What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs

DIANE EBERT-MAY, TERRY L. DERTING, JANET HODDER, JENNIFER L. MOMSEN, TAMMY M. LONG, AND SARAH E. JARDELEZA

Professional development (PD) workshops designed to help faculty move from teacher- to learner-centered science courses for undergraduates are typically evaluated with self-reported surveys that address faculty's satisfaction with a workshop, what they learned, and what they applied in the classroom. Professional development outcomes are seldom evaluated through analysis of observed teaching practices. We analyzed videotapes of biology faculty teaching following PD to address three questions: (1) How learner centered was their teaching? (2) Did self-reported data about faculty teaching differ from the data from independent observers? (3) What variables predict teaching practices by faculty? Following PD, 89% of the respondents stated that they made changes in their courses that included active, learner-centered instruction. In contrast, observational data showed that participation in PD did not result in learner-centered teaching. The majority of faculty (75%) used lecture-based, teacher-centered pedagogy, showing a clear disconnect between faculty's perceptions of their teaching and their actual practices.

Keywords: professional development, undergraduate biology, RTOP, learner-centered teaching, student-centered classrooms

National reports continue to advocate for reform of undergraduate biology education in order to better prepare students for professions in the biological sciences (Alberts 2008, Brewer and Smith 2011). One of the major issues identified is that science is not taught as it is practiced—that is, using active, inquiry-based approaches (AAAS 1990, Handelsman et al. 2007). Indeed, researchers have confirmed that active, learner-centered instructional approaches are associated with enhanced learning in science (Prince 2004, Knight and Wood 2005, Michael 2006, Wood 2009, Blanchard et al. 2010, Derting and Ebert-May 2010).

The importance of student engagement as a means of enhancing student learning is the underpinning of many educational workshops implemented annually in the United States that offer faculty opportunities to learn to improve undergraduate science instruction (DeHaan 2005, Holdren 2009, Pfund et al. 2009). These workshops are typically brief and intense, ranging from several hours to several days, and competencies are addressed that experts deem appropriate for effective instruction, such as active and collaborative learning, team teaching, responding to diversity, and assessing student learning (Garet et al. 2001, Connolly and Millar 2006). Faculty motivation for participation in such workshops is often prompted by dissatisfaction with their teaching, course goals, instructional practice, or student learning outcomes (Sunal et al. 2001, Clarke and Hollingsworth 2002, Gess-Newsome et al. 2003). Therefore, the overall goal in the majority of teaching reform workshops is to change and sustain faculty beliefs and instructional practices in order to improve opportunities for student learning (Connolly and Millar 2006).

To date, the efficacy of workshops has been evaluated principally through self-reported surveys of faculty satisfaction, learning, and examples of applications in their classrooms (Connolly and Millar 2006). In a recent study about undergraduate education in biology (Pfund et al. 2009), the participating faculty reported that they created more effective, learner-centered environments using proven, effective instructional strategies after participating in workshops. Notably, the report identified the need for an independent assessment of classroom practice to supplement self-reported survey data. In fact, there is still an absence of objective measures for evaluating teaching practices of faculty following professional development (PD). In response, we used videotapes (direct observation) along with survey data to study teaching practices of faculty after their completion of a PD program. Specifically, we used the Reformed Teaching Observation Protocol (RTOP) following PD as a means of quantifying the degree to which an undergraduate science classroom is learner centered (Sawada et al. 2002).

BioScience 61: 550–558. ISSN 0006-3568, electronic ISSN 1525-3244. © 2011 by American Institute of Biological Sciences. All rights reserved. Request permission to photocopy or reproduce article content at the University of California Press's Rights and Permissions Web site at *www.ucpressjournals.com/reprintinfo.asp.* doi:10.1525/bio.2011.61.7.9

Our research was designed as a post hoc analysis of two PD programs to address three questions: (1) How learner centered was the pedagogy that faculty implemented following PD? (2) Did the self-reported data about the faculty's teaching practices differ from the data compiled by independent observers? (3) What variables predict teaching practices of faculty following PD?

The study and participants

We focused on faculty participants in two national PD programs: Faculty Institutes for Reforming Science Teaching (FIRST II) and the National Academies Summer Institute on Undergraduate Education in Biology (SI) at the University of Wisconsin. The overarching goals for faculty in both workshops were to increase knowledge about the principles of active learning and scientific teaching (Handelsman et al. 2007) and the corresponding instructional practices; to collaboratively develop active, inquiry-based instructional materials; and to gain experience and confidence in implementing active-learning pedagogies during the workshop and in their classrooms after the workshop. The expectation of both programs was that by providing baseline knowledge and opportunities to practice, faculty would reflect this knowledge and experience in their subsequent classroom instruction.

The FIRST II program extended over three years. Teams of faculty from the same department or institution met three to six different times, for a total of 6–12 workshop days during the program. The SI program involved faculty teams from a department or institution for a five-day workshop during the summer. The SI program occurred annually, with new faculty teams each year (Pfund et al. 2009). During both workshops, the participant faculty designed instructional units that included learning objectives; assessment of student learning that was aligned with those objectives; and active, learner-centered teaching strategies, such as cooperative learning and interactive classroom activities (Lundmark 2002, Pfund et al. 2009). Both workshops focused on introductory biology courses, which often serve large enrollments and diverse student populations.

