

## Article

# Teaching Assistant Professional Development in Biology: Designed for and Driven by Multidimensional Data

Sara A. Wyse,\* Tammy M. Long,<sup>†</sup> and Diane Ebert-May<sup>†</sup>

\*Department of Biological Science, Bethel University, St. Paul, MN 55112; <sup>†</sup>Department of Plant Biology, Michigan State University, East Lansing, MI 48824

Submitted June 11, 2013; Revised March 18, 2014; Accepted March 18, 2014

Monitoring Editor: Jeff Schinske

Graduate teaching assistants (TAs) are increasingly responsible for instruction in undergraduate science, technology, engineering, and mathematics (STEM) courses. Various professional development (PD) programs have been developed and implemented to prepare TAs for this role, but data about effectiveness are lacking and are derived almost exclusively from self-reported surveys. In this study, we describe the design of a reformed PD (RPD) model and apply Kirkpatrick's Evaluation Framework to evaluate multiple outcomes of TA PD before, during, and after implementing RPD. This framework allows evaluation that includes both direct measures and self-reported data. In RPD, TAs created and aligned learning objectives and assessments and incorporated more learner-centered instructional practices in their teaching. However, these data are inconsistent with TAs' self-reported perceptions about RPD and suggest that single measures are insufficient to evaluate TA PD programs.

## INTRODUCTION

In universities throughout the United States, graduate teaching assistants (TAs) are responsible for a significant proportion (as high as 91%; Sundberg *et al.*, 2005) of undergraduate instruction in science, technology, engineering, and mathematics (STEM) disciplines, particularly in introductory laboratory courses (Rushin *et al.*, 1997; Gardner and Jones, 2011). TA roles vary from complete independence in course instruction to grading and leading recitation sections. As a result of the increasing reliance on graduate TAs, many schools offer programs or workshops that provide "TA training." However, lack of formal, substantive professional development (PD) focused on teaching for TAs represents a stark contrast to the state license requirements K–12 teachers must attain to teach in public schools (Tanner and Allen, 2006). Although graduate students are building expertise in the discipline they teach, evidence suggests that effective teaching requires much

more than knowledge of subject matter (Ball and Bass, 2000; Gardner and Jones, 2011). For example, pedagogical content knowledge, the intersection between one's content knowledge and the ability to think about the best ways to approach teaching that content to students, is also essential for effective instruction (Shulman, 1986).

Despite the reliance on TAs for teaching, few studies explore the design, execution, and evaluation of TA PD programs. What is clear from the literature is that the presence and scope of TA PD programs at institutions vary widely—both within and among disciplines (Gardner and Jones, 2011). Most TA PD programs are non-discipline specific and emphasize course management and logistics (Abbott *et al.*, 1989; Gardner and Jones, 2011). Few PD programs offer extensive training in designing instructional materials or using effective pedagogical practices (e.g., Marincovich *et al.*, 1998; Luft *et al.*, 2004; Baviskar and Beardsley, 2006). Those programs that do offer pedagogical training often provide it in the form of a separate course in which TAs learn about the theory and practice of teaching (e.g., Preparing Future Faculty Program courses, courses that contribute to college teaching certification; Hammrich, 2001; McManus, 2002; Roehrig *et al.*, 2003; Bond-Robinson and Rodrigues, 2006; Baumgartner, 2007). Rarely do TAs, during their appointment to teach a course, get simultaneous and targeted PD relevant to the courses they are currently teaching (e.g., Gormally *et al.*, 2011).

For decades, national reports have advocated the need to reform STEM education (e.g., National Research Council [NRC], 2003; American Association for the Advancement

DOI: 10.1187/cbe-13-06-0106

Address correspondence to: Sara A. Wyse (s-wyse@bethel.edu).

© 2014 Wyse *et al.* CBE—Life Sciences Education © 2014 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-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB<sup>®</sup>" and "The American Society for Cell Biology<sup>®</sup>" are registered trademarks of The American Society for Cell Biology.

of Science [AAAS], 2011; Henderson *et al.*, 2011). Because TAs teach large numbers of STEM undergraduates, they have a substantive role in implementing reform goals, yet few research studies have critically examined the effectiveness of preparation and implementation of teaching by TAs in science classrooms (Gardner and Jones, 2011). The majority of data about TA PD programs come from self-reported surveys or interviews that are often administered to TAs at the conclusion of their training (see Table 1 in Gardner and Jones, 2011). For example, Gardner and Jones (2011) reviewed available literature about research on TA training programs and found 45% of studies reported that TAs gained conceptual understanding of the theory and approach to learner-centered instruction, and 27% found that TAs perceived the learning experience to be effective. Only one study reported data about the teaching *practices* of TAs, although not through direct observation of practices but through students' perception of effectiveness as measured by survey instruments (Hampton and Reiser, 2004).

Our study builds upon existing research about TA PD by offering a descriptive framework for the design and evaluation of a TA training program for college-level introductory biology labs. Our design research method is grounded in how people learn and was a direct response to a larger, programmatic effort to transition toward a more inquiry-based curriculum. We adapted an existing evaluation framework that uses multiple sources of data to characterize TA outcomes and applied it before, during, and following implementation of the reformed curriculum. Our study is unique in its use of a multidimensional evaluation of TA PD that incorporates both TAs' perceptions about training and independent measures of their classroom practices.

## THE STUDY DESIGN

Our study was conducted between the Spring of 2008 and the Spring of 2009 and focused on graduate TAs teaching the laboratory component of Bio1, a four-credit introductory biology course for science, engineering, and social science

majors. Bio1 was the first in a two-course sequence required for majors and focused on genetics, evolution, and ecology. A second course (Bio2) addressed cell and molecular biology. Bio1 enrolled ~2000 students per year and consisted of a 3-h lecture taught by faculty from different departments within natural sciences and a 3-h laboratory taught by TAs.

During the Spring (16-wk semester) and Summer (8-wk semester) of 2008, TAs received PD closely aligned to what the literature reports as common experiences (described below), and what we term "traditional PD (TPD)." In each week of the Spring semester (TPD1), TAs taught two 3-h labs and participated in one 3-h preparatory (i.e., "prep") meeting in which they received their training. In the Summer semester (TPD2), TAs taught four 3-h labs and participated in two 3-h prep meetings each week. Because TPD1 and TPD2 represent two iterations of the same PD (i.e., identical training materials and the same training methods), they were collapsed into one type of PD, named "TPD." Reformed PD (RPD; also described below) was implemented in Fall 2008 (16-wk semester; RPD1) and continued through Spring 2009 (16-wk semester; RPD2). We distinguish RPD1 from RPD2 to account for differences in TA involvement in reform activities (described below). In all cases, PD took place exclusively in the context of 1) a precourse orientation meeting, 2) weekly prep meetings, and 3) a final course meeting; teaching and training time were equivalent for all PD participants.

## STUDY PARTICIPANTS

Graduate students were appointed to teaching assistantships in Bio1 through their home departments. Of 38 TAs teaching Bio1, 31 volunteered to participate in this study, including 23 doctoral students in plant biology, zoology, or fisheries and wildlife departments, and 8 in forensic science and anthropology departments. Of the 31 participating TAs, seven taught Bio1 in more than one semester of the study (Table 1). Repeat TAs were counted in each semester they participated, accounting for a total of 38 TAs. Before the start of the research, TAs had an average of  $2.2 \pm 0.63$  semesters of

**Table 1.** TA cohort information<sup>a</sup>

Group	Total TAs	Total TA participants	Repeat TA participants	Survey participants		Video participants	
				First-time	Repeat	First-time	Repeat
TPD	14	12	–	12	–	9	–
RPD1	17	15	3 (TPD)	12	3 (TPD)	10	1 (TPD)
RPD2	13	11	1 (TPD and RPD1) 4 (RDP1)	6	1 (TPD and RPD1) 4 (RDP1)	5	4 (RPD1)
Total	44	38 <sup>b</sup>	7 <sup>c</sup>	30	7 <sup>c</sup>	24	5

<sup>a</sup> Across the three groups (TPD, RPD1 and RPD2), 44 total TAs taught Bio1. Of those, 38 (86%) participated in some aspect of this research. Of the 38 participating TAs, seven participated in more than one semester, accounting for 31 unique individuals. Those 31 participants participated in surveys, videos, or both. Sample sizes labeled "first-time" represent the number of TAs participating in that portion of the study for the first time. For example, 12 TAs who had not yet participated in surveys participated during RPD1. Sample sizes labeled "repeat" indicate the number of TAs participating in this portion of the study who have previously participated (e.g., in TPD or RPD1), as indicated in parentheses. For example, one TA who completed surveys during RPD2 also participated during TPD and RPD1. Totals represent column sums.

