




# Classroom Observations Indicate the Positive Impacts of Discipline-Based Professional Development

Karen Viskupic<sup>1</sup>  · Katherine Ryker<sup>2</sup> · Rachel Teasdale<sup>3</sup> · Cathryn Manduca<sup>4</sup> · Ellen Iverson<sup>4</sup> · Dori Farthing<sup>5</sup> · Monica Z. Bruckner<sup>4</sup> · Rory McFadden<sup>6</sup>

Published online: 01 May 2019  
© The Author(s) 2019

## Abstract

We compared 236 geoscience instructors' histories of professional development (PD) participation with classroom observations using the Reformed Teaching Observation Protocol (RTOP) that describe undergraduate classes as Student-Centered (score  $\geq 50$ ), Transitional (score 31–49) or Teacher-Centered (score  $\leq 30$ ). Instructors who attended PD ( $n = 111$ ) have higher average RTOP scores (44.5 vs. 34.2) and are more frequently observed teaching Student-Centered classes (33% vs. 13%) than instructors with no PD ( $p < 0.001$ ). Instructors who attended PD that is topically-aligned with content taught during the classroom observation are likely to have RTOP scores that are higher by 13.5 points ( $p < 0.0001$ ), and are 5.6 times more likely to teach a Student-Centered class than instructors without topically-aligned PD. Comparable odds of teaching Student-Centered classes (5.8x) occur for instructors who attended two topical PD events but were observed teaching a different topic. Models suggest that instructors with at least 24 h of PD are significantly more likely to teach a Student-Centered class than instructors with fewer hours. Our results highlight the effectiveness of discipline-specific PD in impacting teaching practices, and the importance of attending more than one such PD event to aid transfer of learning.

**Keywords** Professional development · Reformed teaching observation protocol (RTOP) · Reformed teaching · Active learning

## Introduction

Given national mandates to improve undergraduate STEM education (National Research Council 1996; National Science Foundation 1996; Olson and Riordan 2012), research has focused on studying the use of student-centered teaching in the

---

✉ Karen Viskupic  
[karenaviskupic@boisestate.edu](mailto:karenaviskupic@boisestate.edu)

sciences to improve student learning and academic success (Freeman et al. 2014; Knight and Wood 2005; Michael 2006). Student-centered teaching practices that engage students in peer-to-peer interactions and emphasize higher-order thinking have been shown to result in improved learning over teacher-centered, lecture-focused instruction for a wide array of university courses, students, and STEM subjects in terms of reduced failure rates in university courses (e.g. Freeman et al. 2014), and improved student performance on course assessments (e.g. McConnell et al. 2003; Knight and Wood 2005). There is a clear link between student success and active learning (Freeman et al. 2007; Hake 1998; Handelsman et al. 2004; National Research Council 2012), and such work has helped inform STEM faculty about research-based teaching innovations (Dancy and Henderson 2010).

In their commentary on the future of faculty development in higher education, Austin and Sorcinelli (2013) underscore the ongoing priority of faculty development to help faculty gain teaching skills that support student learning, and argue that increasingly diverse student populations make faculty development critical to higher education. Many science faculty enter their positions with little pedagogical training (Handelsman et al. 2004; Rushin et al. 1997; Walczyk et al. 2007), and lack of access to such training is widely regarded as an important barrier to the use of student-centered teaching practices for university faculty (Brownell and Tanner 2012; Handelsman et al. 2004; Rushin et al. 1997). Results of a national survey of faculty developers in higher education indicate that faculty prioritized professional development (PD) programs that led to teaching changes that help student learning (Sorcinelli et al. 2006). There are several pathways available for faculty PD in the use of student-centered teaching practices including participation in workshops (Clark et al. 2002; Pfund et al. 2009), online webinars (e.g. Schrum et al. 2005), publications on the efficacy of specific teaching practices (e.g. Mulnix 2016), online curriculum resources (Manduca et al. 2017), personal mentoring (Czajka and McConnell 2016), and guidance from teaching and learning centers (e.g. McShannon and Hynes 2005).

Although PD that aims to train university faculty about teaching and learning can be beneficial to instructors (Bouwma-Gearhart 2012; Gibbs and Coffey 2016; Postareff et al. 2007), including when it is discipline-based (Clark et al. 2002; Manduca et al. 2017), one's participation in PD and knowledge of the benefits of student-centered teaching practices do not guarantee the transfer of PD content to the university classroom (Brownell and Tanner 2012; Henderson and Dancy 2007; National Research Council 2003). However, the availability of PD is a factor that academic disciplinary communities can impact, thereby improving our capacity to change how we teach through improved pedagogical content knowledge (PCK), which in turn impacts undergraduate student learning. It is therefore important to understand the most effective characteristics of higher education PD programs and how participation can help faculty transfer PCK to university level classroom practice (Derting et al. 2016; Ebert-May et al. 2015; Henderson et al. 2011; Manduca 2017).

## Research Questions

The goal of this study was to determine the impact of discipline-specific PD workshops on undergraduate teaching in the geosciences by answering the following questions:

1. Does participation in discipline-specific PD correlate with the use of student-centered teaching practices?
2. Is there a correlation between teaching practice and the amount of time that has passed since PD participation?
3. Is student-centered teaching more likely when PD and course content are aligned?
4. Is there a minimum number of PD events (or hours) in which an instructor should participate before student-centered teaching occurs?

To explore these questions, we observed undergraduate geoscience classes using the Reformed Teaching Observation Protocol (RTOP; Sawada et al. 2002) and linked results to compiled histories of PD participation for observed instructors. PD histories were derived from the archive of activities hosted by the Science Education Resource Center (SERC) including On the Cutting Edge (OCE), Building Strong Departments, the National Association of Geoscience Teachers (NAGT), InTeGrate, and Supporting and Advancing Geoscience Education at Two-Year Colleges (SAGE).

The PD events considered in this study have common features such as engaging participants in active learning, incorporating participants' prior experiences and knowledge, and including volunteer participants with common disciplinary interests (Manduca et al. 2017). The PD events differ from each other in terms of length and disciplinary content focus, allowing us to test hypotheses related to these variables. While some PD events focused on specific disciplinary topics (e.g. mineralogy), and others did not (e.g. preparing students to enter the workforce), all events included content and examples from the geosciences. Such discipline-based PD should promote changes to observable teaching practice through the development and transfer of instructors' pedagogical content knowledge (PCK).

In the theoretical framework below, we review empirically identified critical features of PD and connect them to theoretical constructs of PCK, adaptive expertise, transfer, and the importance of time in learning. These connections are important for understanding why, and under what conditions, disciplinary PD may be effective in promoting changes to teaching practice.

## Theoretical Framework

The theory of conceptual change describes a process by which learners integrate new conceptions with their existing experiences; if the combination is viable (rational), it can be accommodated by the learner (Posner et al. 1982; Strike and Posner 1992). Faculty take on the role of learner when participating in PD where the purpose is promoting changes in teaching practice by helping them shift from a teacher-centered approach to a new model of student-centered instruction to support student learning. Viewed through the lens of conceptual change theory, instructors who modify their teaching practice to become more student-centered must be dissatisfied with some aspect of their practices (or the outcomes of those practices), they must understand what student-centered practices are, the use of student-centered practices must seem plausible in their instructional and disciplinary context, and the idea of student-centered teaching must open new ways of thinking that seem productive.

Participation in PD events that help instructors meet the conditions for conceptual change should result in changes to instructors' conceptions of teaching, and therefore result in changes to teaching practices. PD events that are effective in promoting instructional changes have been identified through empirical studies, and the common features of these PD events were described by Desimone (2009) as a core set of PD features that are critical to increasing instructors' knowledge and skills and improving their teaching practice. Given the wide variety of PD programs, and methods of their evaluation, it is difficult to directly correlate specific characteristics of PD with improvements in student learning (e.g. Gersten et al. 2014; Yoon et al. 2007), and not all PD programs that include only some of these critical features have resulted in measurable changes to teaching practice (e.g. Garet et al. 2011). While not all researchers agree on a set of best practices for PD, we use Desimone's (2009) ideas to frame our research on the effects of discipline-based PD events on teaching in the geosciences because a) these critical features were described as comprising a conceptual framework for studying the effects of PD, b) the features are described as being independent of the type of activity (e.g. workshop, webinar, journal club), and c) the features can be readily mapped to the characteristics of a variety of discipline-based PD events. If PD contains these critical features, we propose that it is more likely to result in conceptual change, as measured by a difference in teaching practices for those with differing amounts and types of that PD.

### **Critical Features of Professional Development**

Desimone (2009) reviewed empirical research on effective PD to develop a conceptual framework that includes five critical features demonstrated to make PD effective for improving teaching practice. These critical features include: a focus on subject matter content (content focus), opportunities for PD participants to engage in active learning (active learning), opportunities for PD participants to work collectively where discourse is supported by participants' commonalities such as discipline (collective participation), consistency between PD content and instructors' prior knowledge or experience (coherence), and opportunities for participants to engage in PD over a sufficient period of time (duration; Desimone 2009, and references therein). These critical features that promote conceptual change are aligned with models for discipline-based PD that bring together participants interested in similar content and who work together in an active environment to explore new pedagogies.

The critical features of active learning, collective participation, and coherence are met by the PD events included in this study (e.g. Manduca et al. 2017). However, instructors who participated in these discipline-based PD events experienced different levels of participation in terms of duration (both the number of PD events attended and the length of those events), and the content focus, which varied with respect to its alignment with instructors' disciplinary teaching. We therefore explore the critical features of content focus and duration in greater detail, and connect them to theoretical constructs that help to explain why these features are important.

### **Content Focus and Pedagogical Content Knowledge**

The critical feature of content focus describes PD activities that focus on disciplinary content and how students learn that content and, based on empirical studies, is

identified by Desimone (2009) as being the most influential for improving instructors' teaching practice. For example, a regression analysis of survey results from over 1000 mathematics and science teachers identified content focus as a significant feature of PD that increases teachers' knowledge and skills, and that results in changes to teaching practice (Garet et al. 2001).