All faculty from the two PD programs were invited to participate in our study, and we included faculty who responded to survey data (henceforth called *PD participants*; $n = 211$) and a subset of PD participants who also volunteered to videotape their classes (henceforth called *RTOP study participants*; $n = 77$). The RTOP study participants were evenly split by PD program, gender, and tenure status (table 1). Teaching experience ranged from 2 to 41 years and averaged 13 years. The majority (64%) of study participants came from doctoral/research universities, as categorized by the Carnegie Classification system (Carnegie Foundation for the Advancement of Teaching 2001). The remaining 36% of the participants were divided among associate's colleges (10%), baccalaureate colleges (12%) and master's colleges and universities (14%). The doctoral/research universities were specifically targeted for recruitment into the SI program; therefore, they represent the largest proportion of the sample.

Two types of data were collected from the PD participants: surveys (self-reports) and videotapes (observations) of their teaching. FIRST II participants completed more surveys than SI participants because of program differences. The videotapes were collected when the RTOP study participants taught their courses after the PD program. Not all PD participants submitted all of the surveys.

Four surveys focused on the participants' knowledge and implementation of pedagogy, the PD experience itself, and variables that potentially affected teaching practices (table 2). The surveys included a *baseline survey*, in which faculty were asked to report their knowledge of and experience with active-learning pedagogy; a *midprogram survey*, in which faculty were asked to list the reforms that they implemented in their courses based on PD (including assessments, instructional strategies, and teaching materials); a *course survey*, in which faculty were asked for the number of students enrolled in their courses and to indicate the instructional strategies that they used and the frequency

Table 1. RTOP study participant demographics, the mean and standard error for the number of years of teaching experience with undergraduates, and the average knowledge and experience with active-learning teaching strategies from the final survey.

		Participants		Teaching experience		Active-learning knowledge		Active-learning experience	
Group	N	Female (percentage)	Tenured (percentage)	Mean (years)	SE	Mean	SE	Mean	SE
RTOP	77	53	51	13.4	1.0	39.9	1.1	36.2	1.0
FIRST II	39	59	56	12.3	1.3	40.3	1.5	37.1	1.2
SI	38	47	45	14.5	1.6	39.7	1.7	35.6	1.5

Note: Active-learning knowledge was calculated as the sum of five-point Likert scale responses for knowledge of specific active learning practices; active-learning experience was the sum of five-point Likert scale responses for experience with specific active learning practices.
FIRST II, Faculty Institutes for Reforming Science Teaching professional development program; RTOP, Reformed Teaching Observation Protocol study participants; SE, standard error; SI, National Academies Summer Institute on Undergraduate Education in Biology professional development program at the University of Wisconsin.

Table 2. Survey instruments, sample sizes, and participants.

Survey instrument	Sample size	Completed by FIRST II participants	Completed by SI participants
Baseline survey	143	Pre-PD	n/a
Midprogram survey	112	During PD	n/a
Course survey	96	Each time course was taught post-PD	Each time course was taught post-PD
Final survey	113	After completion of PD	After completion of PD

FIRST II, Faculty Institutes for Reforming Science Teaching professional development program; n/a, not applicable; PD, professional development; SI, National Academies Summer Institute on Undergraduate Education in Biology professional development program at the University of Wisconsin

of their usage. The course survey was administered each time the course was taught. For instructors who submitted more than one course survey, only the survey with the highest frequency of active-learning strategies was used for the analysis. We chose to report the best-case scenario for implementation by each faculty following PD. In the *final survey*, faculty responded to a variety of questions, including those concerning their knowledge of and experience with active-learning pedagogy, their confidence about their preparation as a teacher, the level of departmental support that they received, and the challenges that they encountered. The final survey was administered at the end of PD.

Observations of teaching by the RTOP study participants were accomplished using videotapes of class sessions after the participants' completion of PD. The faculty submitted one or more videotapes for analysis of their classroom teaching practice ($n = 192$). The faculty were asked to focus the video recordings on what the faculty and students did during the class, including all interactions among students, those between teacher and students, and any visual materials used. In many cases, there was not a cameraperson making decisions about where to direct the taping; rather, the camera was fixed in place, focused on the instructor. Direct videotaping of students' faces was avoided in order to protect anonymity. Taping began within two semesters after completion of the PD program. We requested videotapes from the faculty four times in two years—twice during each iteration of a course. In fact, not all RTOP study participants submitted a full set of videotapes; 65 faculty submitted two or more videotapes. On average, the first videotaping of a class occurred 40 days after the participants' completion of PD, and the last taping of a class occurred during the second year after PD. Within a course, successive tapings occurred two to six weeks apart. A minimum of one semester elapsed between the first and last taping (average interval was 258 days). We did not obtain videotapes of teaching prior to PD.

Use of the RTOP to rate videos

The videotapes were rated using the RTOP (Sawada et al. 2002), which allows a trained observer to characterize the degree to which faculty implement active-learning teaching techniques in their courses. The RTOP defines and allows the assessment of learner-centered teaching and is aligned with the theoretical underpinnings of inquiry-based teaching (MacIsaac and Falconer 2002). The RTOP is a highly reliable instrument with strong predictive validity for student achievement (Lawson et al. 2002).