<sup>b</sup> The number of repeat TAs is seven. While eight repeats exist (three plus one plus four), one TA who taught in RPD2 also repeated from RPD1. He/she was the only TA to repeat across more than one group, which brings the number of repeat TAs to seven.

<sup>c</sup> The total number of unique participants is equivalent to the total number participating less the repeats:  $38 - 7 = 31$ .

**Table 2.** Demographic information for TA PD participants<sup>a</sup>

Criteria	TPD	RPD1	RPD2
Number of first-time TAs	12	12	6
Percent female (%)	67*	33*	50*
Percent working toward a PhD (%)	59	75	67
Percent with a prior course in teaching (%)	25	17	33
Percent with a prior seminar on teaching (%)	33	33	33
Average number of semesters teaching in Bio1, including present	1.92 ± 0.38	1.67 ± 0.37	2.33 ± 0.66
Average number of prior semesters teaching at the university level (excluding Bio1)	3.92 ± 1.09	3.08 ± 1.01	4.00 ± 1.15
Average number of semesters of teaching outside the current institution	3.67 ± 1.48*	1.00 ± 0.44*	2.33 ± 0.95*

\*Indicates a significant difference ( $\alpha = 0.05$ ).

<sup>a</sup>For the first-time participants in the study (based on survey data, Table 1), the demographic characteristics are displayed broken down by semester of PD (TPD, RPD1, and RPD2). For the average rows, values represent mean ± SE. The percent of TAs who were female (i.e., gender) differed among the TPD and RPD groups ( $\chi^2 = 13.16$ ,  $df = 2$ ,  $p = 0.001$ ). The only other significant difference identified was the number of semesters of teaching experience outside the current institution (ANOVA:  $F_2 = 4.17$ ,  $p = 0.0206$ ). Here, TPD differed from RPD1 and RPD2, and RPD1 differed from RPD2, but not from TPD.

teaching experience that included teaching at nature centers (nine TAs) and in K–12 classrooms (two TAs).

We tested for equivalence between RPD and TPD TA populations (Table 2) using a chi-square test for independence for categorical data (gender, degree [e.g., MS versus PhD candidate], department, undergraduate institution type, undergraduate major, enrollment in a course or seminar focused on teaching) and analysis of variance (ANOVA) tests for continuous data (e.g., number of years in program, semesters of prior teaching experience, semesters teaching Bio1). We found that two demographic variables yielded significant values: 1) Gender ( $\chi^2 = 13.16$ ,  $df = 2$ ,  $p = 0.001$ ) was significantly skewed toward males in RPD1 (Table 2). However, when TAs in the study were counted once (i.e., as 31 individuals), the number of males ( $n = 16$ ) and females ( $n = 15$ ) were roughly equivalent. 2) Semesters of teaching experience outside the current institution was significantly higher for TPD (ANOVA,  $F_2 = 4.17$ ,  $p = 0.0206$ ); however, this difference was largely driven by one TA who taught 3 yr of middle/high school biology accounting for the larger mean and SE (Table 2). No other significant differences were detected.

## DESCRIPTION OF PD

### TPD

In Bio1, TPD is defined as the PD approach used for TAs during the previous 5 yr. In TPD, TAs began with a presemester orientation in which they learned about procedural expectations (e.g., lab safety) and lab policies regarding office hours, lab exams, and grades. TAs received laboratory manuals, had opportunities to meet other TAs teaching the same course, and set up their lab syllabi; student learning and pedagogical practice were not major emphases in TPD orientations.

During the semester, TAs met weekly for prep meetings lasting 3 h. The first hour included a lab overview provided by an experienced TA or prep meeting leader (e.g., university staff). The overview was a lecture format with supporting slides that focused on laboratory content TAs were expected to teach their students during the subsequent week, relevant points to emphasize with students, and suggestions about concepts or techniques that were typically challenging for

students. Although TAs occasionally asked questions of the presenter, the format was not discussion based. Similarly, materials presented provided few, if any, questions for TAs to ask their students or suggestions for engaging students in discussion. TAs spent the remaining 2 h of the prep meeting becoming familiar with the laboratory equipment and practicing lab procedures. In TPD, TAs were responsible for constructing all assessments (e.g., quizzes, homework assignments, and midterm and final exams) for their lab sections. In the weeks preceding the midterm and final exams, prep meetings provided opportunities for TAs to give and receive feedback about exam items from instructional staff and peers and to discuss grading and/or classroom management issues.

### RPD

Beginning in Fall 2009, the presemester orientation shifted toward a workshop-style experience (described in the Supplemental Material) in which TAs learned about backward design (Wiggins and McTighe, 1998) and scientific teaching—that is, teaching science as it is practiced using tested pedagogies (Handelsman *et al.*, 2004). As part of RPD1, TAs worked in small teams to modify the existing labs of Bio1 to make them more inquiry based. These changes included revising learning objectives to focus on higher-order cognitive skills, removing extraneous information in laboratory descriptions, devising new assessments, and ultimately revising labs from procedural step-by-step directions to guided inquiry. TAs in RPD2 received a similar workshop experience but did not participate in revising or developing instructional materials for the lab. Instead, RPD2 TAs implemented materials revised and/or developed by RPD1 TAs and instructional staff.

In practice, the TPD and RPD models are comparable in seven dimensions (Table 3) and offer two distinct approaches to PD.

**Dimension 1. Theoretical Framework.** The RPD model is based on the learning theories of constructivism (von Glasersfeld, 1989) and cooperative learning (Johnson *et al.*, 2000), which state that students learn more effectively when provided opportunities for student interaction and collaboration; the combination may also be referred to as *social*

**Table 3.** Dimensions of RPD model in this study and TPD models<sup>a</sup>

Dimension	TPD model	RPD model
1. Theoretical framework	Behaviorism (transmission model)	Constructivism, cooperative learning
2. Instructional design	Lecture based, answer driven, protocol oriented	Collaborative, inquiry driven, process oriented
3. Leader role	Information provider	Facilitator
4. TA role	Passive, listener, recipient of knowledge	Active, participant
5. Goals of preparation	Content focused	Learning focused
6. Formal reflection	End-of-semester surveys	Iterative, embedded in weekly discussions
7. Length of preparation	Continuous and longitudinal	Continuous and longitudinal

<sup>a</sup>Seven dimensions highlight key differences between the two types of PD explored in this research. TPD includes PD that was focused on answers and protocols and in which leaders were viewed as the source of information and TAs were seen as passive recipients of knowledge. In contrast, RPD focused on processes and inquiry, allowing TAs to experience learner-centered instruction during their PD; this shifted TAs from passive consumers to active participants in their PD.

*constructivism* (Vygotsky, 1978). These learning theories contrast with models of TA training reported in the literature (and our TPD model), which focus on procedural information and preparing for course logistics and content, rather than on effective pedagogy (e.g., Fiszer, 2004; see also Gardner and Jones, 2011). In these models, if pedagogical instruction is included, facilitators typically tell TAs about effective techniques (Gardner and Jones, 2011) and expect them to subsequently implement those techniques in their classroom. Research, however, has shown that transmission of information in workshops or similar PD venues is no guarantee of either TA comprehension or ability to apply the new knowledge in practice (Michael, 2006). Further, the transmission approach contrasts with theories about adult learners that suggest that the best learning occurs when the learners interact with the material and others to build conceptual understanding (Taylor *et al.*, 2000). Therefore, in our RPD model, TAs worked in small teams of three to four members throughout all aspects of training, and teams served as units of feedback for teaching issues and practices (e.g., ideas for engaging students in class discussions, how to effectively use groups in lab tasks, reflecting on classroom experiences).