The importance of content as a critical feature can be understood in the context of pedagogical content knowledge, or PCK, which combines instructors' knowledge of pedagogical practices with disciplinary knowledge (Shulman 1986) to help instructors teach disciplinary content using the most effective means. Magnusson et al. (1999) expanded the conceptualization of PCK to encompass assessment, pedagogy, content, knowledge of students, and curricular knowledge. Amplifiers and filters to this model of PCK were identified by Gess-Newsome (2015), including teaching beliefs, prior knowledge, and context (e.g. class size, institution type).

Using PCK as a framework for interpreting the success of faculty PD considers faculty as learners about teaching in a particular discipline (e.g. Fraser 2016; Major and Palmer 2006; Mulnix 2016; Sunal et al. 2001). Faculty participating in discipline-general PD may see or practice with examples of how to implement particular pedagogical practices that are unrelated to the disciplinary content they teach; for example a geoscientist may attend a workshop about using case studies, but examples from a humanities course may not seem plausible to use in a geoscience course. However, faculty participating in discipline-specific PD will see or practice with examples from their discipline that directly relate to the disciplinary content they teach, or that, if not directly related, can be understood in a disciplinary context.

Discipline-focused PD offers the opportunity for participants to learn and practice activities that are directly tied to what they teach, or at least similar enough that translation to their own teaching is more direct (e.g. Garet et al. 2001; Henderson et al. 2011). This means that instructors are applying their newly developed ideas and practices in conditions similar to those in which they are learned. When faculty try to apply newly developed ideas and practices (PCK) in conditions that are similar to those in which they are learned, the barriers to implementing PCK are likely to be perceived as more porous (easy to overcome) rather than dense (more difficult to overcome) (Gess-Newsome 2015). Therefore, discipline-specific PD may increase the ease of applying new PCK by increasing the porosity of barriers (such as the perceived relevance of a humanities case study activity to a geoscience course) by minimizing the effort needed to implement new PCK, or transferring PD content to one's teaching practice. As learners in the PD environment, instructors wanting to reform their teaching must build their knowledge of new teaching strategies and also organize that knowledge to allow deep and transformative learning that can be transferred to their teaching practice.

### Transfer of Learning to New Contexts

Transfer of one's learning about new teaching methods within a subject (e.g. from discipline-based PD) falls under the domain of acquiring expertise, which includes disciplinary content as well as understanding why and how the subject matter is taught to facilitate student learning (National Research Council 2000; Kreber 2002). Disciplinary content knowledge is generally accepted for faculty teaching in the area of their disciplinary expertise, which means that PD aligned with that expertise can

focus on pedagogical areas related to that discipline. In cases where the learning (PD) and application (instruction) are closely aligned, the transfer is best described as near transfer (Cree and Macaulay 2000). In situations where the PD content and teaching context are less aligned (e.g. a geoscientist learning about the use of case studies in a humanities context), participants must use far transfer to apply their new PCK to a relatively unique situation. PD that brings instructors together with disciplinary commonalities (e.g. collective participation; Desimone 2009) allows instructors to work together to contextualize the pedagogical practices explored, and can result in improved transfer (Gee and Gee 2007; Webster-Wright 2009).

However, most instructors do not teach only a single course in which they have specific content expertise. For such instructors, it is not sufficient to be an expert in one domain, they must be able to combine different specializations and broaden their expertise to be able to teach in other domains (Van der Heijden 2002), for example from having expertise in mineralogy to teaching volcanology. Transfer is known to be especially difficult when new information is presented in a single context rather than in multiple contexts (Bjork and Richardson-Klavehn 1989). Learning new information in a context distinct from an instructor's disciplinary context means that the learner is required to use far transfer, which is more difficult than near transfer (Perkins and Salomon 1992). In the case of discipline-based PD, this suggests that adaptation and application of pedagogical strategies learned in the context of one topic that need to be applied to another topic, may require additional PD to improve transfer.

Transfer can also be considered in terms of adaptive expertise in which instructors (in the role of learners building their PCK) are able to transfer knowledge learned in PD to domains other than the content area of the PD (e.g. Carbonell et al. 2014). There is a positive correlation between the extent of one's knowledge base and the ability to adapt that knowledge to other areas (Chen et al. 2005), known as adaptive performance or adaptive expertise (e.g. Carbonell et al. 2014). Adaptive expertise can develop from accumulated experiences, including prior knowledge; with additional PD, faculty integrate additional experiences and prior knowledge to address more complex problems (Hatano and Inagaki 1986). When faculty apply newly gained PCK, they build their adaptive expertise by modifying known procedures that have proven effective, and can respond with flexibility to contextual variations (e.g. Hatano and Oura 2003). For example, a faculty member who learns to use an activity with high instructional utility and efficacy (e.g. think-pair-share; McConnell et al. 2017) may be able to apply that method as an effective activity for a new subject they teach because they are able to apply their PCK and transfer it to a different context. Thus, PCK developed as part of discipline-based PD allows instructors to learn about then implement new pedagogies on the same PD topic. The fact that the PD is discipline-based may reduce the difficulty of transfer by making barriers to implementing teaching reform more porous. With additional PD, instructors build accumulated experiences, further developing their adaptive expertise, which allows them to transfer their PCK to teaching topics that were not the focus of PD activities.

## Duration of Professional Development

In addition to content, duration is another critical feature of PD explored by this study. Based on empirical studies of the effectiveness of PD in changing instructors'



knowledge, skills and teaching practices, Desimone (2009) concluded that effective PD efforts occur over extended periods of time. For example, Allen et al. (2011) saw improved student learning with participation of secondary teachers in a 20 h PD program that took place over 13 months, and Supovitz and Turner (2000) used K-12 teacher survey results to associate 40+ hours of PD participation with inquiry-based teaching practices. Ho et al. (2001) suggested PD should be dispersed across at least four weeks, while others have suggested more than a semester (Emerson and Mosteller 2000; Gibbs and Coffey 2016), or more than one year (Postareff et al. 2007). Many other studies also point to the importance of PD that extends over time or is of “sufficient” or “substantial” duration (Ebert-May et al. 2015; Garet et al. 2001; Pelch and McConnell 2016; Postareff et al. 2007; Wilson 2013).

In a review of 191 articles about promoting instructional changes in undergraduate STEM courses, Henderson et al. (2011), conclude that effective strategies for promoting instructional change involve interventions lasting at least one semester. Likewise, studies suggest that an individual’s ideas about teaching, and their teaching practice, change over time, rather than as a result of “one-shot” workshops (Loucks-Horsley et al. 2009; Postareff et al. 2007). In a review of 36 papers on the effects of PD in higher education, Stes et al. (2010) also concluded that PD that takes place over time results in more positive changes in instructor behavior than one-time events. They noted, however, that their sample contained only a small number of studies investigating the impact of one-time events, and that further studies were needed (Stes et al. 2010).

The importance of time as a variable for learning was conceptualized by Carroll (1963, 1989) for the school classroom setting, but can be applied to learning in all contexts. Bloom (1974) begins his paper “Time and Learning” with the statement, “All learning, whether done in school or elsewhere, requires time” (p. 682). The idea that learning takes time is implicit in many learning theories, including conceptual change theory. In the context of PD activities, instructors need time to learn about and accept new pedagogical strategies as having value, try them in their courses, reflect on their success, and develop new methods to integrate them into their teaching practice (Henriques 1998; Loucks-Horsley et al. 2009). This can occur during, after, or between professional development events.

Duration is only one of the identified critical features of PD, and while empirical studies suggest that long term PD efforts are more strongly associated with positive outcomes, the possibility of effective short duration PD events is not ruled out (e.g. Henderson et al. 2011). There may be circumstances, perhaps when all of the other critical features of effective PD are met, including content focus that supports the development of PCK, that makes the duration of an event is less critical.

### Connecting the Theoretical Framework and Research Questions

In this study, we combine classroom observations with compilations of instructors’ PD histories to test the hypothesis that participation in discipline-specific PD results in the transfer of PD content to classroom instruction as evidenced by use of student-centered teaching practices (RQ 1). We also test the hypothesis that PD experiences that are discipline-specific and aligned with the content being taught are most likely to result in student-centered teaching (RQ 3). These hypotheses are underpinned by the notion that

participation in discipline-based PD results in changes to instructors' PCK, which in turn results in changes in teaching practice that can be observed using the RTOP.

The PD events considered in this study have participants engage in active learning, emphasize collective participation through the inclusion of instructors with common disciplinary interests, and support coherence in ways such as gathering information about participants' ideas and practices in advance of the PD event, or by attracting volunteer participants who have common interests or experiences (e.g. Manduca et al. 2017). Because the features of the examined PD programs are consistent with Desimone's (2009) critical features of active learning, coherence, and collective participation, but vary in terms of their degree of content focus and duration, we are able to test hypotheses related to these variables.

Our study includes PD events that focused on specific disciplinary topics (e.g. OCE's Teaching Geomorphology) as well as events that are not specific to a disciplinary topic (e.g. PD regarding student motivation), but all PD events had content and examples connected to the geosciences. We are therefore able to examine the relationship between participation in PD with varying degrees of content focus and instructors' teaching practice (RQ 1, 3). Likewise, the duration of the PD events varies, as does the number of PD events that each participant attended; therefore, we can examine the relationship between time spent in PD and instructors' teaching practice (RQ 1, 4), as well as the relationship between elapsed time since PD participation and instructors' teaching practice (RQ 2).

## Methods

To investigate the impact of geoscience-specific PD on undergraduate teaching, we compared participation in PD events archived by SERC (hereafter referred to as PD) with observed teaching practices for 236 undergraduate geoscience instructors at a variety of institutions across the United States. Specific types of events in which geoscience instructors participated are described in Online Resource 1 along with participation levels of observed instructors. Consent was obtained from all participants, and research protocols were reviewed and approved by the Carleton College Institutional Review Board.