An RTOP score is an indicator of the degree of active-learning instruction and student involvement observed in a classroom (Sawada et al. 2002). A final score is obtained by summing five subscales, which quantitatively rate an instructor on (1) *lesson design and implementation*, how well the instructor elicits students' prior knowledge, to what extent the instructor engages students and allows them to influence the focus and direction of the lesson; (2) *propositional knowledge*, how a teacher implements discipline-specific content; (3) *procedural knowledge*, the inquiry processes students use in the lesson; (4) *communicative interactions*, the proportion and variety of discussion that occurs among students and between the teacher and students; and (5) *student–teacher relationship*, attitudes of the teacher and students toward each other and the quality of their interactions. RTOP scores range from 0 to 100, with higher scores representing more learner-centered classrooms (i.e., students actively participate, take primary responsibility for their own learning, interact with each other, and shape the direction of the discussion), whereas lower scores indicate lecture-based, teacher-centered classrooms (i.e., lecture is the primary mode of communication, with minimal roles for students beyond note taking or the use of personal response systems).

Total RTOP scores are classified into one of five categories: categories I and II represent teacher-centered classrooms, and categories III–V represent classrooms that are learner-centered to varying extents (Sawada 2003; table 3). A reformed, learner-centered introductory undergraduate science course typically falls into category III. Categories IV and V are difficult to achieve, in part because the physical classroom—a theatre-style lecture hall with large numbers of students—constrains activities and interaction (Sawada et al. 2002). Laboratory settings or state-of-the-art SCALE-UP classrooms (Gaffney et al. 2008) remove many of the barriers to learner-centered instruction, resulting in students engaged in asking and exploring their own scientific questions (Handelsman et al. 2007). In such instances, reformed teaching can fall into categories IV and V.

The RTOP instrument (Sawada et al. 2002) was used to rate each of the 192 videotapes, with each tape rated twice by different individuals. The raters were randomly chosen from a pool of eight science faculty trained in using the RTOP protocol. All raters were trained and their ratings calibrated during an in-person training session, using an intraclass correlation that estimates the internal consistency among raters' scores (Shrout and Fleiss 1979). After all eight reviewers achieved a significant intraclass correlation of .70 among RTOP scores with a sample of 10 videotapes, they began reviewing the tapes of the study participants. The videotapes were identified only by code; therefore, demographic data about the instructor, information about the course, and the date of taping were not available to the raters. Each tape was rated in a blind review, and the analysis of the tapes occurred over 24 months. Twice during the 24-month period, all eight raters were recalibrated by viewing and rating two selected videos that were not included in the study. This procedure maintained the intraclass reliability at or above its initial level.

Table 3. Categorization of RTOP scores.

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	76–100	Active student involvement in open-ended inquiry, resulting in alternative hypotheses, several explanations, and critical reflection.

Source: Adapted from Sawada 2003.
RTOP, Reformed Teaching Observation Protocol

Statistical analyses

Self-reported data were compiled from the surveys (table 2). Because there were no significant differences on the baseline survey (pre-PD) in the responses of the FIRST II participants who did and did not participate in the RTOP study (one-tailed Wilcoxon two-sample test, $p > .05$), all responses were retained in the baseline data set. Likewise, the final survey data (post-PD) were pooled when responses from the FIRST II and SI faculty did not differ significantly. The responses on specific items of the baseline survey were compared with those on the final survey using one-tailed Wilcoxon two-sample tests (figure 1).

Ten independent variables that characterized faculty responses to PD and faculty demographics were evaluated as potential predictors of faculty teaching practices (box 1). These variables included teaching experience before and after PD, teaching environment (e.g., class size, departmental support), and influences on the extent to which faculty implemented active-learning strategies (e.g., confidence, challenges