**Dimension 2. Instructional Design.** RPD was designed so that TAs had opportunities to engage with the content in the same way they were expected to teach it—that is, by considering the perspectives of both teacher and learner. In TPD, TAs obtained information about the labs they were to teach by receiving information delivered by the prep meeting leader and by practicing lab procedures. These activities provided opportunities for TAs to learn or relearn content or protocols that were new or less familiar. In this way, their engagement with the material was similar to what would be experienced by students in their classrooms—students would receive information from the TAs, learn new content, and conduct unfamiliar lab procedures.

Early in RPD1, lab facilitators adopted a similar model, but included pedagogical practices as a part of the lab instruction for TAs. Facilitators described learner-centered methods for engaging student discussion, creating opportunities for inquiry, and managing group activities. Within a few weeks of this training, TAs expressed frustration about the new approach and felt that they were unable to effectively implement the learner-centered practices that were a focus of the train-

ing. One TA stated, “I agree with your general approach, but I just don’t know what I’m supposed to do when I have no idea what [this kind of teaching] looks like.” In response to TA concerns, the facilitators changed to an approach called the “fishbowl” or “inner circle” (McKeachie, 1999). In subsequent meetings, TAs *observed* desired instructional practices as facilitators transparently modeled learner-centered approaches as they taught the lab to a group of undergraduate learning assistants (see Supplemental Material). Because the facilitators were familiar with the literature on teaching, student learning, and instructional design, they were able to demonstrate reformed teaching practices in their own instruction. As instruction was being modeled, TAs were free to stop “class” and ask questions about pedagogical methods, discuss alternative approaches, and reflect on the instruction taking place (Mezirow, 2003). Importantly, TAs could openly discuss concerns or issues they had in thinking about how they would translate what they observed into the context of their own classroom.

The fishbowl model also permitted the facilitator to ask TAs questions about what they were seeing—with an emphasis on questions related to instructional decisions. For example, the facilitator asked, “Why might you start with this question?” or “How do you think your students will answer that question?” We know that many TAs experience anxiety about teaching concepts for which they have little or fragmented understanding (Muzaka, 2009). In training contexts, TAs may be reluctant to ask content-related questions for fear of appearing less knowledgeable than their peers. Situating the discussion as one about *student* learning enabled facilitators to explore the origins of and reasons behind known misconceptions and barriers to student learning in hopes of improving the learning experience for *both* students and TAs.

Our revised model of PD extends beyond TAs simply doing the lab activity (as in TPD), which is necessary but insufficient as training for effective teaching. RPD creates time and space for TAs to engage with the lab material from a purely pedagogical perspective. Although we continued to allocate time during prep meetings for TAs to practice lab techniques and use equipment, we purposefully separated this from training time dedicated to teaching. We observed that following our implementation of the revised approach, TAs expressed less anxiety about teaching and were far more likely to ask questions about student learning and content understanding. We speculate that by situating TAs as observers in the mock

classroom, we may be reducing their cognitive burden. In TPD, TAs may be learning new material while at the same time thinking forward about how they are going to teach it in the following week. In RPD, TAs still learn new material, but we provide a visual example of what the desired instruction might look like so that TAs might mimic it, rather than create it anew.

**Dimensions 3 and 4. Roles of TA and Facilitator.** Because the RPD in this study is based on a constructivist approach, this required a fundamental shift in facilitator and TA roles. Facilitators were no longer the only source of knowledge, and TAs were no longer passive recipients of information. Functionally, this translated into TAs taking an active role in learning instructional strategies and changed the role of the meeting leader from that of an information source to a facilitator of TA development. The design of RPD was such that TAs were provided with numerous opportunities to actively participate in prep meetings. Each prep meeting began with an opportunity for TAs to talk with their teams and formally reflect on how the previous week's teaching experience went for them. TAs were provided guiding questions to catalyze group discussions (e.g., What went well? What surprised you? What did you learn about yourself as a teacher, about your students as learners?). TAs provided written feedback to facilitators that captured highlights or relevant concerns that emerged from the small-group discussions. Throughout each prep meeting in RPD, TAs had opportunities to participate in ways that drove the focus and duration of training. For example, during fishbowls, TAs could interrupt "class" to ask questions about the pedagogy or content and talk with their teams about expected student responses or preconceptions. Experienced TAs were asked to take on leadership roles by facilitating specific aspects of training sessions, teaching lessons in the fishbowl, and providing informal mentoring for novice TAs. In the design of RPD, time is intentionally allocated for TAs to observe, think, and discuss, so they can more fully engage with both the content and pedagogy of the lab. In TPD, although TAs are occasionally asked to present the lab overview, TA involvement in prep meeting activities is largely organized around learning procedures following presentations.

**Dimension 5. Goals of Preparation.** In backward design (Wiggins and McTighe, 1998), teaching and lesson design begins with clear, measurable goals. In RPD, the goals of weekly prep meetings were made explicit to TAs at the beginning of each meeting in the form of written objectives on a weekly handout. Ultimately, the overarching goal each week was to make sure that TAs felt prepared and equipped to teach the lab the following week. To achieve this, RPD prep meeting activities and discussions were focused on TA learning of both the content and the pedagogy.

While the goals of TA training in TPD may have been clear to facilitators, they were often implicit to TAs. Because TPD focused heavily on logistics, classroom management skills, and laboratory procedure, one might reasonably infer that these represent the goals of TPD. Although we expect that TPD facilitators cared about the quality of teaching in labs, it was not clear to what extent effective pedagogy was an objective of TPD.

**Dimension 6. Formal Reflection.** The RPD model asked TAs to explore their prior conceptions of teaching in the context of learning the dimensions of the new PD model. Reflection was a key component of RPD, because TAs bring well-developed conceptions about what makes an effective teacher, how to teach, the purpose of teaching, and how students learn that are challenged by RPD (Addy and Blanchard, 2010). In RPD, TAs worked in cooperative learning groups in which they developed a community of practice that provided regular, ongoing support (van Driel *et al.*, 2001) and a context for formal reflection (e.g., Richardson, 1996; Mezirow, 2003; Tanner, 2012) that often served as the basis for changes in the lab itself and in the design of the weekly prep meeting.

**Dimension 7. Form of Preparation.** TA PD often takes one of five forms: no training, meetings between TAs and the course instructor, presemester workshops (typically 1–2 d in length), a formal course or seminar that is separate from the TAs' teaching experience, or regular meetings that are concurrent with the course the TAs teach (Rushin *et al.*, 1997). Research to date is clear that longitudinal (meeting a number of different times) and continuous PD is more effective than shorter, one-time workshops in terms of promoting conceptual change and effective practice (Rogers, 1995; Fiszler, 2004). Therefore, we maintained the form of PD that had been previously established in TPD (regular, weekly meetings throughout the semester) rather than converting the RPD to a course-based approach or stand-alone workshop. In this way, TPD and RPD employ practices supported by available literature and do not differ in form.

## ANALYZING OUTCOMES OF TA PD

Although little information is available about evaluating TA PD, there are many different metrics and instruments for evaluating PD more generally (see Boulmetis and Dutwin, 2000). We selected Kirkpatrick's Four Level Evaluation Framework (Kirkpatrick, 1994) because of its multifaceted, data-driven approach to evaluation. Kirkpatrick's framework is the most widely used evaluation model in the business sector and has been used to evaluate faculty PD in higher education as well (e.g., Thackwray, 1997; Steinert *et al.*, 2006). Kirkpatrick's four levels include: 1) reaction to the program, 2) learning occurring as a result of the program, 3) application of content of the program, and 4) impact of the program on outcomes. For the purposes of evaluating TA PD, we do not address level 4 in this study, because students were concurrently enrolled in lecture sections that varied among instructors; thus, isolating student learning relative to TA PD was challenging. Kirkpatrick's framework provided guidance about the types of data and instruments selected to address each of these levels (Table 4).