### Professional Development Participation

After all classroom observations were completed, we compiled a history of PD events attended by each of the 236 instructors prior to the date of their RTOP observation. The PD events considered for this study are those for which agendas and participants are archived by SERC and published on public web pages. These programs share similar organizational elements established by OCE (Manduca et al. 2010; Manduca et al. 2017), and align with the PD critical features of content focus, active learning, coherence, and collective participation as described by Desimone (2009). Compiled PD events occurred from November, 2008 through August, 2015. We recognize that observees could have participated in PD outside of the events we compiled so PD events archived by SERC are used here as a minimum representation of PD participation.



We used published agendas to determine the total number of hours associated with each event, the mode of delivery (face-to-face or virtual), whether participants met over consecutive days or had punctuated meetings, and the amount of time between instructors' RTOP observation and their most recent pre-observation PD event (referred to as "elapsed time"). For all in-person events, the number of hours assigned to the workshop are for synchronous sessions and do not account for asynchronous time spent on activities such as preparatory or homework assignments. For virtual events, the number of hours assigned to the workshop includes all synchronous hours and only includes asynchronous time indicated by event agendas.

## Observed Teaching Practice

While teaching practice can be measured by methods including self-report data (e.g. Ebert-May et al. 2015; Walczyk and Ramsey 2003) and direct observations (e.g. Stains et al. 2018; Teasdale et al. 2017), direct observation of teaching practice is often considered the most unbiased measure (e.g. Desimone 2009). Many classroom observation protocols exist (see Teasdale et al. 2017 for a review), but the RTOP is a validated instrument designed to assess the use, and quality of use, of research-based pedagogies (Sawada et al. 2002) that are consistent with the reformed teaching practices promoted by the PD programs archived by SERC. The RTOP measures the degree to which classroom instruction employs different aspects of active learning including interactions among students, interactions between students and the instructor, emphasis on fundamental concepts, and the incorporation of student ideas into class trajectory (Lawson et al. 2002; MacIsaac and Falconer 2002; Sawada et al. 2002). The RTOP contains 25 items that are scored from 0 to 4, resulting in possible total scores from 0 to 100, with higher scores resulting from more reformed, student-centered teaching practices. Although the RTOP was developed for use in the K-12 setting (MacIsaac and Falconer 2002; Sawada et al. 2002), the instrument has strong predictive validity for student achievement in university settings, with higher RTOP scores correlating with more student-centered teaching and with higher student achievement (Bowling et al. 2008; Falconer et al. 2001; Lawson et al. 2002). The RTOP has been used in many contexts including describing teaching practices across a discipline in higher education (Budd et al. 2013; Teasdale et al. 2017), and describing changes to teaching practice associated with PD programs for higher education faculty and future faculty (Ebert-May et al. 2011, 2015; Manduca et al. 2017). While there are many conceptions of teaching quality that could be considered (e.g. teaching to promote inclusivity), the alignment of the RTOP items with the reformed teaching practices promoted by the PD programs we considered, the high interrater reliability for the instrument, and the numerical scores that are well suited to a large-scale quantitative research project, led us to choose the RTOP to measure the degree of reformed teaching in participating instructors' classes.

A cohort of trained observers used the RTOP to make in-person classroom observations between November, 2008 and May, 2016. Observers used a modified version of the RTOP rubric (Budd et al. 2013; Teasdale et al. 2017). Several cohorts of observers were trained using a three-stage process of observing and scoring videos of university science classes and discussing scores with a calibrated training leader. Trainees only advanced from one stage to the next if their scores were within one

standard deviation of the accepted score. The final stage of training required observers to score two final calibration videos. Calibration of two videos resulted in a Cronbach's alpha of 0.996, which exceeds the threshold for inter-rater reliability ( $\alpha > 0.7$ ; Multon 2010; as reported in Teasdale et al. 2017). Other researchers have also found high inter-rater reliability using the RTOP (Amrein-Beardsley and Popp 2012; Marshall et al. 2011). The quality of the observer pool was maintained through an annual calibration process in which all observers were asked to score a video, and only those observers with scores within one standard deviation of the accepted score were able to contribute observations to the project. Additional details of the training and calibration procedures for the observer cohort are described by Teasdale et al. (2017).

Reproducibility of observations by individual observers (intra-rater reliability) was calculated using 14 repeated analyses of videotaped classrooms by three observers using the method of Bland and Altman (2003). The Bland-Altman test of intra-rater reliability showed good agreement between repeated observations. The mean difference (bias) was  $-0.5$  points out of 100 over a year or more. All repeated observations fell within the 95% upper and lower limits of agreement with a standard deviation of 3.48 points. This intra-rater reliability is the first reported for the RTOP and represents an additional measure of confidence in the utility of classroom observations using the RTOP.

The sample of instructors used in this study is an expansion of the 204 instructors observed by Teasdale et al. (2017). Observed instructors represent a convenience sample in that observers contacted instructors to request an observation visit on any of the following bases: instructors were either geographically nearby (easy to travel to), instructors had asked to be observed, or instructors fit a geographic location or institution type that were under-represented in the observed pool. Additional efforts were made to include instructors across a range of demographic factors including institution type, class size, and course level. We did not target instructors with specific records of PD participation; instructors' pre-observation PD histories were determined after they were observed. As documented by Manduca et al. (2017), a large proportion of study participants attended no workshops and did not use website resources associated with SERC, so our convenience sample indicates no self-selection.

Total RTOP score was used to assign each observation to an instructional category according to the classification established by Budd et al. (2013): Teacher-Centered (RTOP score  $\leq 30$ ), Transitional (RTOP score 31–49) or Student-Centered (RTOP score  $\geq 50$ ).

### Self-Report of Teaching Practice and Professional Development Participation

General teaching practices of observed instructors were evaluated using an instructor survey (Manduca et al. 2017; Teasdale et al. 2017) in which instructors reported information about themselves (e.g. years teaching), their course (e.g. grade level), their pedagogical practice, and resources they used to learn about teaching methods (Manduca et al. 2017). Results from these surveys, taken at the time of the observation, were compared with results of observations of single class periods for each instructor (Teasdale et al. 2017), and were used to compare instructors' self-report of PD participation with compiled PD histories.

## Topical and Non-Topical PD

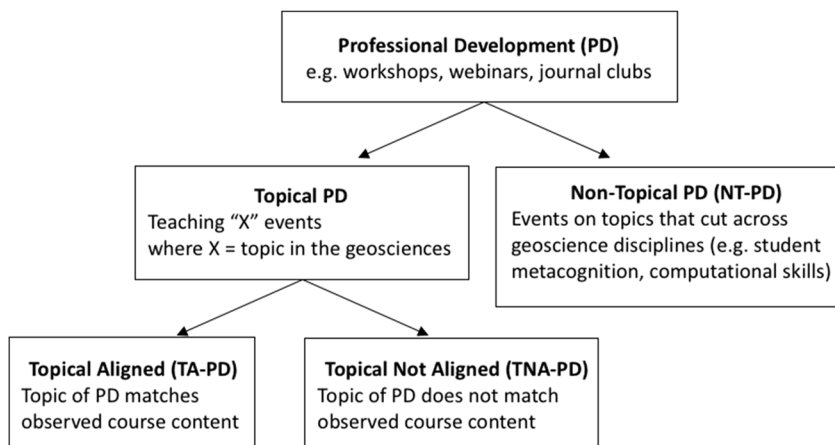
PD events were identified as having topical themes if the event focused on a content area of instruction such as hydrology or structural geology, and were identified as non-topical if the event focused on broader themes or program-wide abilities such as teaching with large data sets or developing students' metacognitive skills. Thus, an instructor was classified as having non-topical PD (NT-PD) if they attended only workshops that did not focus on specific disciplinary content.

## Alignment of PD Content and Observed Course Content

We established the relationship between the disciplinary content of an observed course and the disciplinary content of each instructor's PD by comparing the topics taught during observations with the PD each instructor attended. Instructors were classified as having topical and aligned PD (TA-PD) if they attended PD about teaching specific disciplinary content and were observed teaching a course about that content (Fig. 1). For example, an instructor who attended the Teaching Mineralogy workshop and was observed teaching a mineralogy course was considered to have had TA-PD. If an instructor attended a topical workshop but was observed teaching a different topic, they were classified as having topical but not aligned PD (TNA-PD; e.g. an instructor who attended a Teaching Hydrology workshop but was observed teaching a paleontology course).

Attending a Course Design workshop was considered topical PD because attendees work on developing a specific course of their choosing, and there is a record of course names or descriptions for each participant. An instructor who was observed teaching the same course that they developed during a Course Design workshop was classified as having TA-PD. Instructors who were observed teaching a different course than the one they developed during a Course Design workshop were classified as having TNA-PD. In a few cases we could not determine what course an instructor developed during a Course Design workshop and they were classified as having TNA-PD.

Assignments of topical alignment were made by a single investigator and reviewed by two additional researchers. In many cases, there was not an exact match between



**Fig. 1** Alignment of PD content and disciplinary content of observed course

course and workshop names, but the equivalency in disciplinary content was determined based on the collective experience and disciplinary expertise of the researchers. In cases where there was uncertainty about alignment, we erred on the side of assuming non-alignment.

Instructors classified as having TA-PD may also have attended TNA-PD and/or NT-PD. Likewise, instructors classified as having TNA-PD may also have attended NT-PD.

## Statistical Analyses

All statistical analyses were conducted using SPSS version 22. Two-sample t-tests were used to compare RTOP scores for instructors with no record of PD participation and those who attended one or more PD events (comparison of “No PD” and “1+ PD” groups). Relationships between RTOP score and PD events, PD hours, and elapsed time between PD participation and classroom observation were assessed using Spearman correlations for nonparametric data.

