Direct Observations of the Outcomes of Faculty Professional Development

Diane Ebert-May, *Michigan State University*
Terry Derting, *Murray State University*
Janet Hodder, *University of Oregon*

Introduction

Professional development (PD) of faculty is an integral component of curriculum reform efforts in STEM. Traditionally, PD occurs through workshops that last from hours to several days. Regardless of the particular model of PD used during a workshop, its effectiveness is usually assessed through self-reported surveys of faculty satisfaction, perceived learning, and reports of applications by faculty in their classrooms. We focus on ways of assessing the effectiveness of PD models, with an emphasis on the need for objective measures of change in faculty teaching (Ebert-May et al. 2011).

The data we present raise two significant questions about professional development of faculty who teach undergraduates in STEM. Are traditional approaches to faculty PD effective in changing classroom practices and improving student learning? What evidence is needed to determine the effectiveness of different PD models?

Self-reported data are useful in identifying variables that can influence the extent to which faculty implement new teaching strategies (Henderson et al. 2012). These variables include faculty beliefs about student learning, self-efficacy, level of dissatisfaction with student learning, departmental rewards for teaching and learning, time limitations and peer interactions. Self-reported data do not, however, provide a complete or necessarily accurate assessment of the impacts of PD on classroom practices and student learning. Objective assessment of teaching and learning is also necessary, yet seldom conducted.

In this report, we summarize and disseminate recently published data (Ebert-May et al 2011) to illustrate one of several approaches that we used to assess professional development, specifically, direct observations of faculty teaching. We focused on two national professional development programs: Faculty Institutes for Reforming Science Teaching (FIRST II; Hodder and Ebert-May, NSF DUE 88847) and the National Academies Summer Institutes (SI) at the University of Wisconsin, Madison (Handelsman and Wood, funded by the Howard Hughes Medical Institute). In the FIRST II program, faculty attended workshops for a total of 6-12 days over a period of three years. These faculty were from all types of institutions, ranging from community colleges to research institutions. In the SI program, faculty attended a 5-day institute during the summer and all participants were from research institutions. The workshop goals were similar for both programs and were based on the principles of scientific teaching (Handelsman et al. 2004). The courses targeted for reform were at the introductory biology level (e.g., cell/molecular biology, organismal/population biology, ecology, genetics, and evolution). Teams of faculty participated in the workshops and designed instructional units that included learning objectives, assessments of student learning that were aligned with the objectives, and active, learner-centered teaching strategies such as cooperative learning.

How did faculty teach following professional development programs?

First, we examined the self-reported and directly observed teaching practices by faculty after completion of the FIRST II and SI workshops. For FIRST II faculty, self-reported data were used to determine change in faculty knowledge of and experience with different aspects of active-learning pedagogy. As expected, there were significant improvements in knowledge of each pedagogical area. There were also significant improvements in faculty perceptions of their first-hand experience with each pedagogical area (e.g., science education reform, use of technology in instruction, assessment, cooperative learning), except for course and curriculum planning. Faculty already had substantial experience with the latter variable before the workshops (Ebert-May et al., 2011). For SI faculty, there were significant increases in pedagogical knowledge before and after PD, as well (Pfund et al. 2009). These self-reported data sets from faculty participants in FIRST II and SI were remarkably consistent.

Faculty from both FIRST II and SI also reported perceptions of their use of active-learning strategies after completing their PD workshops. A majority of the faculty reported use of specific inquiry-based or learner-centered teaching practices (Fig. 1). These practices were used at least weekly or monthly, and often in each class period. So according to these data, more than half of the faculty were using active-learning techniques one year after completing professional development (Ebert-May et al. 2011, Pfund et al. 2009).

We collected two to four videos from faculty in FIRST II and SI teaching their students after the workshops (see Ebert-May et al. 2011 for details). "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, learner-centered 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)" (Ebert-May et al. 2011 p. 552). An RTOP score is an indicator of the degree of active learning instruction and student involvement observed in a classroom (Sawada et al. 2002) and can be categorized into five progressive categories of teaching practice. Categories I and II indicate teacher-centered classrooms that range from