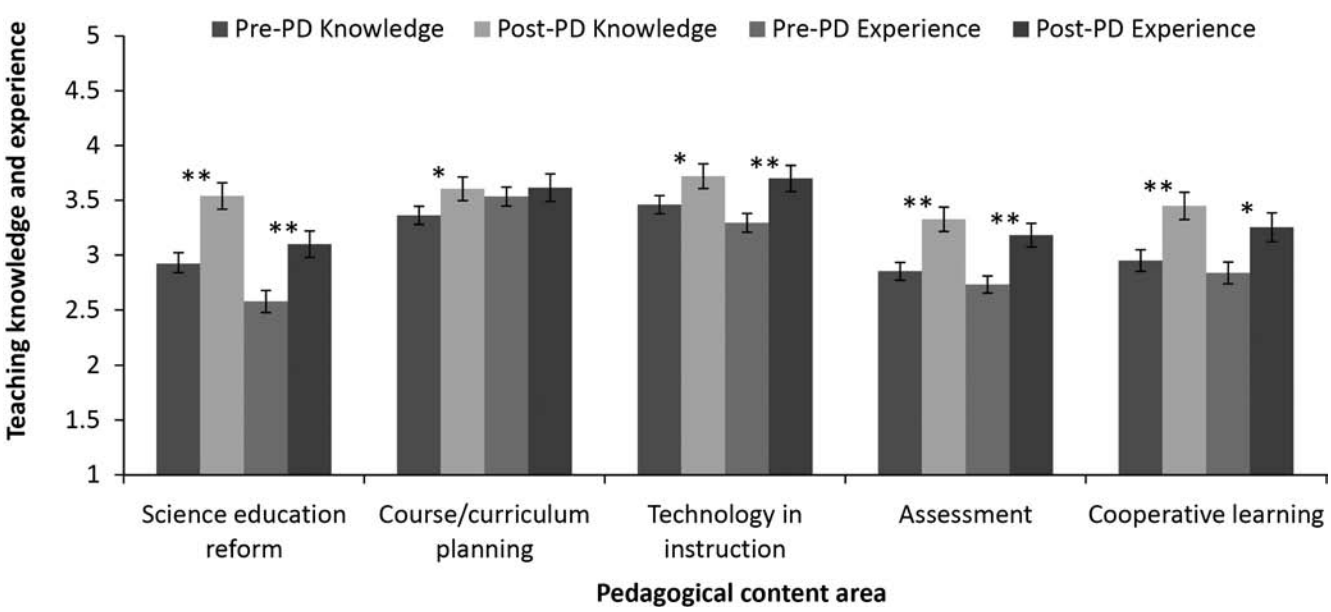

Figure 1. Mean reported gains in knowledge of and first-hand experience with different aspects of active-learning pedagogy from professional development (PD) participants. Error bars represent the standard error. The baseline (pre-PD) data were reported for Faculty Institutes for Reforming Science Teaching professional development program participants (n = 143). The final survey (post-PD) data were obtained from the Reformed Teaching Observation Protocol study (RTOP) participants (n = 77). The responses were based on a five-point Likert scale with 5 being the highest rating and 1 being lowest rating.
****p < .05, **p < .01, 1-tailed Wilcoxon test.***

Box 1. Variables derived from the self-report survey questions that were used in data analysis.

Continuous/ordinal variables

1. Years teaching at the college level
2. Active-learning knowledge (sum of Likert scale responses for knowledge of specific active-learning practices)
3. Active-learning experience (sum of Likert scale responses for experience with specific active-learning practices)
4. Percentage of the faculty's present appointment dedicated to teaching activities and responsibilities
5. Course size (number of students enrolled in a class)
6. Challenges to using active learning in the classroom (sum of Likert scale responses for challenges)
7. Departmental support (sum of Likert scale responses of departmental support of educational reform)

Categorical variables

8. Tenure status (tenured or not)
9. Professional development program completed by a participant (Faculty Institutes for Reforming Science Teaching professional development program or National Academies Summer Institute on Undergraduate Education in Biology professional development program at the University of Wisconsin)
10. Confidence in teaching (high or low)

encountered; box 1). All of the variables were derived directly from the goals of the workshops and were consistent with research on high-quality PD (Garet et al. 2001, Connolly and Millar 2006, Postareff et al. 2008). Three variables (i.e., proportion of the participants' appointments devoted to research, teaching, and service) were all highly correlated, so only the proportion of the appointment devoted to teaching was retained in the analysis. We then used a general linear models effect selection analysis (PROC GLMSELECT, SAS, Cary, North Carolina) to identify variables potentially useful as predictors of mean RTOP and subscale scores. We used the forward selection method with adjusted r^2 as the criterion for the addition of variables to the model. Data were tested for normality prior to the analyses. All statistical analyses were conducted using SAS Version 9.2 (SAS, Cary, North Carolina). Results of statistical analyses were considered to be significant at $p < .05$.

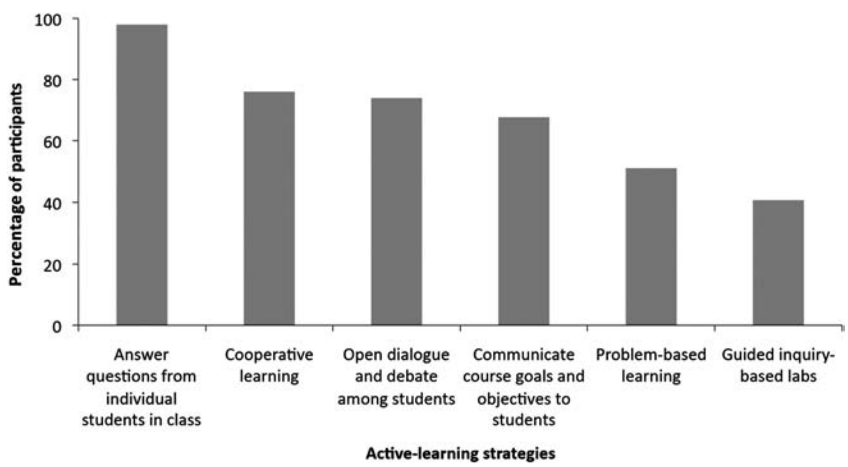

Figure 2. Percentage of professional development (PD) participants who reported the use of specific active-learning strategies on the course survey after their participation in PD (n = 96). The frequency values of strategy use included in the data were monthly, weekly, or each class period.

Results of the analyses

Depending on whether self-reported or observational data were used, the consequences of faculty participation in PD workshops differed. Survey data indicated that the PD workshops resulted in significant gains in faculty knowledge of and firsthand experience with specific aspects of reformed teaching (figure 1). On the midprogram survey, FIRST II participants were asked to list the reforms that they implemented in each of their classes and what they felt they took away from their participation in the PD project. At the end of PD, 89% of the FIRST II respondents ($n = 112$) stated that they implemented reforms in the prior academic year that stemmed from their participation in PD activities. A majority of the FIRST II and the SI faculty ($n = 96$) reported use of specific inquiry-based and learner-centered teaching practices after PD in each class period, weekly, or monthly (figure 2).

