washington, DC.
- Amrein-Beardsley A, Osborn Popp SE (2012). Peer observations among faculty in a college of education: investigating the summative and formative uses of the Reformed Teaching Observation Protocol (RTOP). *Educ Assess Eval Account* 24, 5–24.
- Amundsen C, Wilson M (2012). Are we asking the right questions? A conceptual review of the educational development literature in higher education. *Rev Educ Res* 82, 90–126.
- Anderson LW, Krathwohl DR (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, complete ed., New York: Longman.
- Anderson WA, Banerjee U, Drennan CA, Elgin SCR, Epstein IR, Handelsman J, Hatfull GF, Losick R, O'Dowd DK, Olivera BM, *et al.* (2011). Changing the culture of science education at research universities. *Science* 331, 152–153.
- Association of American Universities (2014). Undergraduate STEM Education Initiative. <https://stemedhub.org/groups/aaui> (accessed 1 October 2014).
- Austin AE, Connolly M, Colbeck CL (2008). Strategies for preparing integrated faculty: the Center for the Integration of Research, Teaching, and Learning. In: *Educating Integrated Professionals: Theory and Practice on Preparation for the Professoriate*, New Directions for Teaching and Learning, ed. CL Colbeck, KA O'Meara, and AE Austin, San Francisco: Jossey-Bass, 69–81.
- Baecher L, Kung S-C, Jewkes AM, Ro C (2013). The role of video for self-evaluation in early field experiences. *Teach Teacher Educ* 36, 189–197.
- Bloom BS (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*, New York: McKay.
- Bowling BV, Huether CA, Wang L, Myers MF, Markle GC, Dean GE, Acra EE, Wray FP, Jacob GA (2008). Genetic literacy of undergraduate non-science majors and the impact of introductory biology and genetics courses. *BioScience* 58, 654–660.

- Brownell SE, Tanner KD (2012). Barriers to faculty pedagogical change: lack of training, time, incentives, and ... tensions with professional identity? *CBE Life Sci Educ* 11, 339–346.
- Budd DA, Kraft KJ, McConnell DA, Vislova T (2013). Characterizing teaching in introductory geology courses: measuring classroom practices. *J Geosci Educ* 61, 461–475.
- Connolly M, Millar S (2006). Using workshops to improve instruction in STEM courses. *Metropolitan Universities* 17, 453–65.
- Creswell JW, Clark VLP (2007). *Designing and Conducting Mixed Methods Research*, Thousand Oaks, CA: Sage.
- Crowe A, Dirks C, Wenderoth MP (2008). Biology in Bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7, 368–381.
- Desimone L, Porter A, Garet M, Yoon K, Birman B (2002). Effects of professional development on teachers' instruction: results from a three-year longitudinal study. *Educ Eval Policy Anal* 24, 81–112.
- Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE (2011). What we say is not what we do: effective evaluation of faculty professional development programs. *BioScience* 61, 550–558.
- Emerson JD, Mosteller F (2000). Development programs for college faculty: preparing for the twenty-first century. In: *Educational Media and Technology Yearbook*, vol. 25, ed. RM Branch and MA Fitzgerald, Englewood, CO: Libraries Unlimited, 26–42.
- Falconer K, Wyckoff S, Joshua M, Sawada D (2001). Effect of reformed courses in physics and physical science on student conceptual understanding. Paper presented at the Annual Conference of the American Educational Research Association, held 13 April 2001, in Seattle, WA.
- Feldman A (2000). Decision making in the practical domain: a model of practical conceptual change. *Sci Educ* 84, 606–623.
- Garet MS, Porter AC, Desimone L, Birman BF, Yoon KS (2001). What makes professional development effective? Results from a national sample of teachers. *Am Educ Res J* 38, 915–945.
- Gibbs G, Coffey M (2004). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learn Higher Educ* 5, 87–100.
- Guskey TR (2000). *Evaluating Professional Development*, Thousand Oaks, CA: Corwin.
- Gwet KL (2010). *How to Compute Intraclass Correlation with MS EXCEL: A Practical Guide to Inter-Rater Reliability Assessment for Quantitative Data*, Gaithersburg, MD: Advanced Analytics.
- Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, Gentile J, Lauffer S, Stewart J, Tilghman SM, *et al.* (2004). Scientific teaching. *Science* 304, 521–522.