To assess whether PD events or PD hours were a stronger predictor of Student-Centered classrooms when adjusting for the type of PD, we ran three binary logistic regression models. In all models the dependent variables were RTOP scores of 50+ (Student-Centered) vs. 0–49 (not Student-Centered) and the predictor variables were doses of PD in each of the three PD types (TA-, TNA-, NT-PD). Neither events nor hours were tested as continuous variables due to their non-normal distributions. For the first model, events were categorized into 0, 1, 2+ PD events based on the variable distributions. There were only  $n = 2$  participants with 2 TA-PD events, so TA-PD event category was dichotomized to no TA-PD vs. any TA-PD. The second model tested hours categorized using 20 h as a cut point based on Allen et al. (2011) and Desimone (2009). To get a model to fit with the hours predictors, we had to combine no NT-PD hours with 1–19 NT-PD hours. Combining the 0 h and 1–19 h categories was a result of not having any participants with 1–19 h of NT-PD events in the Student-Centered category. The final model used a combination of events and hours by segmenting the one event categories into one event with 1–19 h and one event with 20+ hours.

An optimal cut-point for the number of PD hours needed to increase the likelihood of instructors using Student-Centered teaching was determined using the minimum  $p$  value approach with Bonferroni correction (Mazumdar and Glassman 2000; Williams et al. 2006). In this approach we performed chi-square tests of Student-Centered (vs. not) and multiple possible cut-points of hours and looked for the lowest  $p$  value. For example, we performed a chi-square test of Student-Centered (vs. not) by 0 to 11 h (vs. 12 or more); and then tested Student-Centered (vs. not) by 0 to 15 h (vs. 16 or more); etc. We limited the selection area for the multiple tests by excluding the top and bottom 10% range of hours and binned hours into four-hour increments. The result was a selection area from 12 to 144 h in four-hour increments, or 34 tests. Using the Bonferroni correction, we set  $\alpha$  at  $0.05/34 = 0.0015$ .

Another binary logistic regression model was used to predict Student-Centered classrooms among those who had either Student-Centered (RTOP score 50+) or Teacher-Centered (RTOP score 0–30) classrooms. This model used a subset of the study population, excluding those in the Transitional category (RTOP score 31–49), with events and type of PD as predictors.

## Limitations

Although efforts were made to observe instructors with varying demographic characteristics, the sample includes a small percentage of all United States geoscience instructors, and a small percentage of PD participants. Manduca et al. (2017) estimate that there are approximately 10,000 geoscience faculty teaching undergraduate courses. Likewise, from 2002 to 2012, more than 2800 individuals participated in OCE events alone (Manduca et al. 2017). Many more participated in subsequent OCE and other PD events compiled here.

Workshops offered through programs archived by SERC are not the only source of PD. Participants in this study may have benefitted from PD at their home institution, through professional organizations, online resources, publications, personal mentoring, or other means. The quantity of PD workshops we report should be considered a minimum, or under-representation, of the PD each instructor experienced.

Observations of each instructor were made in a single “lecture” class period and did not include observation of any laboratory, field, or recitation sections. No online courses were observed. Both Ebert-May et al. (2015) and Teasdale et al. (2017) report good agreement between RTOP observations and self-report of instructors’ general teaching practices providing confidence that results of single RTOP observations are accurate measures of teaching practice. In addition, we attempted to mitigate the limitation of a single observation by communicating with observed instructors in advance of observations to ensure that they would describe the observed class as “typical” of their teaching practice.

Our assignment of the number of hours associated with each PD event assumes full participation of attendees for both in-person and virtual workshops. We cannot determine whether specific individuals were present for entire events, or the degree to which participants were mentally engaged in programs. Our assignment of hours does not consider time spent on preparatory assignments, or discussions and work that happened during unscheduled time.

## Results

### Demographics

Instructor survey data indicate that observed instructors ( $n = 236$ ) are 41% female and 59% male, with 34% reporting a rank of full professor, 22% associate professor, 21% lecturer or instructor, 14% assistant professor, 8% adjunct professor, and 1% other. Observations occurred at institutions categorized as research/doctoral (47%), masters granting (28%), bachelors granting (6%), and associates granting (18%), based on the Carnegie Classification System for Institutions of Higher Education (Carnegie Classification 2015). Observed classes were introductory geoscience (59%) and upper-level geoscience courses (39%) that were categorized as small ( $\leq 30$  students, 51%), medium (31–79 students, 30%), or large ( $\geq 80$  students, 19%).

## PD Histories

PD histories include 53% of instructors ( $n = 125$ ) with no PD record before the date they were observed, and 47% of instructors ( $n = 111$ ) with at least one PD event before they were observed. Of the instructors with PD, participation included one event (41%;  $n = 45$ ), two events (18%;  $n = 20$ ), three events (11%;  $n = 12$ ), four events (8%;  $n = 9$ ) or at least five events (22%;  $n = 25$ ). Virtual PD events were attended by 27 observees, but only three instructors participated in only virtual events (1 event each). Individual PD events range in duration from 1 to 43 h, and individuals who participated in PD events have between 1 and 503 h of PD (1 to 22 events).

## RTOP Scores

Total RTOP scores for observed classes range from 13 to 89 with an average score of 39.0. Observed classes include 74 in the Teacher-Centered instructional category (31% of observations), 109 in the Transitional instructional category (46%), and 53 in the Student-Centered instructional category (22%; Fig. 2). As was established by Teasdale et al. (2017), RTOP scores do not vary systematically with demographic variables such as class size, institution type, or course level.

## Survey Responses

Survey results were used to investigate self-reported use of workshops for teaching-related PD. Participants responded to the question, “How do you learn about new teaching methods?” by choosing all that apply from these options: professional meetings or workshops; publications; discussion with faculty members in my department; discussions with colleagues at other institutions; online resources; my own research. Professional meetings or workshops (meetings/workshops) was selected by 64% of participants, and was the third most frequently selected option behind “discussions with faculty members in my department” (77%) and “discussions with colleagues at other institutions” (66%).

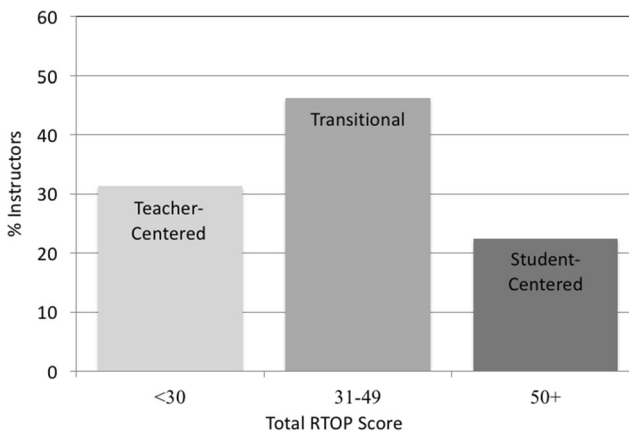


Fig. 2 Instructional categories based on total RTOP scores for observed classes ( $n = 236$ )



We compared self-reported importance of PD participation to compiled PD histories. Survey responses and compiled PD history match for 67% of study participants; 39% of participants report learning about new teaching methods at meetings/workshops and have a record of attending PD, and 28% of participants do not report learning about new teaching methods at meetings/workshops and have no record of attending PD.

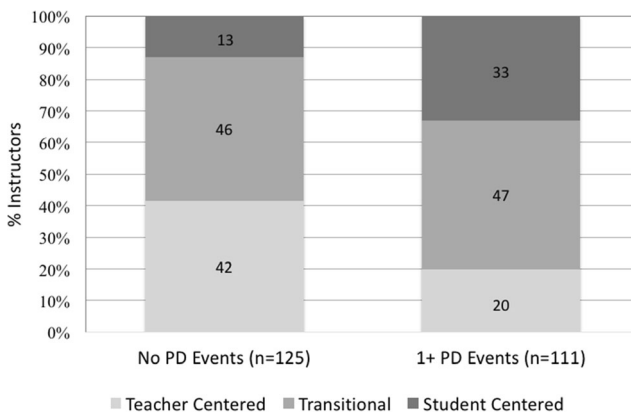
Survey responses and compiled PD history do not match for 33% of study participants; 25% of participants report learning about new teaching methods at meetings/workshops but have no record of attending PD, and 8% of participants do not report learning about new teaching methods at meetings/workshops but have a record of attending PD.

### Participation in Discipline-Specific PD and Teaching Practice (RQ1)

Average RTOP scores are higher for instructors who attended at least one PD event ( $p < 0.0001$ ). Instructors with no PD events ( $n = 125$ ) have an average RTOP score of 34.2 (42% Teacher-Centered, 46% Transitional, 13% Student-Centered). Instructors who participated in at least one PD event ( $n = 111$ ) have an average RTOP score of 44.5 (20% Teacher-Centered, 47% Transitional, 33% Student-Centered; Fig. 3).

Positive correlations between RTOP score and PD participation are observed for both virtual (online) and face-to-face (in person) modes of PD delivery. This relationship (Spearman's  $\rho = 0.3192$ ;  $p < 0.05$ ) was found to exceed Cohen's (1988) convention for a medium effect size ( $d = 0.50$ ), and represents a statistically significant relationship. When face-to-face and virtual workshops are separated, the statistical ( $p$  value) and practical (effect size) significances both remain valid for each sub-group. With only three faculty who completed only virtual events, the sample is too small to attempt additional statistical analyses, so participation in any event type, whether face-to-face, hybrid, or virtual, are grouped together for the remainder of our analyses.