The participants also reported that they encountered multiple challenges that affected their use of learner-centered methods in their classrooms (figure 3). Over half of the faculty in the FIRST II and SI projects reported that they were *somewhat* (rating = 2) or *very* (rating = 3) challenged by each of the factors shown in figure 3. Lack of time for teaching activities was clearly the dominant challenge, whereas a lack of cooperation and support from staff, colleagues, or other entities were minor challenges. In fact, when asked about specific aspects of departmental support for undergraduate education, faculty responses were largely positive or neutral (figure 4).

In contrast with the self-reported data, observations of faculty classrooms indicated that a majority of faculty (75%)

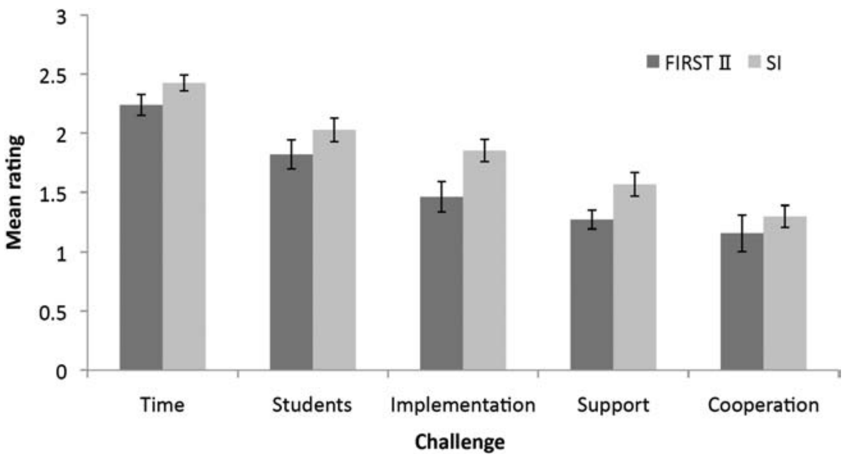

Figure 3. Mean rating of challenge categories for professional development participants on the final survey (0, n/a*; 1,* not challenging*; 2,* somewhat challenging*; 3,* very challenging*;* n *= 113). Error bars represent the standard error. We organized the challenges into five categories: (1) time constraints (time to plan, develop, or adapt materials; grade and give feedback; train colleagues or teaching assistants; and balance teaching with other activities), (2) students (attitudes toward teaching methods and course evaluation feedback), (3) implementation (classroom infrastructure and use of technology), (4) support (through the campus administration, teaching rewards, tenure, financial, staff), and (5) cooperation (of departmental and other faculty, teaching assistants, and staff). Abbreviations: FIRST II, Faculty Institutes for Reforming Science Teaching; SI, National Academies Summer Institute on Undergraduate Education in Biology.*

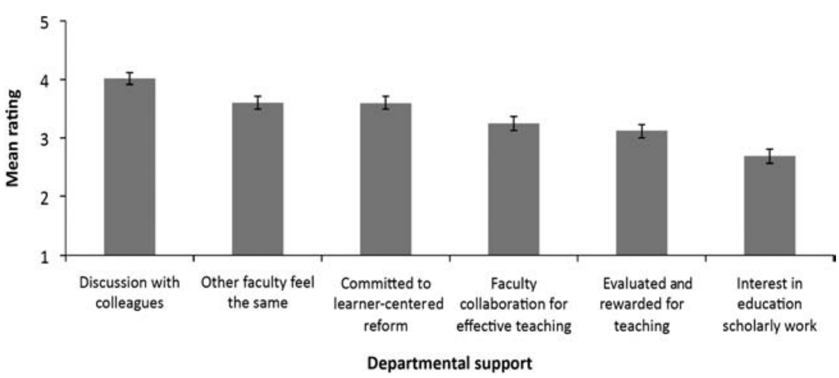

Figure 4. Mean agreement of professional development participants with final survey items that characterize their department's commitment toward undergraduate education (1, strongly disagree*; 2,* moderately disagree*; 3,* neutral*; 4,* moderately agree*; 5,* strongly agree*;* n *= 113). Error bars represent the standard error. To better understand the role of departmental support in teaching practice, we also asked the participants to indicate their level of agreement with six survey items: (1) They frequently discuss issues about improvement of teaching and learning with departmental colleagues; (2) other departmental faculty also feel a need to improve college science teaching and learning; (3) their department is committed to reforming curricula and courses to enhance active learning and inquiry-based teaching; (4) departmental faculty collaborate to achieve effective teaching; (5) departmental faculty are recognized, evaluated, and rewarded for effective teaching; and (6) departmental faculty are interested or engaged in scholarly work about teaching and learning.*

implemented a lecture-based, teacher-centered pedagogy, which was determined by mean RTOP scores for the videotapes submitted (figure 5). Furthermore, in the two years following PD, we observed no major shift in faculty practices. Fifty-seven percent of the participants were in the same RTOP category from their first to their final videotape. Of the remainder, 23% moved to a lower RTOP category following their first tape, whereas 20% moved to a higher RTOP category. There were no significant differences in the total RTOP scores or the subscale scores between faculty who participated in the FIRST II program and those who participated in the SI program (*t*-test, $p > .05$).