For the purposes of our study, we applied this three-level evaluation framework before (TDP), during (RPD1), and after (RPD2) implementing RPD. Through this, we are able to contrast outcomes resulting from TPD and RPD. For each of the levels we include in our study, we discuss our approaches and rationale, as well as the nature of evidence we used for each PD program. It is important to note that our adaptation of Kirkpatrick's framework is from the perspective of designers of RPD, not TPD. Although we apply the framework to

**Table 4.** Evaluation framework for TA PD<sup>a</sup>

Level	Kirkpatrick framework	TA PD context	Evaluation data
1	Reaction to the program	How did TAs react to PD? How did they perceive the effectiveness of their training?	Written responses to survey questions
2	Learning as a result of the program	What did TAs learn about teaching this PD? What learning did TAs expect of their students?	Cognitive level of TA-designed classroom artifacts (e.g., assessment items, learning objectives). Alignment of goals with assessments.
3	Application of the content of the program	How did TAs apply their PD to the context of their teaching?	Video analyses of TAs' classroom practice
4	Impact of the program on outcomes	Impact of PD on student learning outcomes	Not collected

<sup>a</sup>Kirkpatrick's Four Levels Evaluation framework was adapted for the context of the study on TA PD (Kirkpatrick, 1994). Each level is described, and associated data collected for each level are described.

TPD, we do so for the purpose of characterizing TA outcomes resulting from alternative approaches to PD, not evaluating how well TPD achieved TPD goals.

### **Level 1. What Was the Response of the TAs to Their PD?**

At the end of each semester of teaching, TAs completed surveys (see Supplemental Material) that provided data about their perceptions of PD. In total, 36 surveys were collected from TAs across semesters of the study (Table 1). Analyses focused on three agree/disagree statements from the surveys: 1) "Prep meetings prepared me to teach next week's lab exercise." 2) "Prep meetings increased my confidence as a teacher." 3) "Prep meetings improved my teaching skills." Results from these three questions were compared among TPD, RPD1, and RPD2 using a Kruskal-Wallis test, and significant differences were evaluated using a pairwise Mann-Whitney test (i.e., Wilcoxon rank-sum; R Core Development Team, 2008).

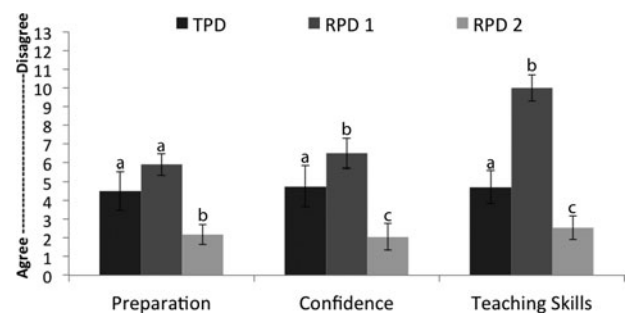
Our analyses indicated that TAs perceived TPD to be somewhat helpful in terms of their preparation to teach, confidence, and teaching skills (Figure 1). Results were not consistent between RPD1 and RPD2. TAs in RPD2 indicated significant agreement that RPD2 increased their confidence, preparation, and teaching skills (Kruskal-Wallis test:  $p < 0.05$ ), however, TAs in RPD1 did not hold the same views (Figure 1). TAs in RPD1 reported the greatest disagreement that RPD1 helped with their preparation, confidence, or teaching skills (Figure 1).

### **Level 2. What Did TAs Learn about Teaching in Their PD? What Learning Did TAs Expect of Their Students?**

In TPD, TAs were required to create their own assessments, including a midterm and final exam, as well as various assignments (e.g., quizzes, homework) throughout the semester. Facilitators and experienced TAs provided tips about assessment creation, including but not limited to: how to write a multiple-choice question, how to make questions challenging, how to write and set up a lab practical (it was expected that TAs offer 1 lab practical), and how to write assessments to reduce cheating. TAs were afforded opportunities to have

their assessments reviewed by peers for clarity, difficulty, and fairness. Collectively across TPD, TAs created 78 assessments, including exams (54%), quizzes (31%), in-class assignments (13%), and homework (2%).

One of the goals of the lab restructure was to reduce variation among sections by establishing common assessments that would be used by all TAs. In RPD, TAs were trained in backward design and learned about the necessity of aligning learning objectives and assessments. In addition, TAs learned about Bloom's taxonomy (Bloom and Krathwohl, 1956) as a way to compare assessments and objectives in terms of what they asked students to know and do. Collaboratively, TAs created 16 assessments, composed primarily of pre- and postlab assignments. Prelabs asked students to identify the purpose of the lab, summarize the methods through a written or visual representation, and state any questions they had about the lab content or procedures. Postlabs asked students to summarize results, represent data as graphs or tables, and write conclusions and reflections based on what they had learned.



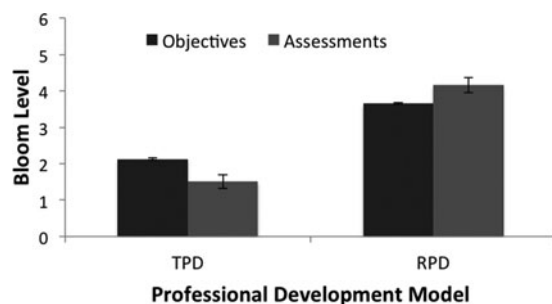
**Figure 1.** TA reactions to PD models. Specific TA responses about how well they thought their PD: 1) prepared them to teach, 2) increased their confidence, and 3) improved their teaching skills during three contrasting semesters of PD. y-axis represents a continuum of responses from agree to disagree; 0 represents 100% agreement with the statement, and 13 represents 0% agreement. TPD, traditional professional development; RPD1, reformed professional development (1); RPD2, reformed professional development (2). Bars represent means of TA responses (TPD:  $n = 12$ ; RPD1:  $n = 15$ ; RPD2:  $n = 11$ ) along the continuum with SE. Letters represent significant differences based on Kruskal-Wallis tests, where  $p < 0.05$  within each question (i.e., set of three bars). Identical letters (e.g., "a," "a") are not statistically different from one another.

Both TPD and RPD required TAs to create assessments to measure students' achievement on learning objectives for the lab. In TPD, learning objectives were predefined in the lab manual as a part of each lab activity. In RPD, TAs collaboratively developed learning objectives that reflected the new emphasis on inquiry.

We evaluated TA learning (level 2) as alignment between the cognitive level targeted by learning objectives and the assessments created by TAs within each type of PD. Two trained independent raters blindly rated the cognitive level of each objective and assessment item using Bloom's *Taxonomy of Educational Objectives* (Bloom and Krathwohl, 1956). Bloom's taxonomy describes a six-level hierarchy of cognitive skills, including: 1) knowledge, 2) comprehension, 3) application, 4) analysis, 5) synthesis, and 6) evaluation. Levels progress from less to more cognitively complex but are unrelated to difficulty (Wyse and Wyse, unpublished data). Raters followed the methods used by Momsen *et al.* (2010). The two raters exhibited an intraclass correlation (ICC) of 0.78, indicating that the ratings assigned by the two raters possessed high inter-rater reliability. Bloom ratings for each learning objective and assessment item were averaged to obtain a single rating for each objective and assessment.

We compared Bloom's levels of objectives and assessments within TPD and RPD using nonparametric ANOVA tests (e.g., Kruskal-Wallis), followed with nonparametric pairwise comparisons using Mann-Whitney (i.e., Wilcoxon rank-sum test) with a continuity correction to determine alignment (i.e., did TAs assess students at cognitive levels that matched the stated learning objectives?). All statistical analyses were performed using R (R Core Development Team, 2008).