RTOP scores are generally higher for instructors who attended a larger number of PD events (Spearman's  $\rho = 0.3428$ ,  $p < 0.05$ ;  $d = 0.70$ ). A comparison of RTOP scores for instructors who attended only a single PD event ( $n = 45$ ) to the total number of hours spent at that event shows a weak but positive correlation between RTOP score and event length ( $R^2 = 0.1134$ ,  $p < 0.01$ ;  $f^2 = 0.128$ ). Correlations between RTOP score



**Fig. 3** Comparison of instructional category for instructors with no PD history and instructors who attended at least one PD event

and quantity of PD (number of PD events or hours) were statistically significant, but have small effect sizes and correlation coefficients.

### Elapsed Time (RQ2)

There is a slight negative correlation between an instructor's RTOP score and the number of days since participation in their most recent pre-observation PD event, with wide scatter. The relationship (Spearman's  $\rho = -0.162$ ) represents a statistically insignificant relationship ( $p > 0.05$ ).

### Disciplinary Content of PD and Teaching Practice (RQ3)

#### Alignment of PD and Course Content

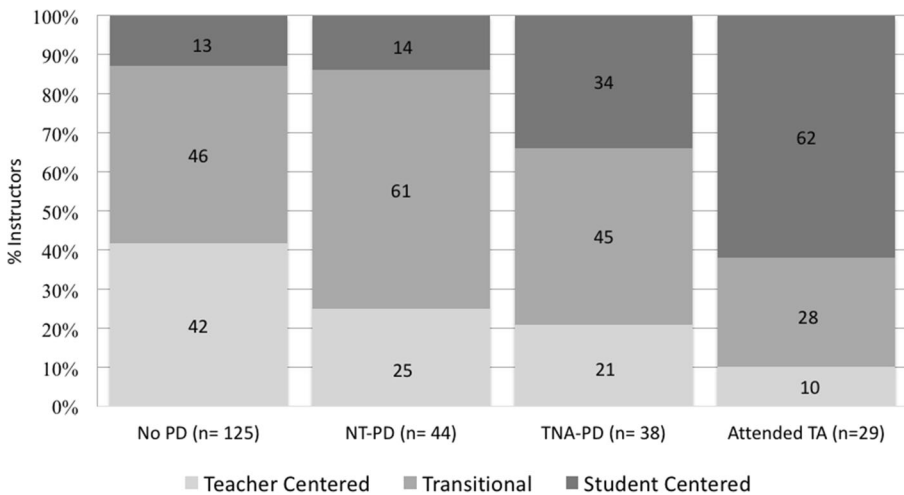
To examine the influence of specific PD content on instructors' teaching practices we compared RTOP scores for instructors in three groups: those who attended PD with the same topical content that was taught during their observation (TA-PD); those who attended PD on a topic different from that taught during their observation (TNA-PD); and those who attended PD that was not focused on disciplinary content (NT-PD; Fig. 1). RTOP scores of the 111 instructors who attended PD are described in Table 1 and Fig. 4.

The average RTOP score is highest for instructors with TA-PD, the majority of whom are in the Student-Centered instructional category (62%). Percentages of instructors in the Student-Centered instructional category who had TNA-PD (34.2%) and NT-PD (13.6%) are lower. The percentage of instructors in the Teacher-Centered instructional category is lowest for those who had TA-PD (10.3%) and highest for instructors who had NT-PD (25.5%). In contrast, 42% of instructors who did not participate in PD are in the Teacher-Centered instructional category.

Six of the 29 instructors who attended TA-PD attended only one workshop; five were observed teaching a Student-Centered class and one was observed teaching a Transitional class. Of the instructors who attended TA-PD plus additional PD ( $n = 23$ ), 57% were observed teaching Student-Centered classes, 30% were observed teaching Transitional classes, and 13% were observed teaching Teacher-Centered classes.

**Table 1** RTOP scores and instructional categories for instructors based on the topical alignment of PD attended and the course observed. Scores and instructional categories for instructors with no PD are included for reference

Type of Professional Development (PD)	n (236 total)	Average RTOP Score	Teacher- Centered (% instructors)	Transitional (% instructors)	Student- Centered (% instructors)
Topical, Aligned (TA-PD)	29	53.2	10.3	27.6	62.1
Topical, Not Aligned (TNA-PD)	38	43.2	21.1	44.7	34.2
Non-Topical (NT-PD)	44	39.8	25.0	61.4	13.6
No PD	125	34.2	41.6	45.6	12.8



**Fig. 4** Comparison of instructional category for instructors who experienced different alignment between PD content and observed course

A multiple linear regression predicted RTOP score based on number of PD events attended and the alignment of those events, resulting in a significant regression eq. ( $F(2,233) = 22.4322, p < 0.0001$ ), with an  $R^2$  of 0.161. Participants' predicted RTOP score is equal to  $35.80 + 1.34$  (number of PD events) +  $13.51$  (aligned PD), where number of events is coded as a simple count, and alignment is coded as a binary statement of whether or not the individual participated in aligned PD. The number of events and alignment were both significant predictors of RTOP score.

### Logistical Regression Models for Student-Centered Teaching

Results of three logistic regression models indicate that the number of PD events attended is a stronger predictor of a Student-Centered classroom (Nagelkerke adjusted  $R^2 = 0.241$ ) than hours (adjusted  $R^2 = 0.198$ ) or than events and hours combined (adjusted  $R^2 = 0.234$ ). Based on the superior predictive strength, outputs of the events model are presented and discussed below; however, all three are included in Online Resource 2.

A binary logistic regression using all 236 observations was used to predict inclusion in the Student-Centered instructional category (RTOP score 50+ vs. 0–49) using the number and alignment of PD events an instructor attended. Alignment of PD was categorized by TA-PD events (yes/no), TNA-PD events (0, 1, 2+), and NT-PD events (0, 1, 2+). After controlling for other types of PD, the odds of instructors with one or more TA-PD events ( $n = 29$ ) teaching a Student-Centered class are 5.6 times higher ( $p < 0.001$ ) than instructors without any TA-PD ( $n = 207$ ). The odds of instructors who participated in two or more TNA-PD events ( $n = 18$ ) teaching a Student-Centered class are 5.8 times higher than instructors without any TNA-PD ( $n = 181$ ). The odds of instructors with one TNA-PD event ( $n = 37$ ) teaching a Student-Centered class are not significantly higher than instructors without any TNA-PD events ( $n = 181$ ). The odds of instructors with one NT-PD event ( $n = 31$ ) or two or more NT-PD events ( $n = 52$ ) teaching a Student-Centered class are not significantly higher than instructors with no PD events ( $n = 153$ ).

## Unaligned PD and Instructional Categories

Instructors with TNA-PD or NT-PD are most frequently observed teaching Transitional classes (44.7%, 61.4% respectively; Table 1), and are observed teaching Teacher-Centered classes less frequently than instructors with no PD (25.0% vs. 41.6%). Faculty with no PD are most frequently observed teaching Transitional classes (45.6%), and slightly less frequently observed teaching Teacher-Centered classes (41.6%). If instructors attended any PD, the percentage of Teacher-Centered instruction decreases (from 41.6% to 25.0%) and the percentage of Transitional instruction increases (from 45.6% to 61.4%). Because of the low number of instructors with only TNA-PD ( $n = 16$ ), it was not possible to determine the odds of Transitional teaching based on NT-PD, TNA-PD or TA-PD participation. However, these results, along with the multiple linear regression showing small increases to RTOP score with each additional PD event, suggest that instructors with any PD are more likely to have a Transitional classroom than a Teacher-Centered classroom.

## Amount of PD Needed to Make a Difference (RQ4)

The optimal minimal number of PD hours needed to substantially affect the likelihood of teaching a Student-Centered class was a tie between 24 and 28. These two cut points had the lowest  $p$  value in the minimum  $p$  value approach. Therefore, we use the lower of the two (24 h) as a minimum number of hours needed to meaningfully impact the likelihood of teaching a Student-Centered class. Among instructors with 0 to 23 h of PD, 12.6% (19 of 151) were observed teaching Student-Centered classes; compared to 40.0% (34 of 85) of instructors with 24 or more hours of PD. Of the 236 instructors observed, 85 had had at least 24 h of PD; 23 of those (27.1%) accumulated those hours through one workshop, and an additional 16 (18.8%) through two workshops. Instructors with a high number of PD hours also attended a high number of PD events ( $R^2 = 0.82$ ;  $d = 4.27$ ), which prevents us from distinguishing the impact of the number rather than the duration of events.

## Discussion

### Discipline-Specific PD and Teaching Practice

Attending any of the PD events examined in this study increases the likelihood that an instructor will teach in a student-centered manner, as measured by RTOP score (RQ1). While our logistical regression models focused on predicting inclusion in the Student-Centered instructional category, teaching a Transitional class is also a desirable outcome of PD, especially for instructors whose original teaching practices may have been Teacher-Centered. The percentage of instructors observed teaching Teacher-Centered classes decreases with participation in any discipline-based PD, which is considered a measure of PD success. These results are consistent with the importance of disciplinary context and examples in developing PCK and decreasing the density of barriers to teaching reform (Gess-Newsome 2015), thereby increasing the likelihood that practices learned in PD will be transferred to one's classroom.

The target audience for all PD events considered in this study were broadly-defined geoscientists and our results highlight the positive impact of workshops that target instructors who teach similar content, consistent with the critical features of collective participation and content (Desimone 2009). Our results contrast with those of Stes et al. (2010), who, in a review of 36 studies on the effect of PD in higher education, investigated whether PD initiatives that target specific groups, such as those in a particular discipline, have more impact than initiatives without a target group, and found comparable outcomes for discipline-based and discipline-general studies. However, our results from direct observations of classrooms are in agreement with those of other authors who attribute positive impacts of PD to disciplinary content and the common disciplinary interest of PD participants. For example, observations, interviews and surveys by Marbach-Ad et al. (2015), indicate that successful implementation of PD is linked to disciplinary content and PCK, and is provided in the context of professional communities within academic departments. In another example from higher education, Ebert-May et al. (2015) attribute the success of a PD program for future biology faculty, at least partially, to the participants' common trait of being inexperienced teachers. The effectiveness of PD for K-12 teachers is reported to increase when PD is focused on a specific academic subject (e.g. Garet et al. 2001; Wilson 2013), and when the participants have characteristics in common such as teaching in the same discipline or at the same school (Garet et al. 2001). Participants' common academic discipline, or interest in teaching similar content, helps develop professional communities of practice (Manduca et al. 2017). Developing ideas about how to teach specific content may be easier when fellow PD participants share a common language and similar set of experiences (critical features of collective participation and coherence), and applying those new ideas may be easier when they are taught within a disciplinary context, consistent with the idea that discipline-based PD increases the porosity of potential barriers to PCK implementation.