We next conducted analyses to identify potential predictors of faculty RTOP scores. The results of a final general linear regression were statistically significant and incorporated four variables that accounted for 19% of the variation in RTOP score (table 4). The first variable entered into the model was the participant's number of years of teaching at the undergraduate level. This value was inversely related to the mean RTOP score, accounting for 8% of the score variation. The second variable added to the model was class size, which, like teaching experience, was inversely related to RTOP score. The proportion of a faculty member's appointment devoted to teaching activities was added as the third variable and was also inversely related to RTOP score. The final model variable was the faculty member's direct experience with active-learning teaching strategies. The more experience that faculty had with diverse active-learning strategies, the higher their RTOP score. None of the remaining variables (box 1) contributed positively to the adjusted r^2 for the model, so these variables were not selected for inclusion. The use of other criteria (e.g., Akaike information criterion, Schwarz's Bayesian information criterion) or model selection approaches (e.g., least angle regression) for the inclusion or exclusion of variables into the models were more conservative, usually including only the first two predictor variables (table 4) for each RTOP component and accounting for less of the variation in the mean RTOP and subscale scores.

A multivariate analysis of the five subscales provided further insights into the total RTOP score and the predictor variables. As with the total RTOP score, 4 of the 10 potential predictor variables (i.e., years of teaching experience, class size, experience with active-learning teaching strategies, and proportion of appointment devoted to teaching) dominated the model for each subscale except propositional knowledge (table 4). Although these four independent variables were incorporated into most of the models, no single variable accounted for a high proportion of variation in a subscale (≤11%). Likewise, a major proportion of the variation in subscales and in the total RTOP score (≥78%) was not accounted for by any of the linear regression models. All of the subscale models were significant at $p \leq .1$, but only three of the models met our criteria for statistical significance at $\alpha < .05$. One additional variable—departmental support—was included in the model for propositional knowledge, but it only explained 2% of the variation in the subscale scores, and the overall model results were not statistically significant (table 4).

Final RTOP category	I	II	III	IV	V
V					0%
IV		1.5%		0%	1.5%
III	4.6%	3.1%	9.2%	1.5%	
II	10.8%	21.5%	4.6%	6.2%	
I	26.2%	7.7%	1.5%		
	Initial RTOP category				
	Teacher centered → Learner centered				

Figure 5. Matrix of Reformed Teaching Observation Protocol study (RTOP) score categories between the initial video and the final video indicate the type of teaching along a learner-centered continuum (see table 3 for an explanation of the categories). The diagonal indicates faculty whose scores remained the same over time. The numbers above the diagonal indicate gains in RTOP category, whereas the numbers below the diagonal indicate losses.

How learner centered was the pedagogy that faculty implemented following PD?

Our results were consistent with the "paradox of change without difference" (Woodbury and Gess-Newsome 2002). Faculty desire change (Samuelowicz and Bain 2001) and those in the present study stated that they had implemented active-learning teaching (figures 1 and 2). However, the direct observations indicated that the implementation of learner-centered teaching did not occur (figure 5, table 4). Teacher-centered, transmission-of-information practices prevailed, despite the faculty's participation in PD workshops designed to help them implement active-learning teaching. The scores on four of the RTOP subscales, each of which is focused heavily on student–teacher and student–student dynamics in the classroom, were low (table 4). In contrast, the faculty's scores were almost twice as high on the propositional knowledge subscale as they were on the other subscales (table 4). The propositional knowledge subscale is focused on faculty knowledge about content, deriving connections with real-life examples, and promoting coherent understanding across topics and disciplines (MacIsaac and Falconer 2002). High propositional knowledge scores reflect the ability of faculty to demonstrate their content knowledge in interdisciplinary contexts and their use of appropriate abstractions.

Table 4. Mean participant scores and the results of general linear model analyses for total RTOP score and the five subscale scores.

		Variable addition step										
		Step 1		Step 2		Step 3		Step 4				
Variable	Mean	Step	r^2	Step	r^2	Step	r^2	Step	r^2	Model r^2	F	p
RTOP score	37.12	Y	.08	C	.05	P	.03	E	.03[a]	.19	2.65	.04
Subscale scores												
Lesson design	5.99	C	.08	Y	.04	P	.03	E	.02[a]	.17	2.29	.07
Propositional knowledge	12.91	P	.08	D	.02					.10	2.54	.09
Procedural knowledge	4.77	C	.10	Y	.06	P	.03	E	.03[a]	.22	3.16	.02
Communicative interactions	6.35	C	.11	Y	.06	E	.03	P	.03	.15	3.20	.02
Student–teacher interaction	6.96	Y	.08	C	.05	E	.03[a]	E	.03	.16	3.10	.03

Note: Step number refers to the sequential step of variable addition during the model-building process.
C, class size; D, departmental support (see box 1 for details); E, cumulative firsthand experience with science education reform, course and curriculum planning, theories of learning, technology in instruction, interdisciplinary approaches to inquiry and problem solving, assessment, cooperative learning, case studies, independent projects, problem-based learning, inquiry-based laboratories, inquiry-based field projects, and teaching portfolios; P, the proportion of faculty appointment devoted to teaching activities; RTOP, Reformed Teaching Observation Protocol study; Y, the total number of years of undergraduate teaching experience.
[a]Following the r^2 value for variable addition steps indicates a positive relationship; all other r^2 values for variable addition steps indicate negative relationships.

Did self-reported data about faculty's teaching practices differ from data compiled by independent observers?