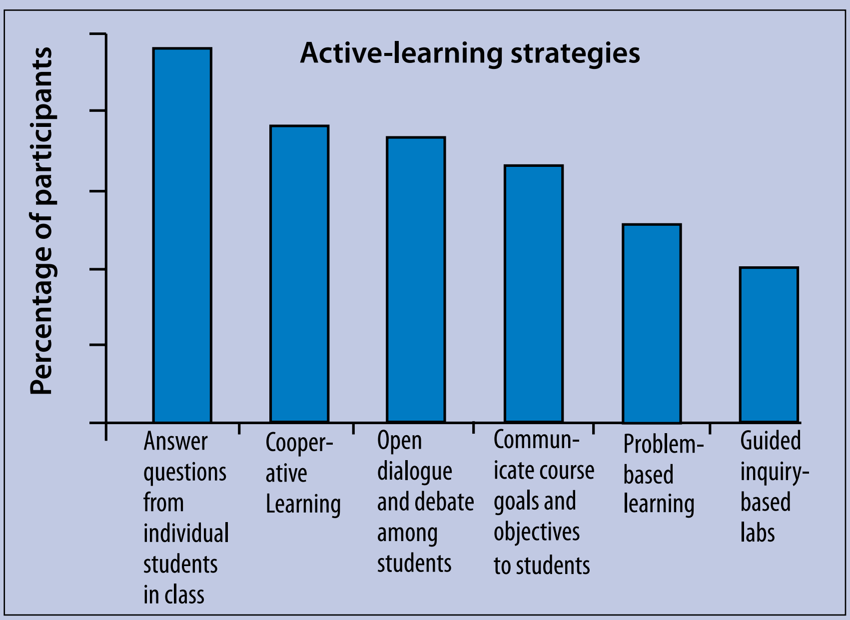

Fig. 1. Reported use of active learning strategies by FIRST II faculty after professional development (n=96; Ebert-May et al. 2011 p. 554)

straight lecture to minor student involvement. Categories III-V indicate increasing levels of learner-centered classrooms.

We examined the change in participants' RTOP scores from when they taught right after they completed the first PD workshop to when they taught one to two years later. "In contrast with the self-reported data, observations of faculty classrooms indicated that a majority of faculty (75%) implemented a lecture-based, teacher-centered pedagogy, which was determined by mean RTOP scores for the videotapes submitted. 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, $p > .05$)." (Ebert-May et al 2011, p. 554-555).

The directly observed data conflicted with the self-reported data from participants who indicated that they implemented the active-learning practices they learned in the FIRST II and SI workshops. What happened? Here we note that faculty did not intentionally report misinformation; in fact, they were truly motivated and excited about changes in their courses. However, their ***perceptions*** of teaching did not match their teaching practice. We hypothesize that the faculty did not fully understand what active, learner-centered teaching is and how to implement it, nor did faculty change their beliefs about how students learn and how they teach (Henderson et al. 2011). Thus, although faculty did implement components of learner-centered teaching, it was not of sufficient depth and breadth to transform the basic nature of their teaching practice.

In terms of types of data used to evaluate professional development programs, we claim that self-reported data are useful, especially for formative evaluation, but are incomplete. Direct measures of faculty practice are necessary.

What variables predict teaching practices of faculty?

To help us better understand faculty teaching practice, we used the data collected to try to predict the type of teaching implemented by faculty. We wanted to know what variables predict teaching practice. Based on our experience with PD and relevant published literature, we predicted that the following variables were important: (1) ***experience***, defined by number of years teaching, knowledge and practice with active learning, type of professional development program, and self-confidence; (2) ***class size***, including all of the associated challenges with implementing anything new with a large number of students; and (3) ***faculty appointment***, in terms of percent teaching, tenure status, and departmental support for teaching.

The results indicated that the predictor variables for teaching practices in our model accounted for 19% of the variation in mean RTOP score that contributed to explaining observed classroom teaching after PD (Table I). For example, faculty with less teaching experience engaged in more learner-centered teaching compared with faculty with more years of teaching experience. Also, department and peer support for faculty use of non-lecture approaches to teaching had no significant relationship with the classroom practice used by faculty. These data suggest that assumptions about the

Table I. Results of a general linear model analysis for total RTOP score. The sign in parentheses indicates whether the relationship between the variable and RTOP score was positive or negative.

Variable Entered into Model	*Model* r^2
Teaching experience (-)	0.08
Class size (-)	0.13
Proportion for teaching[1] (-)	0.16
Experience with reform[2] (+)	0.19

1. The proportion of faculty appointment devoted to teaching activities.
2. 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 project, problem-based learning, inquiry-based laboratories, inquiry-based field projects, and teaching portfolios (from Ebert-May et al. 2011).

effectiveness of traditional models of professional development need validation using both objective and subjective measures. The data also indicate a need for new models of PD for STEM faculty that include multiple-year programs, formative feedback about teaching (e.g., mentoring), and reform of an entire course. New models of PD must be coupled with valid and reliable measures of student performance. Evaluation of the effectiveness of models of faculty professional development must be rigorous and data-driven.

What is the reformed model of professional development?