Bouwma-Gearhart (2012) identified an interest in gaining confidence with teaching strategies as a primary motivator for faculty to seek out and participate in STEM higher education PD. Discipline-specific PD opportunities may be especially attractive to instructors who encounter dissatisfaction related to teaching disciplinary content, or who are seeking solutions to seemingly discipline-specific teaching challenges. This is consistent with Gess-Newsome's (2015) model that posits pedagogical dissatisfaction drives participation in PD. It may also appeal to those seeking out a supportive community of practice (Wenger 2000), wherein members of the community work toward developing their own understanding, thus nurturing the growth of the larger community (Bianchini et al. 2002). In a qualitative study of geoscience PD, Manduca et al. (2017) report that the majority of faculty attributed greater confidence in using specific teaching strategies to participation in PD workshops, with nearly half reporting shifts related to a change in their teaching beliefs. These findings are consistent with the Gess-Newsome (2015) model of amplifiers and filters moderating the influence of PD on practice, and the Park and Oliver (2008) construct of PCK including self-efficacy.

### Topical Alignment of PD Content and Transfer to Teaching Practice

Attending any discipline-based PD increases the likelihood of student-centered teaching, but PD focused on a specific disciplinary topic has a greater impact than PD

without a disciplinary topic focus (RQ3), consistent with the importance of disciplinary examples and context in developing PCK and transferring it to classroom practice. The high proportion of Student-Centered classes (83%) taught by instructors who only attended TA-PD suggests that attending TA-PD may be a threshold event, which is able to significantly impact one's teaching practice. While we do not know participants' teaching practices prior to their PD participation, only 35 instructors (15% of total population) who have not had TA-PD were observed teaching a Student-Centered class, so there is a strong connection between student-centered teaching and TA-PD.

Instructors who attended topical PD, but were observed teaching a different topic (TNA-PD) were most frequently observed teaching Transitional (44.7%) or Student-Centered (34.2%) classes. These results are consistent with research showing that individuals are able to transfer material that is similar to topics learned in training, and when topics are directly aligned with work activities (e.g. topically aligned) than material that is dissimilar (Blume et al. 2010). However, our data also show that two TNA-PD events are needed to be statistically comparable with just one TA-PD event, indicating that with additional topical PD, transfer of information from the topic of the PD to other topics is more likely. This is consistent with findings that near transfer is easier than far transfer (e.g. Blume et al. 2010), and implies that by having multiple TNA-PD events, transfer that would have been perceived as far is now perceived as near, as instructors recognize the similarities between the learning situation and transfer setting.

With participation in at least two TNA-PD events, it appears that instructors are able to use learned pedagogies in courses beyond the focus of the PD event. This suggests that during a topical workshop, instructors are focused on one topic and then, as they apply new pedagogical strategies within that topic (PCK), the reformed methodologies become ingrained across their teaching practice, and are more easily transferred to other courses that we captured in the form of TNA-PD observations. As is known for student learning, deepening one's knowledge not only involves acquisition of information, but also includes organizing and connecting new information to put it into context and make it useful when needed (e.g. National Research Council 2000; Chi 2008). As learners in the PD environment, instructors looking to reform their teaching must build their knowledge of new teaching strategies and also organize that knowledge, or their instructional PCK, to allow deep and transformative learning that can be transferred to their teaching practice. The idea of reformed methodologies becoming ingrained is well aligned with our results showing that as instructors attend more PD their ability to apply new knowledge to different areas of their teaching practice increases, consistent with the development of adaptive expertise. Considering that instructors are likely to teach more than one topical course in the discipline, attending more than one topical workshop is an effective mechanism for improving transfer.

### **Time Spent in PD, Elapsed Time, and High-Impact Single Events**

Our cut point analysis suggests that a minimum of 24 h of PD is needed to affect the likelihood of teaching a Student-Centered class (RQ4), a result consistent with the minimum 20 h of PD suggested by Desimone (2009) as a potential threshold for effecting change in teaching practices. Our analysis does not indicate whether those 24 h should ideally be split across multiple short events or completed in one longer



workshop. Of the 85 instructors who participated in at least 24 h of PD, only one had a maximum event length of less than 10 h, and 39 of the 85 participants reached the 24 h minimum with only one or two events.

Our result confirms that learning takes time; at least 24 h of PD, with the critical features described above, are most likely to result in Student-Centered instruction.

In agreement with Manduca et al. (2017), we suggest that one-time participation in a workshop may lead to changes in teaching practice, especially if the workshop is topical and at least 24 h in length. Of the 53 instructors observed teaching Student-Centered classes, 11 attended only a single topical PD event (5 TA-PD, 6 TNA-PD). A limitation of the study is that we do not know why instructors choose to attend PD or not, and we cannot rule out the possibility that instructors who attended PD were motivated to do so by the desire to make a specific instructional change which they then followed through in making. However, inclusion in the Student-Centered instructional category is most likely with participation in one to two topical events, rather than comparable participation in non-topical events, and this correlation is consistent with the idea that the single PD event can be effective (RQ4).

While Henderson et al. (2011) concluded that long-term interventions are more successful at promoting instructional change, they note an exception may be interventions designed to make specific or localized changes to instruction, such as using new technology. The two studies cited as illustrating this exception both include workshops that focus on use of technology, but they do so with disciplinary focus, consistent with the development of PCK. Kahn and Pred (2001) report on a series of two-day workshops to improve faculty use of technology in both STEM and humanities fields. Each workshop was targeted to faculty in a specific discipline (or set of related disciplines), and was designed to demonstrate the use of technology in teaching disciplinary content (Kahn and Pred 2001). While the use of technology in teaching would not be considered topical in our study, focusing on teaching specific disciplinary content with technology would be, and would theoretically promote the development of PCK. Campbell et al. (2007) describe workshops designed to bring modern genomics techniques to undergraduate biology classes, with a focus on using specific technology. Faculty participated in a 1.5-day workshop, with the option of taking an additional 2.5-day workshop, and in a follow-up survey one year after the PD, 79% of respondents reported using the workshop materials in at least one class (Campbell et al. 2007). While these workshops were focused on technology, they had a disciplinary focus consistent with supporting the development of PCK.

We have no evidence that elapsed time since PD participation affects teaching practice. There is no significant correlation ( $p > 0.05$ ) between the amount of time that passed since PD participation and RTOP score (RQ2). Hypotheses that either more elapsed time helps instructors integrate and implement their new knowledge, or that more elapsed time leads faculty to forget or dismiss new pedagogical ideas, are not supported.

### Factors beyond PD Participation

While we found a positive correlation between the number of PD events an instructor attended and their RTOP score, attending more events does not always correlate with high RTOP scores, and instructors with a wide range of PD experience were observed

teaching classes in all instructional categories. We recognize that implementing change in one's classroom is related to a wide variety of factors, beyond simply participating in PD and the state of the instructor's PCK (Blume et al. 2010; Borrego and Henderson 2014). In addition to PD, there are numerous additional logistical conditions that must be dealt with in order for PD content to become ingrained in one's teaching practice such as the extent of an instructor's autonomy, opportunities to implement change, support for implementation, and class contexts (e.g. Blume et al. 2010). Eighteen categories of barriers to implementing change to teaching practice were identified by Shadle et al. (2017). A simplistic reduction of such a wide array of factors can focus on intrinsic factors, those that are related to an instructor's choices, and external, or situational, factors, which are those beyond the instructor's control.

Potential intrinsic factors include one's cognitive ability to incorporate the learning outcomes from PD into one's teaching (e.g. Blume et al. 2010), one's professional identity (Brownell and Tanner 2012), and one's motivation to change and one's self-efficacy or tolerance for change. As noted previously, intrinsic factors such as teaching beliefs, self-efficacy and personal motivation may act as amplifiers or filters to how PCK is enacted in practice (Gess-Newsome 2015).

External factors that are outside the purview of the instructor are also known as possible factors that inhibit transfer of PD to one's teaching practice. Examples include departmental or institutional support for change, workloads that facilitate (or not) implementation (Parker et al. 2016; Manduca et al. 2017), content coverage expectations, class sizes and classroom arrangements (Doyle 1986) and the perceived level of student buy-in.

Professional identities and institutional and departmental culture can also play a role in decisions (or motivation) to integrate PD to one's teaching practice (Henderson and Dancy 2007; Brownell and Tanner 2012; Austin 2011). Even when reformed teaching is valued by one's department peers or the university administration, instructors may still feel a need to justify the use of methods learned in PD (Hatano and Inagaki 1986). Alternatively some faculty have found resistance in their attempts to modify their teaching practice from their colleagues (Brownell and Tanner 2012).

While instructor decisions related to their teaching practice are multifaceted, our study does find a significant relationship between discipline-based PD participation and observed classroom practices, consistent with PD participation strengthening instructors' PCK. Exploring exceptions to this relationship is an avenue for future research.