The self-reported data that we analyzed presented a very different picture of faculty teaching from that described by the observational data. On the basis of the self-reported data alone, our conclusions about faculty's teaching practices following PD would parallel those of the study of SI faculty (figures 1 and 2; Pfund et al. 2009). Specifically, faculty stated that they did implement active learning in their classrooms and the majority of the FIRST II participants (89%) stated that they reformed their teaching practices following PD. The discrepancy in results between our self-reported data and the observational data, however, supports the concerns of Kane et al. (2002) about the value and interpretation of self-reported information as a measure of teaching practices and the effectiveness of PD. We do not question whether the faculty in our study believed that they implemented active, student-centered teaching practices; the self-reported data and the faculty's written statements at the end of PD did indicate these beliefs. Indeed, most of the faculty did interact with students and used some active-learning strategies during some class sessions (i.e., RTOP category II; table 3).

Disconnects between teachers' conceptions of their practice and the reality of that practice are not new (Fung and Chow 2002). Yet the question remains, why did faculty use teacher-centered rather than student-centered practices after PD? Traditional beliefs, self-efficacy, values, level of dissatisfaction with student learning, and attitudes about teaching and learning probably influence the degree to which faculty create learner-centered classrooms, despite participation in workshops that are designed to promote such pedagogies (Gess-Newsome et al. 2003, Henderson and Dancy 2007). Rewards within departments for teaching and learning compared with those for research also influence individuals' interest in and motivation to spend time redesigning courses and implementing different pedagogies (Henderson 2008). Peer interactions about teaching set a positive or negative tone within departments, thereby encouraging or discouraging commitment to teaching as scholarship. Likewise, concerns about students' willingness to adapt to different teaching approaches and the impact of student evaluations influence faculty decisions about whether to implement learner-centered classrooms (Henderson 2008). In our study, peer and departmental support and cooperation were not perceived as barriers to implementing student-centered teaching, but practical matters of time, practical aspects of teaching implementation, and student attitudes and evaluations were cited as substantive challenges (figure 3). However, even with these reported challenges, the faculty perceived themselves as frequently using active, learner-centered teaching practices (figure 2).

What variables predict teaching practice?

The potential predictive variables that we examined (box 1) accounted for only 19% of the variation in mean RTOP score, even though they were derived from factors traditionally thought to explain teaching differences among faculty (Gess-Newsome et al. 2003, Henderson and Dancy 2007). The statistical significance of most of our models and the consistency of the predictive variables included in them indicate that those variables warrant consideration in relation to teaching reform and PD. Specifically, the most significant and consistent predictors of RTOP score were the number of years of teaching experience and class size. Of secondary importance were the proportion of the faculty's appointment devoted to teaching and their experience with curriculum reform and active-learning teaching practices (table 4). The negative relationship between years of teaching and RTOP scores indicated that the novice teachers implemented inquiry-based, learner-centered instruction to a greater extent than experienced teachers. Our results support the claim of Gibbs and Coffey (2004) that PD in teaching should occur early in one's career so it becomes the standard rather than the exception in departmental cultures. One possible explanation for our finding is that faculty who have taught longer may have greater difficulty implementing or are less inclined to implement learner-centered teaching. The class-size results, in which teaching of large classes was correlated with a lower RTOP score (i.e., the teaching was teacher centered), are consistent with the findings of Murray and MacDonald (1997). They posited that the teaching practices of instructors are influenced by their belief that once the number of students exceeds 40, the class is perceived as a lecture.

Re-envisioning faculty PD

In general, PD workshops provide some elements of what is needed to reform undergraduate science teaching, such as knowledge of why and how to use active-learning strategies (Pfund et al. 2009), use of instructional principles such as backward design (Wiggins and McTighe 2005) to develop teaching materials and classroom activities, and dealing with self-confidence and implementation challenges (Postareff et al. 2008). Clearly, more work is needed. We posit that true understanding and implementation of learner-centered teaching cannot be taught without direct practice and feedback on that practice, which parallels how students learn (Bransford 2000). The faculty in our study learned what was taught in the workshop, but they were left alone to successfully develop and implement active-learning teaching strategies. There was no on-site network of expert support. Our expectation that faculty teams would synthesize and expand on what was learned in PD workshops and would teach learner-centered courses was not met, at least not within our study period. Prather and Brissenden (2008) recognized this problem and suggested a model of *situated apprenticeships* to increase the frequency and success of faculty attempts at teaching change. Communities of practice composed of individuals with similar goals can support one another with implementation strategies (Rogan 2007). Regular and timely feedback from experts is fundamental to the PD process as faculty work to improve their classrooms (Henderson et al. 2009). Instant replays of teaching, with expert commentary, may become our most powerful tool as we strive to improve the outcomes of PD programs that could ultimately improve student learning.

Acknowledgments

We thank Debra Linton and Everett Weber for data collection and Ken Fairbanks for statistical advice. Special thanks to the National Science Foundation for the long-term support of the FIRST project (Faculty Institutes for Reforming Science Teaching) under Division of Undergraduate Education Award 0088847. This research was reviewed by Michigan State University's Institutional Review Board and classified as expedited.

References cited

[AAAS] American Association for the Advancement of Science. 1990. The Liberal Art of Science: Agenda for Action. AAAS.

Alberts B. 2008. Considering science education. Science 319: 1589.

Blanchard MR, Southerland SA, Osborne JW, Sampson VD, Annetta LA, Granger EM. 2010. Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. Science Education 94: 577–616.

Bransford J. 2000. How People Learn: Brain, Mind, Experience, and School, expanded ed. National Research Council.