- Handelsman J, Miller S, Pfund C (2006). *Scientific Teaching*, New York: Freeman.
- Hatzipanagos S, Lygo-Baker S (2006). Teaching observations: promoting development through critical reflection. *J Further Higher Educ* 30, 421–431.
- Henderson C (2008). Promoting instructional change in new faculty: an evaluation of the Physics and Astronomy New Faculty Workshop. *Am J Phys* 76, 179.
- Henderson C, Beach A, Finkelstein N (2011). Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. *J Res Sci Teach* 48, 952–984.
- Henderson C, Dancy M (2007). Barriers to the use of research-based instructional strategies: the influence of both individual and situational characteristics. *Phys Rev ST Phys Educ Res* 3, 020102.
- Henderson C, Dancy M, Niewiadomska-Bugaj M (2012). Use of research-based instructional strategies in introductory physics: where do faculty leave the innovation-decision process? *Phys Rev ST Phys Educ Res* 8, 020104.
- Hilborn RC (2012). *The Role of Scientific Societies in STEM Faculty Workshop, Meeting Overview*, Washington, DC: Council of Scientific Society Presidents/American Chemical Society.
- Hill HC, Beisiegel M, Jacob R (2013). Professional development research: consensus, crossroads, and challenges. *Educ Res* 42, 476–487.
- Ho A, Watkins D, Kelly M (2001). The conceptual change approach to improving teaching and learning: an evaluation of a Hong Kong staff development programme. *Higher Educ* 42, 143–169.
- Hubball H, Collins J, Pratt D (2005). Enhancing reflective teaching practices: implications for faculty development programs. *Can J High Educ* 35, 357–81.
- Kober L (2015). *Reaching Students: What Research Says about Effective Instruction in Undergraduate Science and Engineering*, Washington, DC: National Academies Press.
- Lawson AE, Benford R, Bloom J, Carlson MP, Falconer KF, Hestenes DO, Judson E, Piburn MD, Sawada D, Turley J, *et al.* (2002). Evaluating college science and mathematics instruction: a reform effort that improves teaching skills. *J Coll Sci Teach* 31, 388–393.
- Light G, Calkins S, Luna M, Drane D (2009). Assessing the impact of a year-long faculty development program on faculty approaches to teaching. *Int J Teach Learn Higher Educ* 20, 168–181.
- Lindblom-Ylänne S, Trigwell K, Nevgi A, Ashwin P (2006). How approaches to teaching are affected by discipline and teaching context. *Stud Higher Educ* 31, 285–298.
- MacIsaac D, Falconer K (2002). Reforming physics instruction via RTOP. *Phys Teach* 40, 16–22.
- Manduca CA, Iverson E, McConnell DA, Bruckner M, Greenseid L, Macdonald RH, Tewksbury B, Mogk DW (2014). On the cutting edge: combining workshops and on-line resources to improve geoscience teaching. Paper presented at the Geological Society of America Annual Meeting, held 19–22 October 2014 in Vancouver, BC.
- Manduca CA, Mogk DW, Tewksbury B, Macdonald RH, Fox SP, Iverson ER, Kirk K, McDaris J, Ormand C, Bruckner M (2010). SPORE: Science Prize for Online Resources in Education: On the Cutting Edge: teaching help for geoscience faculty. *Science* 327, 1095–1096.
- Marshall JC, Smart J, Lotter C, Sirbu C (2011). Comparative analysis of two inquiry observational protocols: striving to better understand the quality of teacher-facilitated inquiry-based instruction. *School Sci Math* 111, 306–315.
- McAlpine L, Weston CB (2000). Reflection: issues related to improving professors' teaching and students' learning. *Instr Sci* 28, 363–385.
- Momsen JL, Long TM, Wyse SA, Ebert-May D (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE Life Sci Educ* 9, 435–440.
- Osborne J, Simon S, Christodoulou A, Howell-Richardson C, Richardson K (2013). Learning to argue: a study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *J Res Sci Teach* 50, 315–347.
- Pfund C, Manske B, Austin AE, Connolly M, Moore K, Mathieu R (2012). Advancing STEM undergraduate learning: preparing the nation's future faculty. *Change* 44, 64–72.