## Implications

Our results suggest that PD focused on teaching specific disciplinary content is highly effective in promoting student-centered teaching. In addition, topical events may be an accessible way to attract faculty to participate in PD because they already have interest in the topic and in interacting with other scientists in their discipline. The combination of discipline-focused pedagogy, learning theory, course and activity design, and community sharing results in the highest transfer of PD content to classroom practice. PD providers should therefore prioritize the development and offering of topic-focused PD in the discipline, and leaders such as department chairs, deans, and others looking to promote teaching reform should encourage participation in such events. Because

instructors are likely to teach more than one disciplinary course, and because our study suggests that attending at least two topic-focused PD events results in greater transfer to classroom practice, participation in more than one topical PD event should be encouraged.

Participation in topical PD appears to be most impactful to instructors in our study, but participation in any pedagogically-focused PD is also likely to result in more student-centered teaching. While PD focused on specific disciplinary content may be attractive to some instructors, others may be interested in learning about specific pedagogical tools, or in the development of specific skills. Instructor motivations for participating in PD and for enacting pedagogical reform are complex, so offering a variety of entry points for instructors to participate in PD is prudent.

Participation in a minimum of 24 h of PD increases the likelihood of teaching a Student-Centered class. Most participants in our study achieved that minimum by attending a small number of longer workshops rather than many short events, but our analysis does not indicate an ideal distribution of the timing of PD participation. PD providers who hope to impact teaching practice through one or two workshops should consider a minimum target of 24 h of synchronous time. There is no indication that PD participation either “grows in” and becomes more impactful with elapsed time nor “dies out” and becomes less impactful.

While many questions remain about instructors’ motivations for participating in PD and the conditions under which PD content gets transferred to classroom practice, this work shows the benefits of discipline-specific PD participation for promoting implementation of reformed teaching practices and provides guidelines to help PD developers and promoters make decisions to best achieve their goals.

**Acknowledgements** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. On behalf of all authors, the corresponding author states that there is no conflict of interest. The authors wish to thank the team of observers who invested much time and effort in being trained to use the RTOP instrument, and in making the classroom observations for this project. We also thank the 236 instructors who agreed to have their classrooms observed. Funding for this work comes from National Science Foundation Division of Undergraduate Education grant 1022844. Three anonymous reviewers provided helpful and critical comments that greatly improved this manuscript; we thank them for their time and attention.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037.
- Amrein-Beardsley, A., & Popp, S. E. O. (2012). Peer observations among faculty in a college of education: Investigating the summative and formative uses of the reformed teaching observation protocol (RTOP). *Educational Assessment, Evaluation and Accountability*, 24(1), 5–24.

- Austin, A. E. (2011). Promoting evidence-based change in undergraduate science education. Presented at the fourth committee meeting on status, contributions, and future directions of discipline-based education research, Washington, D.C.
- Austin, A. E., & Sorcinelli, M. D. (2013). The future of faculty development: Where are we going? *New Directions for Teaching and Learning*, 2013(133), 85–97.
- Bianchini, J. A., Hilton-Brown, B. A., & Breton, T. D. (2002). Professional development for university scientists around issues of equity and diversity: Investigating dissent within community. *Journal of Research in Science Teaching*, 39(8), 738–771.
- Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes, the Tulane Flowerree symposium on cognition* (pp. 313–344). New Jersey: Lawrence Erlbaum Assoc. Publishers.
- Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: Analyses of measurement studies. *Ultrasound in Obstetrics and Gynecology*, 22(1), 85–93.
- Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29(9), 682–688.
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36(4), 1065–1105.
- Borrego, M., & Henderson, C. (2014). Increasing the use of evidence-based teaching in STEM higher education: A comparison of eight change strategies. *Journal of Engineering Education*, 103(2), 220–252.
- Bouwma-Gearhart, J. (2012). Research university STEM faculty members' motivation to engage in teaching professional development: Building the choir through an appeal to extrinsic motivation and ego. *Journal of Science Education and Technology*, 21(5), 558–570.
- Bowling, B. V., Huether, C. A., Wang, L., Myers, M. F., Markle, G. C., Dean, G. E., Acra, E. E., Wray, F. P., & Jacob, G. A. (2008). Genetic literacy of undergraduate non-science majors and the impact of introductory biology and genetics courses. *BioScience*, 58(7), 654–660.
- Brownell, S. E., & Tanner, K. D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and...tensions with professional identity? *CBE—Life Sciences Education*, 11(4), 339–346.
- Budd, D. A., van der Hoeven Kraft, K. J., McConnell, D. A., & Vislova, T. (2013). Characterizing teaching in introductory geology courses: Measuring classroom practices. *Journal of Geoscience Education*, 61(4), 461–475. <https://doi.org/10.5408/12-381.1>.
- Campbell, A. M., Ledbetter, M. L. S., Hoopes, L. L., Eckdahl, T. T., Heyer, L. J., Rosenwald, A., Fowls, E., Tonidandel, S., Bucholtz, B., & Gottfried, G. (2007). Genome consortium for active teaching: Meeting the goals of BIO2010. *CBE—Life Sciences Education*, 6(2), 109–118.
- Carbonell, K. B., Stalmeijer, R. E., Konings, K. D., Segers, M., & van Merriënboer, J. J. G. (2014). How experts deal with novel situations: A review of adaptive expertise. *Educational Research Review*, 12, 14–29.
- Carnegie Classification (2015). The Carnegie Classification of Institutions of Higher Education. Retrieved from <http://carnegieclassifications.ie.edu/>
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64(8), 723–733.
- Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, 18(1), 26–31.
- Chen, G., Thomas, B., Wallace, J. C. (2005). A Multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *Journal of Applied Psychology*, 90(5), 827–841.
- Chi, M. T. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61–82). Hillsdale: Erlbaum.
- Clark, G., Blumhof, J., Gravestock, P., Healey, M., Jenkins, A., Honeybone, A., et al. (2002). Developing new lecturers: The case of a discipline-based workshop. *Active Learning in Higher Education*, 3(2), 128–144.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Vol. 2). Hillsdale: Erlbaum.
- Cree, V. E., & Macaulay, C. (Eds.). (2000). *Transfer of learning in professional and vocational education*. London: Routledge.
- Czajka, C. D., & McConnell, D. (2016). Situated instructional coaching: A case study of faculty professional development. *International Journal of STEM Education*, 3(1), 542–514.
- Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78(10), 1056–1063.
- Derting, T. L., Ebert-May, D., Henkel, T. P., Maher, J. M., Arnold, B., & Passmore, H. A. (2016). Assessing faculty professional development in STEM higher education: Sustainability of outcomes. *Science Advances*, 2(3), e1501422–e1501422.

- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 392–431). New York: Macmillan.
- 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 professional development programs. *BioScience*, 61(7), 550–558.
- Ebert-May, D., Derting, T. L., Henkel, T. P., Middlemis Maher, J., Momsen, J. L., Arnold, B., & Passmore, H. A. (2015). Breaking the cycle: Future faculty begin teaching with learner-centered strategies after professional development. *CBE-Life Sciences Education*, 14(2), ar22.
- Emerson, J. D., & Mosteller, F. (2000). Development programs for college faculty: Preparing for the twenty-first century. *Educational Media and Technology Yearbook*, 25, 26–42.
- Falconer, K., Wyckoff, S., Mangala, J., Sawada, D. (2001). Effect of reformed courses in physics and physical science on student conceptual understanding. Paper presented at the annual conference of the American Educational Research Association, Seattle.
- Fraser, S. P. (2016). Pedagogical content knowledge (PCK): Exploring its usefulness for science lecturers in higher education. *Research in Science Education*, 56(1), 141–161.
- Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., Dirks, C., & Wenderoth, M. P. (2007). Prescribed active learning increases performance in introductory biology. *CBE-Life Sciences Education*, 6(2), 132–139.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945.
- Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation*. Washington, DC: U.S. Department of Education. Retrieved from [https://www.mdrc.org/sites/default/files/full\\_442.pdf](https://www.mdrc.org/sites/default/files/full_442.pdf). Accessed 20 March, 2019.
- Gee, J., & Gee, J. P. (2007). *Social linguistics and literacies: Ideology in discourses*. London: Routledge. <https://doi.org/10.4324/9780203944806>.
- Gersten, R., Taylor, M. J., Keys, T. D., Rolhus, E. & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches*. (REL 2014-010). Washington, DC: US Department of education, Institute of Education Sciences, National Center for education evaluation and regional assistance, regional educational laboratory southeast. Available at [https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL\\_2014010.pdf](https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2014010.pdf). Accessed 20 March, 2019.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 38–52). London: Routledge.
- Gibbs, G., & Coffey, M. (2016). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education*, 5(1), 87–100.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74.
- Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Stewart, J., Tilghman, S. M., & Wood, W. B. (2004). Scientific teaching. *Science*, 304(5670), 521–522.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). New York: WH Freeman and Co.
- Hatano, G., & Oura, Y. (2003). Commentary: Reconceptualizing school learning using insight from expertise research. *Educational Researcher*, 32(8), 26–29.
- Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics - Physics Education Research*, 3(2), 020102.
- 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.
- Henriques, L. (1998). Maximizing the impact of your inservice: Designing the inservice and selecting participants (pp. 1–16). Paper presented at the Annual Meeting of the Association for the Education of Teachers of Science, Minneapolis.