Assessment items designed by TAs in TPD ( $n = 922$  items) primarily asked students to demonstrate knowledge of concepts (mean Bloom level =  $1.51 \pm 0.02$ ), with a few application-level questions found on homework and in-class assignments (Figure 2). Quiz and exam (midterm and final) items had mean Bloom levels below the average of 1.51 and



**Figure 2.** Learning objectives and assessments. TPD had 11 labs with stated learning objectives. TAs in TPD received limited formal instruction about assessment development and created their own assessments for the course. RPD had nine labs, and TAs created learning objectives across all levels of Bloom's taxonomy following training in objective writing and assessment design. RPD TAs also created assessments. TPD TAs created assessments ( $n = 922$  items) at significantly lower cognitive levels than their objectives ( $n = 68$  objectives) targeted ( $p < 0.001$ ). RPD TAs had no difference between the Bloom's level of assessments ( $n = 25$  items) and objectives ( $n = 41$  objectives;  $p = 0.102$ ; Wilcoxon rank-sum test), although the trend indicates that these assessment items were at higher Bloom's levels than objectives.

were significantly different from Bloom levels for in-class and homework assignments (Wilcoxon rank-sum test:  $p < 0.001$ ). Because quiz and exam items ( $n = 780$ ) constitute more than 84% of all the assessment items, there was low alignment between their laboratory objectives ( $2.13 \pm 0.04$ ) and their assessment of student learning ( $1.51 \pm 0.02$ ) (Wilcoxon rank-sum test:  $p < 0.001$ ). Objectives targeted comprehension and assessments, especially high-stakes assessments, evaluated student knowledge.

TAs in RPD demonstrated their ability to design assessments through the creation of prelab ( $n = 8$ /semester) and postlab ( $n = 8$ /semester) assignments. During RPD, TAs assessed student learning ( $4.1 \pm 0.21$ ) at the same cognitive level expected based on learning objectives ( $3.65 \pm 0.19$ ; Wilcoxon rank-sum test:  $p = 0.1020$ ). Most objectives for RPD were in the application, analysis, and synthesis levels of Bloom's taxonomy (Figure 2).

### Level 3. How Did TAs Apply Their PD to the Context of Their Teaching?

The nature of pedagogical training TAs received differed between TPD and RPD (Table 3). In RPD, learner-centered teaching was explicitly modeled for TAs, particularly through the facilitators' use of the fishbowl design. In TPD, pedagogical training was more implicit and more closely resembled the type of instruction typical of lecture-based college classrooms. We measured TAs' application of their training (level 3) by directly observing their classroom practice. TA teaching practice is most commonly measured through the use of self-reported surveys or student evaluations (see Gardner and Jones, 2011). However, Ebert-May *et al.* (2011) reported substantive discrepancies between faculty members self-reported perceptions about their teaching practice and data collected through independent observation. We therefore chose to observe and analyze videotapes of TA classroom practice for 29 (of 38) TAs who volunteered to participate in this study (Table 1). Each TA was videotaped twice during a semester (Table 1), with the exception of one TA who was taped only once. Dates selected for video recording were based on two criteria: 1) time during the semester (e.g., early vs. late) and 2) the type of laboratory (e.g., more prescriptive vs. more inquiry focused); see Table 5. It is

**Table 5.** Videotape selection criteria for TPD and RPD<sup>a</sup>

PD	Time	Method	Content
TPD	1 (early)	1 (prescriptive)	1 (predator-prey dynamics)
	2 (late)	2 (inquiry)	2 (gel electrophoresis)
RPD	1 (early)	1 (inquiry)	1 (cellular reproduction)
	2 (late)	2 (prescriptive)	2 (animal diversity)

<sup>a</sup>Each TA in the study was videotaped twice during the semester(s) they volunteered to participate. All TAs were videotaped for the same lab within each treatment (e.g., RPD). Videotapes of TAs were selected based on time during the semester (early vs. late), pedagogical approach (more prescriptive vs. inquiry), and content (differing scales). It is important to note that pedagogical approach is embedded within the treatment. For example, a prescriptive lab during TPD was one that was a step-by-step lab with a predetermined outcome, whereas a prescriptive lab during RPD was a highly guided inquiry experience.

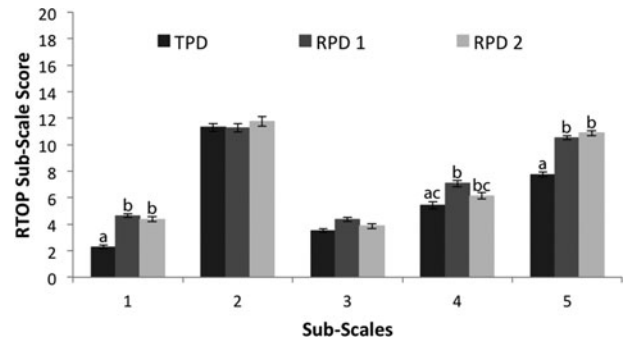
important to note that the labels “prescriptive” versus “inquiry focused” are contextualized during each type of PD. That is, during the TPD model, prescriptive labs were more procedural and inquiry labs were question-and-answer driven. During the RPD model, prescriptive and inquiry labs differed in the degree of guidance and direction provided during the inquiry experience. Labs videotaped within a treatment (e.g., all those taped during RPD) were the same for each TA in the study.

Videotapes were evaluated using the Reformed Teaching Observation Protocol (RTOP; Sawada *et al.*, 2000), a valid and reliable instrument for measuring the degree of learner-centered instruction based on observed classroom practices (Sawada, 1999; Sawada *et al.*, 2002). RTOP scores range from 0 to 100, with higher numbers indicating more learner-centered instruction. The RTOP composite score is based on five subscales: 1) lesson design and implementation, 2) propositional knowledge, 3) procedural knowledge, 4) communicative interactions, and 5) student–teacher relationships. RTOP scores are interpreted through five levels, with a score greater than 45 representing learner-centered instruction (Ebert-May *et al.*, 2011). Two independent raters were trained and calibrated using RTOP training videos, and each conducted a blind review of 57 ( $29 \times 2 = 58$ ;  $58 - 1 = 57$ ) videotapes. The ICC for the two raters was 0.700, indicating high reliability.

We used a linear model to determine what factors from our study might be impacting the degree of learner-centered instruction taking place in these TAs’ classrooms. The linear model included PD model (i.e., TPD, RPD1, RPD2), videotape number (i.e., tape 1, tape 2) and our two significant covariates (i.e., gender and semesters of teaching experience outside the current institution). The linear model (all assumptions held) included testing for interactions. We used a post hoc Tukey’s honest significant difference (HSD) test to identify which semesters and videotape numbers differed from one another. The five subscales were also tested with a linear model (all assumptions held) and assessed with post hoc pairwise comparisons using Tukey’s HSD test. All analyses were performed in R (R Core Development Team, 2008).

Type of PD (i.e., TPD, RPD1 or RPD2) was the only significant predictor of RTOP score (ANOVA:  $F_2 = 6.08$ ,  $p < 0.005$ ). Other covariates (i.e., gender and semesters of teaching experience outside of the current institution) were nonsignificant predictors, as was the time of videotaping (i.e., tape 1 vs. tape 2).

The composite scores for TAs in TPD correspond to the RTOP interpretation (Ebert-May *et al.*, 2011) of “straight lecture” (0–30; Figure 3). Within RTOP subscales, TPD TAs scored the highest in content knowledge (subscale 2) and lowest in lesson design (subscale 1; Figure 3). In contrast, RPD TAs fell into the second cut-score category (31–45) “lecture with minor student participation” (Figure 3). Despite falling short of designation as “learner-centered” (scores greater than 45), TAs in RPD scored higher (ANOVA:  $F_2 = 6.18$ ,  $p = 0.005$ ) on the RTOP overall compared with TPD. Differences were also evident on RTOP subscales, with RPD TAs scoring higher on instructional design and implementation (subscale 1; ANOVA:  $F_2 = 8.92$ ,  $p = 0.0005$ ), classroom culture: communicative interactions (subscale 4; ANOVA:  $F_2 = 3.99$ ,  $p = 0.024$ ), and classroom culture: student–teacher relationships (subscale 5; ANOVA:  $F_2 = 9.88$ ,  $p = 0.0005$ ).