We used the data from the study of FIRST II and SI to revise and implement Faculty Institutes for Reforming Science Teaching, now called FIRST IV. We changed the target audience from faculty to future faculty; that is, FIRST IV focuses on postdoctoral (postdoc) scholars in the biological sciences. FIRST IV is a national professional development program designed to shape postdocs' beliefs about and abilities to implement learner-centered biology courses that result in improved student learning. FIRST IV is currently training 200 postdocs (approximately 75 of whom are now in faculty positions) in learner-centered teaching and course design, impacting ~10,000 undergraduate students annually who learn science by doing science, even in large enrollment courses. To date, objective and self-reported assessment data provide consistent evidence that the professional development model is effective, resulting in postdocs who successfully design, implement, and teach inquiry-based, learner-centered courses. Compared with FIRST II faculty, significantly more FIRST IV postdocs are implementing learner-centered classrooms (Fig. 2).

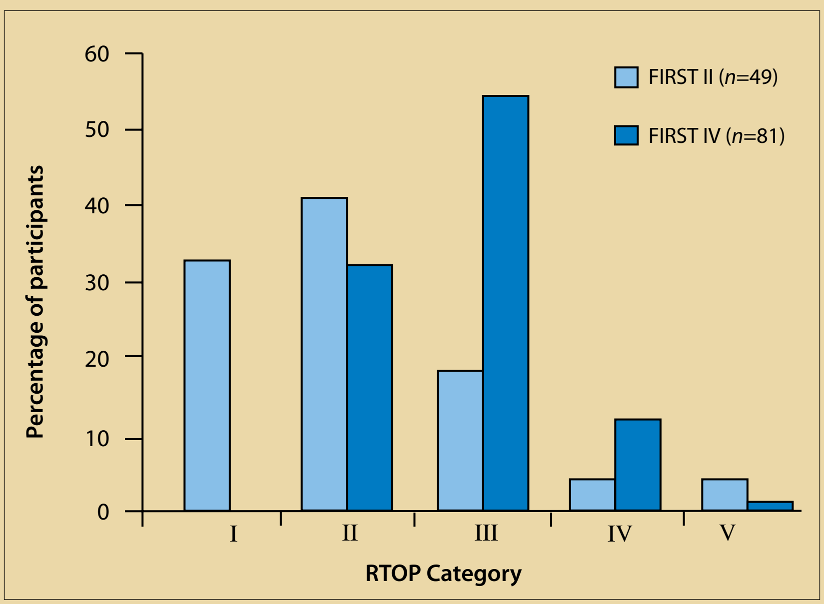

Fig. 2. Distribution of RTOP scores from videos of classroom teaching by FIRST II faculty and FIRST IV post-docs.

The key outcome for any professional development model is student learning. Course materials (syllabi, learning objectives and assessments) can be evaluated in terms of the alignment of objective and assessment item. The degree of correlation in Bloom's scores provides a proven method for measuring alignment of assessments with objectives, an indicator of successful backward design (Freeman et al. 2011, Momsen et al. 2010). What do assessment data tell us about instruction? If faculty members want students to achieve higher-cognitive skills in the process of learning science, students need to practice these skills both in and outside class using student-centered pedagogies. What faculty members want students to know and be able to do is reflected in the exams.

Performance and self-reported data from students indicated improved learning and classroom environments in courses taught by FIRST IV PDs. Evidence also shows that the FIRST IV model of professional development provides a well-established support network for PDs as they practice teaching and begin academic positions.

References

Ebert-May, D., Derting, T.L., Hodder, J., Momsen, J.L., Long, T. M., & Jardeleza, S.E. (2011). "What we say is not what we do: Effective evaluation of faculty development programs," *BioScience* **6**(17), 550-558.

Handelsman, J., Ebert-May, D., Beichner, R., Bruins, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Steward, J., Tilghman, S., & Wood, W. (2004). "Scientific teaching," *Science* **304**, 521-522.

Henderson, C., Beach, A., & Finkelstein, N. (2011). "Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature," *Journal of Research in Science Teaching,* **48**(8), 952–984.

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?" *Physical Review Special Topics. Physics Education Research* **8**, 020104.

Lawson, A.E., Benford R, Bloom, I., Carlson, M.P., Falconer, K.F., Hestenes, D.O., Judson, E., Pilburn, M.D., Sawada, D., & Wycoff, S. (2002). "Reforming and evaluating college science and mathematics instruction," *Journal of College Science Teaching* **31**, 388–393.

MacIsaac, D. & Falconer, K. (2002). "Reforming physics instruction via RTOP," *The Physics Teacher* **40**, 479–485.

Pfund, C., et al. (2009). "Summer Institute to improve university science teaching," *Science* **324**, 470–471.

Sawada, D., Piburn, M.D., 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.