- Ho, A., Watkins, D., & Kelly, M. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. *Higher Education*, 42(2), 143–169.
- Kahn, J., & Pred, R. (2001). Evaluation of a faculty development model for technology use in higher education for late adopters. *Computers in the Schools*, 18(4), 127–153.
- Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, 4(4), 298–310.
- Kreber, C. (2002). Teaching excellence, teaching expertise and the scholarship of teaching. *Innovative Higher Education*, 27, 5–23.
- Lawson, A., Benford, R., Bloom, I., Carlson, M., Falconer, K., Hestenes, D., Judson, E., Piburn, M., Sawada, D., Turley, J., & Wyckoff, S. (2002). Evaluating college science and mathematics instruction. *Evaluating College Science and Mathematics Instruction*, 31(6), 388–393.
- Loucks-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (2009). *Designing professional development for teachers of science and mathematics*. Thousand Oaks: Corwin Press.
- MacIsaac, D., & Falconer, K. (2002). Reforming physics instruction via RTOP. *The Physics Teacher*, 40(8), 479.
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Netherlands: Springer.
- Major, C. H., & Palmer, B. (2006). Reshaping teaching and learning: The transformation of faculty pedagogical content knowledge. *Higher Education*, 51(4), 619–647.
- Manduca, C. A. (2017). Surveying the landscape of professional development research: Suggestions for new perspectives in design and research. *Journal of Geoscience Education*, 65(4), 416–422.
- Manduca, C. A., Mogk, D. W., Tewksbury, B., Macdonald, R. H., Fox, S. P., Iverson, E. R., Kirk, K., McDaris, J., Ormand, C., & Bruckner, M. (2010). On the cutting edge: Teaching help for geoscience faculty. *Science*, 327(5969), 1095–1096.
- Manduca, C. A., Iverson, E. R., Luxenberg, M., Macdonald, R. H., McConnell, D. A., Mogk, D. W., & Tewksbury, B. J. (2017). Improving undergraduate STEM education: The efficacy of discipline-based professional development. *Science Advances*, 3(2), e1600193. <https://doi.org/10.1126/sciadv.1600193>.
- Marbach-Ad, G., Egan, L. C., & Thompson, K. V. (2015). A disciplinary teaching and learning center: Applying pedagogical content knowledge to faculty development. In G. Weaver (Ed.), *Transforming Institutions: 21st century undergraduate STEM education* (pp. 297–311). West Lafayette: Purdue University Press.
- Marshall, J. C., Smart, J., Lotter, C., & Sirbu, C. (2011). Comparative analysis of two inquiry observational protocols: Striving to better understand the quality of teacher-facilitated inquiry-based instruction. *School Science and Mathematics*, 111(6), 306–315.
- Mazumdar, M., & Glassman, J. R. (2000). Categorizing a prognostic variable: Review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Statistics in Medicine*, 19(1), 113–132.
- McConnell, D. A., Steer, D. N., & Owens, K. D. (2003). Assessment and active learning strategies for introductory geology courses. *Journal of Geoscience Education*, 51(2), 205–216.
- McConnell, D. A., Chapman, L. A., Czajka, C. D., Jones, J. P., Ryker, K. D., & Wiggen, J. (2017). Instructional utility and learning efficacy of common active learning strategies. *Journal of Geoscience Education*, 65(4), 604–625.
- McShannon, J., & Hynes, P. (2005). Student achievement and retention: Can professional development programs help faculty GRASP it? *The Journal of Faculty Development*, 20(2), 87–93.
- Michael, J. (2006). Where's the evidence that active learning works? *AJP: Advances in Physiology Education*, 30(4), 159–167.
- Mulnix, A. B. (2016). STEM faculty as learners in pedagogical reform and the role of research articles as professional development opportunities. *CBE-Life Sciences Education*, 15(4), es8–es9.
- Multon, K. D. (2010). Interrater reliability. In D. M. Nougherty & B. Frey (Eds.), *Encyclopedia of research design* (pp. 626–628). Thousand Oaks: Sage.
- National Research Council. (1996). *From analysis to action: Undergraduate education in science, mathematics, engineering, and technology*. Washington, D.C.: The National Academies Press. <https://doi.org/10.17226/9128>.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, D.C.: The National Academies Press. <https://doi.org/10.17226/9853>.
- National Research Council. (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, D.C.: The National Academies Press. <https://doi.org/10.17226/10497>.
- National Research Council (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. (S. Singer, N. R. Nielsen, & H. A. Schweingruber, Eds.). Washington, D.C.: The National Academies Press. <https://doi.org/10.17226/13362>



- National Science Foundation Directorate for Education and Human Resources. (1996). *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology*. Washington, D.C.: National Science Foundation.
- Olson, S., & Riordan, D. G. (2012, February). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President. Executive Office of the President.
- Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284.
- Parker, L. C., Adedokun, O., & Weaver, G. C. (2016). Culture, policy and resources: Barriers reported by faculty implementing course reform. In G. C. Weaver, D. Wilella, A. Burgess, & L. Slakey (Eds.), *Transforming institutions: 21st century undergraduate STEM education* (pp. 370–380). West Lafayette: Purdue University Press.
- Pelch, M. A., & McConnell, D. A. (2016). Challenging instructors to change: A mixed methods investigation on the effects of material development on the pedagogical beliefs of geoscience instructors. *International Journal of STEM Education*, 3(5), 1–18.
- Perkins, D. N., & Salomon, G. (1992). Transfer of learning. *International encyclopedia of education*, 2, 6452–6457.
- Pfund, C., Miller, S., Brenner, K., Bruns, P., Chang, A., Ebert-May, D., Fagen, A. P., Gentile, J., Gossens, S., Khan, I. M., Labov, J. B., Pribbenow, C. M., Susman, M., Tong, L., Wright, R., Yuan, R. T., Wood, W. B., & Handelsman, J. (2009). Professional development. Summer institute to improve university science teaching. *Science*, 324(5926), 470–471.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
- Postareff, L., Lindblom-Ylänne, S., & Nevgi, A. (2007). The effect of pedagogical training on teaching in higher education. *Teaching and Teacher Education*, 23(5), 557–571.
- Rushin, J. W., De Saix, J., Lumsden, A., Streubel, D. P., Summers, G., & Bemson, C. (1997). Graduate teaching assistant training: A basis for improvement of college biology teaching & faculty development? *The American Biology Teacher*, 59(2), 86–90.
- 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(6), 245–253.
- Schrump, L., Burbank, M. D., Engle, J., Chambers, J. A., & Glassett, K. F. (2005). Post-secondary educators' professional development: Investigation of an online approach to enhancing teaching and learning. *The Internet and Higher Education*, 8(4), 279–289.
- Shadle, S. E., Marker, A., & Earl, B. (2017). Faculty drivers and barriers: Laying the groundwork for undergraduate STEM education reform in academic departments. *International Journal of STEM Education*, 4(1), 8.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Sorcinelli, M. D., Austin, A. E., & Eddy, P. L. (2006). *Creating the future of faculty development: Learning from the past, understanding the present*. Bolton: Anker.
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagen, M. K., Edson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., Rodela, T. M., Sanders, E. R., Schimpf, N. G., Schultz, P. M., Smith, M. K., Stetzer, M., Van Valkenburgh, B., Vinson, E., Weir, L. K., Wendel, P. J., Wheeler, L. B., & Young, A. M. (2018). Anatomy of STEM teaching in north American universities. *Science*, 359(6383), 1468–1470.
- Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review*, 5(1), 25–49.
- Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147–176). Albany: State University of New York Press.
- Sunal, D. W., Hodges, J., Sunal, C. S., Whitaker, K. W., Freeman, L. M., Edwards, L., Johnston, R. A., & Odell, M. (2001). Teaching science in higher education: Faculty professional development and barriers to change. *School Science and Mathematics*, 101(5), 246–257.
- Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37(9), 963–980.

- Teasdale, R., Viskupic, K., Bartley, J. K., McConnell, D., Manduca, C., Bruckner, M., Farthing, D., & Iverson, E. (2017). A multidimensional assessment of reformed teaching practice in geoscience classrooms. *Geosphere*, 13(2), 608–627. <https://doi.org/10.1130/GES01479.1>.
- Van der Heijden, B. I. (2002). Individual career initiatives and their influence upon professional expertise development throughout the career. *International Journal of Training and Development*, 6(2), 54–79.
- Walczyk, J. J., & Ramsey, L. L. (2003). Use of learner-centered instruction in college science and mathematics classrooms. *Journal of Research in Science Teaching*, 40(6), 566–584.
- Walczyk, J. J., Ramsey, L. L., & Zha, P. (2007). Obstacles to instructional innovation according to college science and mathematics faculty. *Journal of Research in Science Teaching*, 44(1), 85–106.
- Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research*, 79(2), 702–739.
- Wenger, E. (2000). Communities of practice and social learning systems. *Organization*, 7(2), 225–246.
- Williams, B. A., Mandrekar, J., Mandrekar, S., Cha, S., & Furth, A. (2006). *Finding optimal cutpoints for continuous covariates with binary and time-to-event outcomes*. Rochester: Department of Health Sciences Research, Mayo Clinic.
- Wilson, S. M. (2013). Professional development for science teachers. *Science*, 340(6130), 310–313.
- Yoon, K.S., Duncan, T., Lee, S.W-Y., Scarloss, B., & Shapley, K.L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. (REL 2007-033). Washington, DC: US Department of education, Institute of Education Sciences, National Center for education evaluation and regional assistance, regional educational laboratory southwest. Available at [https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel\\_2007033.pdf](https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033.pdf). Accessed 20 March, 2019.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

**Karen Viskupic<sup>1</sup> · Katherine Ryker<sup>2</sup> · Rachel Teasdale<sup>3</sup> · Cathryn Manduca<sup>4</sup> · Ellen Iverson<sup>4</sup> · Dori Farthing<sup>5</sup> · Monica Z. Bruckner<sup>4</sup> · Rory McFadden<sup>6</sup>**

<sup>1</sup> Department of Geosciences, Boise State University, 1910 University Drive, Boise, ID 83725, USA

<sup>2</sup> School of the Earth, Ocean and Environment, University of South Carolina, 701 Sumter Street, EWS 307, Columbia, SC 29208, USA

<sup>3</sup> Department of Geological and Environmental Sciences, California State University, Chico, 400 W 1st Street, Chico, CA 95929, USA

<sup>4</sup> Science Education Resource Center, Carleton College, One North College Street, Northfield, MN 55057, USA

<sup>5</sup> Department of Geological Sciences, State University of New York at Geneseo, 1 College Circle, Geneseo, New York 14454, USA

<sup>6</sup> Department of Geology, Gustavus Adolphus College, 800 W. College Avenue, St. Peter, MN 56082, USA