**Figure 3.** Mean RTOP scores for TAs by semester of PD. The three bars (with SEs) for each subscale represent the average RTOP score for all TAs on that subscale within each particular semester. TPD:  $n = 9$ ; RPD1:  $n = 11$ ; RPD2:  $n = 9$ . TPD averaged  $29.4 \pm 5.8$ , whereas RPD1 and RPD2 averaged  $37.1 \pm 6.1$  and  $36.5 \pm 8.1$ , respectively, on the overall RTOP scale. Subscales: 1, instructional design and implementation; 2, content knowledge; 3, procedural knowledge; 4, communicative interactions; and 5, student–teacher relationships. Letters above bars represent significant ( $p < 0.05$ ) differences among semesters (as determined by ANOVA and post hoc pairwise comparisons using Tukey’s HSD) within each subscale; identical letters indicate nonsignificant differences.

## DISCUSSION

Graduate TAs teach a large number of undergraduates in STEM, yet receive relatively limited pedagogical PD through their TA training (Gardner and Jones, 2011). Further, TA PD is most often evaluated solely through the use of self-reported data. Here, we adapted and applied Kirkpatrick’s evaluation framework for use in TA PD, because it offers a multi-level approach to evaluation that combines self-reported data with competency analysis (Kirkpatrick, 1994). By applying common evaluation metrics to alternative models of PD, we gained insights about strengths and weaknesses associated with each.

Our evaluation of TPD revealed that TAs did not feel prepared to teach (Figure 1). Their assessments (i.e., quizzes, exams) matched the instruction they received in that they were not taught about alignment and their assessment items did not exhibit alignment with learning outcomes (Figure 2). Finally, RTOP analyses of TAs in TPD revealed that despite their strong content knowledge, TAs were less effective in deploying learner-centered practices in their classrooms.

Our evaluation of RPD outcomes are less straightforward. Our analyses show that, with RPD, TAs designed assessments aligned with objectives and engaged in more learner-centered practices in their classrooms (Figures 2 and 3). These desirable outcomes, however, were inconsistent with TAs’ self-reported perceptions about PD efficacy. In RPD1, TAs felt *less* prepared and *less* confident to teach, and they also felt that their teaching skills *decreased* (Figure 1), yet these very same TAs developed better-aligned assessments and were observed teaching in more learner-centered ways (Figure 3). It is important to note that had we applied conventional evaluation tools (surveys) following RPD1, we would likely have abandoned our approach given the negative feedback regarding TAs’ perception about RPD in improving their teaching skills, confidence, and preparation. However, data from TAs’ assessments and teaching videos suggested otherwise and



indicated improvement in TAs' efficacy in learner-centered teaching methods. In RPD2, TAs reported more positive perceptions about PD (Figure 1) and maintained a degree of learner-centered instruction similar to RPD1 (Figure 3). Our use of a multilevel evaluation framework enabled us to discern patterns we might have missed had we evaluated PD using only conventional, survey-based methods.

### *Nature and Quality of PD Differences Revealed*

We recognize the potential for substantive differences in the experiences among TA cohorts owing to the reform process itself and their roles in it. Our goal of making labs more inquiry based was accomplished, in part, through the contributions of TAs in RPD1 who collaborated with us to design and revise lab materials. In our view, the RPD1 TAs' participation in the transition toward reform served to 1) authentically engage TAs as owners of and collaborators in the reform process itself, and 2) provide a unique form of PD that reflects the experience of faculty members as they design and test learning materials in the context of their own classrooms. It is possible that implementing one's own teaching materials provided RPD1 TAs quite different insights and perspectives about teaching and learning compared with TPD and RPD2 TAs who were implementing materials developed by others.

With respect to differences between TPD and RPD, the fundamental shift toward an inquiry focus meant that TAs trained under TPD were aspiring to quite different goals for their teaching compared with TAs trained under RPD. In our experience, the transition toward inquiry was not simply a matter of changing the materials or approaches to teaching labs, but incorporated a number of cultural and attitudinal factors accompanying the change process. For example, the very nature of the role of TAs—both in prep meetings and in their classrooms—differed substantively in TPD and RPD. As such, we cannot directly link PD as a causal explanation for the differences we observed in TA performance. Variation among TA cohorts in terms of their perceptions and practices is likely the result of multiple interacting factors related to changes in curriculum and sociocultural norms and expectations, as well as PD. Below, we hypothesize about potential reasons that could account for some of the patterns and variability we observed.

**Level 1.** When compared with TAs in TPD, RPD1 TAs did not feel confident or prepared to teach, nor did they believe they made improvements in their teaching skills. There are many possible reasons for this. First, for most TAs, lecture-based instruction is most familiar, as they have had a long apprenticeship of observation (Lortie 1975) in this teaching approach. It may not be surprising that TAs felt ill prepared or uncertain about their classroom practices when confronted with a novel way of teaching that most had not seen or experienced as learners. Lewis *et al.* (1999) found that it was not uncommon for teachers to report feeling unprepared to teach when in the process of learning new pedagogies and approaches to teaching; this could be the case for the TAs as well. Second, TAs in RPD1 were developing revised course materials for a learner-centered lab experience—a task that was new for nearly every TA teaching that semester. It is possible that the process was overwhelming and contributed to TAs' feelings of frustration or lack of progress in teaching

skills. A third explanation considers TAs' beliefs. The RPD approach may have conflicted with some TAs' beliefs about effective teaching and student learning (Richardson, 1996; Wyse, 2010). TAs were experiencing learner-centered instruction during their PD that may not have aligned with their conception of "teaching." For example, if TAs believed that the role of a teacher is to provide information (i.e., answer, rather than ask questions), they may have experienced discomfort in an environment in which the facilitator provided training in the form of questioning, inquiry, and reflection. Therefore, the fact that TAs felt that their teaching skills did not improve may make sense in the context of their beliefs.

In RPD2, however, TAs' perceptions shifted, and many reported that PD was effective in *improving their teaching skills and helping them feel prepared and confident*. So why was there improvement in TA perceptions in RPD2? First, it is possible that the facilitators were more effective in implementing the fishbowl and other aspects of PD during the second iteration, which could have contributed to TA confidence and preparedness (Figure 1). Second, four TAs (Table 1) from RPD1 taught again in RPD2. These TAs brought experience and familiarity with the course that could have positively influenced the dynamic of the weekly prep meetings and the opinions of less-experienced TAs, particularly in the context of small-group discussions. Although removing these four TAs from the RPD2 data set did not change statistical outcomes for any questions, the facilitators observed that repeat TAs appeared "less stressed" and were more cooperative—particularly with respect to helping their peers learn the newer materials and approaches. In this way, repeat TAs brought expertise and energy to RPD2 that may have improved perceptions and confidence for all TAs.

**Levels 2 and 3.** Despite inconsistencies in self-reported perceptions from RPD1 to RPD2, both cohorts of RPD were associated with improvements in teaching-related skills and competencies. With respect to assessment design, TPD TAs created 84% of their assessment items for high-stakes assessments (e.g., quizzes and midterm and final exams). These items were targeted at the knowledge level (Bloom's level 1), whereas their learning objectives targeted a comprehension level (Bloom's level 2); very small SEs indicate little variation among items in the levels targeted. Although some may argue that Bloom's level 1 (knowledge) and Bloom's level 2 (comprehension) are both low-order cognitive processes (Crowe *et al.*, 2008), the cognitive complexity of what students are asked to do during recall compared with comprehension are distinct (Bloom and Krathwohl, 1956). That is, to be asked to know (i.e., recall) rather than comprehend (i.e., understand) requires different cognitive skills.

For TAs in RPD, their high-stakes assessments were better matched with their learning objectives. While both objectives and assessments created by TAs in RPD are higher order (levels 3 and 4), assessment items measured a wider range of cognitive processing levels (as evidenced by the larger SE in Figure 2), leading to better alignment across all assessment items. Further, research suggests that higher-order cognitive processing skills may indeed not be as hierarchical as lower-order skills (Crowe *et al.*, 2008). This distinction is important: lower-order cognitive skills build upon one another, such that asking students to comprehend requires that they also know, so if TAs are asking students to recall knowledge, but the

expectation is that they comprehend the concepts, there is a degree of misalignment between what is expected of students and what is assessed. In RPD, we did not see this misalignment. TAs created assessments that asked students to demonstrate their learning using similar cognitive processing skills to those expected of them as stated in learning objectives. This alignment may be due, in part, to the explicit training on backward design in the presemester orientation and feedback on their assessment development.