Brewer C, Smith D, eds. 2011. Vision and Change in Undergraduate Biology Education: A Call to Action. American Association for the Advancement of Science.

Carnegie Foundation for the Advancement of Teaching. 2001. The Carnegie Classification of Institutions of Higher Education, 2000 ed. Carnegie Foundation for the Advancement of Teaching.

Clarke D, Hollingsworth H. 2002. Elaborating a model of teacher professional growth. Teaching and Teacher Education 18: 947–967.

Connolly MR, Millar SB. 2006. Using workshops to improve instruction in STEM courses. Metropolitan Universities 17: 53–65.

DeHaan RL. 2005. The impending revolution in undergraduate science education. Journal of Science Education and Technology 14: 253–269.

Derting TL, Ebert-May D. 2010. Learner-centered inquiry in undergraduate biology: Positive relationships with long-term student achievement. Cell Biology Education 9: 462–472.

Fung L, Chow LPY. 2002. Congruence of student teachers' pedagogical images and actual classroom practices. Educational Research 44: 313–321.

Gaffney JDH, Richards E, Kustusch MB, Beichner RJ. 2008. Scaling up education reform. Journal of College Science Teaching 37: 18–23.

Garet MS, Porter AC, Desimone L, Birman BF, Yoon KS. 2001. What makes professional development effective? Results from a national sample of teachers. American Educational Research Journal 38: 915–945.

Gess-Newsome J, Southerland SA, Johnston A, Woodbury S. 2003. Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. American Educational Research Journal 40: 731–767.

Gibbs G, Coffey MG. 2004. The impact of training of university teachers on their teaching skill, their approach to teaching and the approach to learning of their students. Active Learning in Higher Education 5: 81–101.

Handelsman J, Miller S, Pfund C. 2007. Scientific Teaching. Freeman.

Henderson C. 2008. Promoting instructional change in new faculty: An evaluation of the physics and astronomy new faculty workshop. American Journal of Physics 76: 179–187.

Henderson C, Dancy MH. 2007. Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. Physics Education Research 3: 020102.

Henderson C, Beach A, Famiano M. 2009. Promoting instructional change via co-teaching. American Journal of Physics 77: 274–283.

Holdren JP. 2009. Science in the White House. Science 324: 567.

Kane R, Sandretto S, Heath C. 2002. Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. Review of Educational Research 72: 177–228.

Knight JK, Wood WB. 2005. Teaching more by lecturing less. Cell Biology Education 4: 298–310.

Lawson AE, Benford R, Bloom I, Carlson MP, Falconer KF, Hestenes DO, Judson E, Pilburn MD, Sawada D, Wycoff S. 2002. Reforming and evaluating college science and mathematics instruction. Journal of College Science Teaching 31: 388–393.

Lundmark C. 2002. The FIRST project for reforming undergraduate science teaching. BioScience 52: 552.

MacIsaac D, Falconer K. 2002. Reforming physics instruction via RTOP. Physics Teacher 40: 479–485.

Michael J. 2006. Where's the evidence that active learning works? Advances in Physiology Education 30: 159–167.

Murray K, MacDonald R. 1997. The disjunction between lecturers' conceptions of teaching and their claimed educational practice. Higher Education 33: 331–349.

Pfund C, et al. 2009. Summer Institute to Improve University Science Teaching. Science 324: 470–471.

Postareff L, Lindlom-Ylanne S, Nevgi A. 2008. A follow-up study of the effect of pedagogical training on teaching in higher education. Higher Education 56: 29–43.

Prather EE, Brissenden G. 2008. Development and application of a situated apprenticeship approach to professional development of astronomy instructors. Astronomy Education Review 7: 1–17.

Prince M. 2004. Does active learning work? A review of the research. Journal of Engineering Education 93: 223–231.

Rogan JM. 2007. How much curriculum change is appropriate? Defining a zone of feasible innovation. Science Education 91: 439–460.

Samuelowicz K, Bain JD. 2001. Revisiting academic's beliefs about teaching and learning. Higher Education 41: 299–325.

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.

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 Science and Mathematics 102: 245–253.

Shrout PE, Fleiss JL. 1979. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin 86: 420–428.

Sunal DW, Hodges J, Sunal CS, Whitaker KW, Freeman LM, Edwards L, Johnston RA, Odell M. 2001. Teaching science in higher education: Faculty professional development and barriers to change. School Science and Mathematics 101: 246–257.

Wiggins G, McTighe J. 2005. Understanding by design, 2nd ed. Association for Supervision and Curriculum Development.

Wood WB. 2009. Innovations in teaching undergraduate biology and why we need them. Annual Review of Cell and Developmental Biology 25: 93–112.

Woodbury S, Gess-Newsome J. 2002. Overcoming the paradox of change without difference: A model of change in the arena of fundamental school reform. Educational Policy 16: 763–782.

Diane Ebert-May (ebertmay@msu.edu) is a professor, Tammy M. Long is an assistant professor, and Sarah E. Jardeleza is a postdoctoral researcher, with the Department of Plant Biology at Michigan State University, in East Lansing. Terry L. Derting is a professor in the Department of Biological Sciences at Murray State University, in Murray, Kentucky. Janet Hodder is a faculty member at the Oregon Institute of Marine Biology, University of Oregon, in Charleston. Jennifer L. Momsen is an assistant professor in the Department of Biological Sciences at North Dakota State University, in Fargo.