- Pfund C, Miller S, Brenner K, Bruns P, Chang A, Ebert-May D, Fagen AP, Gentile J, Gossens S, Khan IM, *et al.* (2009). Summer Institute to improve university science teaching. *Science* 324, 470–471.
- Piburn MD, Sawada D, Falconer K, Turley J, Benford R, Boom I (2000). *Reformed Teaching Observation Protocol (RTOP)*. http://physicshed.buffalostate.edu/AZTEC/RTOP/RTOP_full (accessed 11 March 2015).
- Pintrich PR, Marx RW, Boyle RA (1993). Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Rev Educ Res* 63, 167–199.

- Posner GJ, Strike KA, Hewson PW, Gertzog WA (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Sci Educ* 66, 211–227.
- Postareff L, Lindblom-Ylänne S, Nevgi A (2007). The effect of pedagogical training on teaching in higher education. *Teach Teacher Educ* 23, 557–571.
- 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. www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf (accessed 15 October 2014).
- Prosser M, Ramsden P, Trigwell K, Martin E (2003). Dissonance in experience of teaching and its relation to the quality of student learning. *Stud Higher Educ* 28, 37–48.
- Prosser M, Trigwell K (1997). Relations between perceptions of the teaching environment and approaches to teaching. *Br J Educ Psychol* 67, 25–35.
- Roberson PK, Shema SJ, Mundfrom DJ, Holmes TM (1995). Analysis of paired Likert data: how to evaluate change and preference questions. *Family Med* 27, 671–675.
- Rybarczyk B, Lerea L, Lund PK, Whittington D, Dykstra L (2011). Postdoctoral training aligned with the academic professoriate. *BioScience* 61, 699–705.
- Sandretto S, Kane R, Heath C (2002). Making the tacit explicit: a teaching intervention programme for early career academics. *Int J Acad Dev* 7, 135.
- Sawada D (2003). Reformed Teacher Education in Science and Mathematics: An Evaluation of the Arizona Collaborative for Excellence in the Preparation of Teachers, Arizona State University Document Production Services, Tempe.
- Sawada D, Piburn MD, Judson E, Turley J, Falconer K, Benford R, Bloom I (2002). Measuring reform practices in science and mathematics classrooms: the Reformed Teaching Observation Protocol. *School Sci Math* 102, 245–253.
- Shields PM, Marsh JA, Adelman NE (1998). Evaluation of NSF's Statewide Systemic Initiatives (SSI) Program: The SSIs' Impacts on Classroom Practice, Menlo Park, CA: SRI International.
- Singer SR, Nielsen NR, Schweingruber HA (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, Washington, DC: National Academies Press, 264.
- Smith DJ, Valentine T (2012). The use and perceived effectiveness of instructional practices in two-year technical colleges. *J Excell Coll Teach* 23, 133–161.
- Smith MK, Vinson EL, Smith JA, Lewin JD, Stetzer MR (2014). A campus-wide study of STEM courses: new perspectives on teaching practices and perceptions. *CBE Life Sci Educ* 13, 624–635.
- Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach* 28, 497–526.
- Teasdale R, Budd D, Cervato C, Iverson E, Kraft KJVDH, Manduca C, McConnell DA, McDaris JR, Murray DP, Slattery W (2011). Enhancing student-centered teaching practices: approaches developed on the new Cutting Edge Geosciences RTOP website. Paper presented at the Geological Society of America Annual Meeting, held 9–12 October 2011, in Minneapolis, MN.
- Trigwell K, Prosser M (2004). Development and use of the approaches to teaching inventory. *Educ Psychol Rev* 16, 409–424.
- Trigwell K, Prosser M, Ginns P (2005). Phenomenographic pedagogy and a revised approaches to teaching inventory. *High Educ Res Dev* 24, 349–360.
- Weimer M (2002). *Learner-Centered Teaching: Five Key Changes to Practice*, San Francisco, CA: Jossey-Bass.
- Weiss I, Montgomery D, Ridgway C, Bond S (1998). *Local Systemic Change through Teacher Enhancement: Year Three Cross-Site Report*, Chapel Hill, NC: Horizon Research.
- Wilson SM (2013). Professional development for science teachers. *Science* 340, 310–313.
- Wlodarsky R (2005). The professoriate: transforming teaching practices through critical reflection and dialogue. *Teach Learn* 19, 156–172.