In their classrooms, TAs in RPD demonstrated greater progress toward learner-centered teaching, with many TAs moving from “straight lecture” to “lecture with student engagement” (Figure 3). The use of RTOP subscales enabled us to determine the nature of change. Although RPD TAs made measurable changes in lesson design and implementation (subscale 1), communicative interactions (subscale 4) and student–teacher relationship (subscale 5) showed the most improvement (Figure 3). This was evident in RPD TAs’ decreased lecture time and higher student involvement compared with TPD TAs. Increasing inquiry was a primary objective of reform, therefore training TAs to ask questions and engage student interaction were priorities of RPD. We predict that our lesson plans (Supplemental Material) and the fishbowl approach in PD meetings provided TAs with resources and opportunities to see what communicative interactions and student–teacher relationships might look like in a reformed setting. It was not surprising that subscales 2 (content knowledge) and 3 (procedural knowledge) did not differ between TPD and RPD, because TAs do have knowledge of biological content.

### Changing Teaching Practices of TAs

It is important to note that while RPD TAs made positive gains based on RTOP scores, their gains were small. Our results indicate that even with explicit instruction, modeling, and structural changes to the laboratory, changing teaching practices (even for beginning instructors) to become more learner-centered is challenging. Yelon *et al.* (2004) claim that change requires three essentials: 1) credibility, 2) practicality, and 3) need. It is unusual that a single TA would experience all three of these components in a single-semester TA experience. However, these three ingredients could have played a role in the positive responses to RPD2. For example, TAs may have experienced credibility as they learned about the rationale and evidence base for learner-centered instructional practices (see Prince, 2004, as an example). Once this rationale was established, TAs needed to experience the practical success of such methods. This may mean experiencing it for themselves or hearing of success from trusted peers or colleagues. For example, in RPD2, TAs may have heard other TAs (including the four TAs who taught during RPD1) talking about the learning of their Bio1 students. Therefore, it is possible that some TAs could have experienced the practicality of reformed teaching approaches and thus may have been more motivated or encouraged to try some of these approaches themselves.

Finally, TAs must experience a need for a different way of teaching. Without the cognition of there being a need for change, little change is possible (Yelon *et al.*, 2004). TAs are young instructors with a wealth of personal classroom experiences that were successful (Lortie, 1975). Their personal belief

structure may support the idea that a transmission mode of instruction works, and as such, there is no need to change to learner-centered ways. Many TAs have not yet had numerous teaching experiences that put them face to face with the reality that what worked for them as learners may not work for all students. Here, we assert that TA PD can offer some but not all of the solution. PD can help build the case for the need (Yelon *et al.*, 2004) by providing pedagogical instruction concurrent with reflection on the teaching and learning experiences of others.

### USING MULTIDIMENSIONAL DATA

Self-reported survey data are the standard by which most programs evaluate PD, including TA training (see Table 1 in Gardner and Jones, 2011). Our results confirm the need for integrating multiple sources of data when evaluating outcomes from TA PD and identifying the potential for bias and errant conclusions when relying on more narrow or unidimensional evaluation protocols. This study demonstrates the use of different kinds of data to inform the design and evaluation of TA PD.

We conclude that self-reported data alone do not predict what TAs actually do in their classrooms, nor do they reflect TAs’ knowledge about learner-centered instruction. For example, Ebert-May *et al.* (2011) found that faculty tend to overestimate the degree of learner-centeredness they implement in their classroom. For TAs, self-efficacy may explain the misalignment between their self-reported data and their classroom practice. As TAs experience the difficulties of teaching, their self-efficacy tends to decrease; therefore, it is possible that TAs underestimated their competencies in RPD1 due to their confrontation with the difficult realities of teaching (Prieto and Altmaier, 1994).

Future investigations into evaluations of TA PD need to include more than self-reported data. While there are many different metrics or instruments for evaluating PD (see Boulmetis and Dutwin, 2000), we found Kirkpatrick’s Four Levels Evaluation framework to be particularly conducive for providing robust and diverse data about TA outcomes (Kirkpatrick, 1994). Kirkpatrick’s framework serves to provide program evaluators with guidance about the types of data and instruments necessary for evaluating multiple dimensions of PD. By including multiple data types, we have greater potential for understanding how various types of PD impact the learning and practice of TAs.

### RE-ENVISIONING TA PD

Nationally, recognition of the need to implement TA PD that better prepares TAs to teach is gaining momentum. While some programs provide innovative ways to prepare TAs (see Gardner and Jones, 2011, for a review), most are either stand-alone courses or seminars that are decontextualized from a teaching experience, or so highly context-specific that they are not transferable to other programs or disciplines. In addition, many of these programs use limited methods for evaluating effectiveness. Concurrently, there is considerable attention to improving introductory college-level science courses (NRC, 2003; AAAS, 2011), many of which have laboratory sections

taught by graduate TAs. If programs are serious about efforts to improve the quality of introductory college science courses, they must pay attention not only to what is happening in large class meetings (i.e., lectures) but also to the quality of TA training.

We predict that the framework for RPD outlined here is broadly applicable across a variety of institutions because of the following dimensions. Our RPD framework: 1) views learning from a constructivist perspective (von Glasersfeld, 1989), recognizing that TAs build and construct knowledge of biology through interaction and involvement with the material and one another; 2) acknowledges that TAs play dual roles as learners and as teachers, and the design of the PD reflects this consideration of TA roles; 3) places the leader(s) in the role of guide, allowing TAs to construct their understanding through interactions with leaders and peers; and 4) engages the TAs as collaborators in the process of reform, empowering them to provide input and feedback along the way that fundamentally alters the structure of both the reformed course and the PD program (Table 3). We believe this framework is transferable to other institutions and can be implemented across diverse TA programs in which regular prep meetings can provide a context for sustained and iterative PD training activities (see Supplemental Material).

We strongly advocate the fishbowl approach for TA training and believe it can easily be implemented into an existing TA prep meeting (see Supplemental Material) by choosing a lead TA for each week and using a small group of students (e.g., paid or volunteer undergraduate learning assistants) to demonstrate lab instruction as TAs observe. A facilitator can ask TAs questions for procedural and metacognitive feedback during or after the fishbowl. Additionally, TAs can be organized into groups during prep meetings, with groups enabling support and discussion among peers who are all in the process of learning to teach.

Finally, we urge leaders of TA PD programs to adopt rigorous and multidimensional methods for evaluating TA PD outcomes. Currently, there is no standard evaluation protocol for TA PD; as such, we adapted Kirkpatrick's evaluation framework (Kirkpatrick, 1994) to the specific context of an introductory biology lab course. Having multiple data sources enabled us to independently examine the impacts of TA PD on TAs' perceptions and impacts on teaching practices. Again, we feel that a multilevel evaluation approach is highly transferable across programs and lends itself to customization to suit diverse needs.

Because graduate TAs teach a significant portion of our undergraduates in STEM and many will go on to teach subsequent generations of scientists and citizens, providing and evaluating the quality of TA PD needs to become part of our broader effort to improve the preparation of graduate students for the changing face of higher education. Our model for reformed TA PD is grounded in theory about how people learn and uses data from multiple sources to test its effectiveness. Based on our experience designing, implementing, and evaluating reformed PD, teaching TAs in the same manner we hope they will teach was a step toward building a model of PD that helps TAs gain skills in learner-centered instruction. The negative connotation often associated with the phrase "teach as you are taught" (Lortie, 1975) assumed a positive one—in hopes that TAs *will* teach as they were taught to teach.

## ACKNOWLEDGMENTS

First and foremost, we thank the graduate TAs who participated in and contributed to this study. We also thank the director, staff, and undergraduate learning and research assistants associated with the Bio1 program for supporting this research. We are thankful for constructive comments on drafts from Jennifer Momsen, Terry Derting, and Paula Soneral. This material has IRB approval (MSU, IRB#X07-1196) and is based upon work supported by the National Science Foundation under grant number DUE-0736928. Funding was also provided by the Future Academic Scholars in Teaching program at Michigan State University.

## REFERENCES

- Abbott RD, Wulff DH, Szego CK (1989). Review of research on TA training. In: *Preparing the Professoriate of Tomorrow to Teach: Selected Readings in TA Training*, ed. JD Nyquist, RD Abbot, DH Wulff, and J Sprague, Dubuque, IA: Kendall/Hunt, 111–124.
- Addy TA, 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 program. *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*, Washington, DC.
- Ball DL, Bass H (2000). Interweaving content and pedagogy in teaching and learning to teach: knowing and using mathematics. In: *Multiple Perspectives on Mathematics Teaching and Learning*, ed. J Boaler, Westport, CT: Ablex, 83–104.
- Baumgartner E (2007). A professional development teaching course for science graduate students. *J Coll Sci Teach* 36(6), 16–21.
- Baviskar S, Beardsley P (2006). *Survey Report on Graduate Teaching Assistants at ISU*, Pocatello: Idaho State University.
- Bloom BS, Krathwohl DR (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*, New York: Longmans, Green.
- Bond-Robinson J, Rodrigues RAB (2006). Catalyzing graduate teaching assistants' laboratory teaching through design research. *J Chem Educ* 83, 313–323.
- Boulmetis J, Dutwin P (2000). *The ABCs of Evaluation: Timeless Techniques for Program and Project Managers*, San Francisco: Jossey-Bass.
- 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.
- 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 development programs. *BioScience* 6, 550–558.
- Fiszer EP (2004). *How Teachers Learn Best: An Ongoing Professional Development Model*, Lanham, MD: Scarecrow Education.
- Gardner GE, Jones MG (2011). Pedagogical preparation of the science graduate teaching assistant: challenges and implications. *Sci Educ* 20, 31–41.
- Gormally C, Brickman P, Hallar B, Armstrong N (2011). Lessons learned about implementing an inquiry-based curriculum in a college biology laboratory classroom. *J Coll Sci Teach* 40(3), 45–51.
- Hammrich PL (2001). Preparing graduate teaching assistants to assist biology faculty. *J Sci Teach Educ* 12, 67–82.
- Hampton SE, Reiser RA (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classrooms. *Res High Educ* 45, 497–527.
- Handelsman J *et al.* (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. *J Res Sci Teach* 48, 952–984.
- Johnson DW, Johnson RT, Stanne MB (2000). *Cooperative Learning Methods: A Meta-analysis*, Minneapolis: University of Minnesota. [www.tablelearning.com/uploads/File/EXHIBIT-B.pdf](http://www.tablelearning.com/uploads/File/EXHIBIT-B.pdf) (accessed October 2009).
- Kirkpatrick DL (1994). *Evaluating Training Programs: The Four Levels*, San Francisco: Berrett-Koehler.
- Lewis L, Parsad B, Carey N, Bartfai N, Farris E, Smerdon B, Greene B (1999). *Teacher Quality: A Report on the Preparation and Qualifications of Public School Teachers*, Publication number 1999080, Alexandria, VA: National Center for Education Statistics.
- Lortie DC (1975). *Schoolteacher: A Sociological Study of Teaching*, Chicago: University of Chicago Press.
- Luft JA, Kurdziel JP, Roehrig GH, Turner J (2004). Growing a garden without water: graduate teaching assistants in introductory science laboratories at a doctoral/research university. *J Res Sci Teach* 41, 211–233.
- Marincovich M, Prostko J, Stout F (eds.) (1998). *The Professional Development of Graduate Teaching Assistants*, Bolton, MA: Anker.
- McKeachie WJ (1999). *Teaching Tips: Strategies, Research and Theory for College and University Teachers*, Boston, MA: Houghton Mifflin.
- McManus DA (2002). Developing a teaching assistant preparation program in the School of Oceanography, University of Washington. *J Geosci Educ* 50, 158–168.
- Mezirow J (2003). Transformative learning as discourse. *J Transform Educ* 1, 58–63.
- Michael J (2006). Where's the evidence that active learning works? *Adv Psychol Educ* 30, 159–167.
- Momsen JL, Long TM, Wyse SA, Ebert-May D (2010). Just the facts? Undergraduate biology courses focus on low-level cognitive skills. *CBE Life Sci Educ* 9, 435–440.
- Muzaka V (2009). The niche of graduate teaching assistants (GTAs): Perceptions and reflections. *Teach High Educ* 14, 1–12.
- National Research Council (2003). *Improving Undergraduate Education in Science, Mathematics, Engineering and Technology*, Washington, DC: National Academies Press.
- Prieto LR, Altmeyer EM (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. *Res High Educ* 35, 481–497.
- Prince M (2004). Does active learning work? A review of the research. *J Eng Educ* 93, 223–231.
- R Core Development Team (2008). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. [www.R-project.org](http://www.R-project.org) (accessed January 2008).
- Richardson V (1996). The role of attitudes and beliefs in learning to teach. In: *The Handbook of Research on Teacher Development*, 2nd ed., ed. J Sikula, New York: Macmillan, 102–119.
- Roehrig GH, Luft JA, Kurdziel JP, Turner JA (2003). Graduate teaching assistants and inquiry-based instruction: implications for graduate teaching assistant training. *J Chem Educ* 80, 1206–1210.
- Rogers EM (1995). *Diffusion of Innovation*, 4th ed., New York: Free Press.
- Rushin JW, DeSaix J, Lumsden A, Streubel DP, Summers G, Bernson C (1997). Graduate teaching assistant training—a basis for improvement of college biology teaching and faculty development? *Am Biol Teach* 59, 86–90.
- Sawada D (1999). *Psychometric Properties of RTOP*, Tempe: Arizona Collaboration for Excellence in the Preparation of Teachers.
- Sawada D, Pilburn M, Falconer K, Turley J, Benford R, Bloom I (2000). *Reformed Teaching Observation Protocol (RTOP)*, Report number IN00-1, Tempe, Arizona Collaborative for Excellence in the Preparation of Teachers.
- Sawada D, Pilburn M, Judson E, Turley K, Falconer K, Benford R, Bloom I (2002). Measuring reform practices in science and mathematics classrooms: the Reformed Teaching Observation Protocol. *Sch Sci Math* 102, 245–253.
- Shulman L (1986). Those who understand: knowledge growth in teaching. *Educ Res* 15(2), 4–14.
- Steinert YK, Mann, 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 8. *Med Tech* 28, 497–526.
- Sundberg MD, Armstrong JE, Eischusen EW (2005). A reappraisal of the status of introductory biology laboratory education in US colleges and universities. *Am Biol Teach* 67, 525–529.
- Tanner K (2012). Promoting student metacognition. *CBE Life Sci Educ* 11, 113–120.
- Tanner K, Allen D (2006). Approaches to biology teaching and learning: on integrating pedagogical training into the graduate experiences of future science faculty. *Cell Biol Educ* 5, 1–6.
- Taylor K, Marienau C, Fiddler M (2000). *Developing Adult Learners*, San Francisco, CA: Jossey-Bass.
- Thackwray B (1997). *Effective Evaluation of Training and Development in Higher Education*, London: Kogan Page.
- van Driel JH, Beijaard D, Verloop N (2001). Professional development and reform in science education: the role of teachers' practical knowledge. *J Res Sci Teach* 38, 137–158.
- von Glasersfeld E (1989). Cognition, construction of knowledge and teaching. *Syntheses* 80, 121–141.
- Vygotsky L (1978). *Mind in Society*, London: Harvard University Press.
- Wiggins G, McTighe J (1998). *Understanding by Design*, Alexandria, VA: Association for Supervision and Curriculum Development.
- Wyse SA (2010). *Breaking the mold: Preparing graduate teaching assistants to teach as they are taught to teach*, PhD Dissertation, East Lansing: Michigan State University.
- Yelon S, Sheppard L, Slight D, Ford JK (2004). Intention to transfer: how do autonomous professionals become motivated to use new ideas? *Perform Improv Q* 17, 82–103.