

RESEARCH

Open Access



Challenging instructors to change: a mixed methods investigation on the effects of material development on the pedagogical beliefs of geoscience instructors

Michael A. Pelch* and David A. McConnell

Abstract

Background: STEM educational reform encourages a transition from instructor-centered passive learning environments to student-centered, active learning environments. Instructors adopting these changes incorporate research-validated teaching practices that improve student learning. Professional development that trains faculty to implement instructional reforms plays a key role in supporting this transition. The most effective professional development experiences are those that not only help an instructor redesign a course, but that also result in a permanent realignment of the teaching beliefs of participating instructors. Effective professional development features authentic, rigorous experiences of sufficient duration. We investigated changes in the teaching beliefs of college faculty resulting from their participation in the Interdisciplinary Teaching about the Earth for a Sustainable Future (InTeGrate) project that guided them in the development of reformed instructional materials for introductory college science courses. A convergent parallel mixed methods design was employed using the Teacher Belief Interview, the Beliefs About Reformed Science Teaching and Learning survey, and participants' reflections on their experience to characterize pedagogical beliefs at different stages of their professional development.

Results: Qualitative and quantitative data show a congruent change toward reformed pedagogical beliefs for the majority of participants. The majority of participants' Teacher Belief Interview (TBI) scores improved toward more student-centered pedagogical beliefs. Instructors who began with the most traditional pedagogical beliefs showed the greatest gains. Interview data and participants' reflections aligned with the characteristics of effective professional development. Merged results suggest that the most significant changes occurred in areas strongly influenced by situational classroom factors.

Conclusions: The process of materials development employed in the InTeGrate project is comprised of rigorous, authentic, and collaborative experiences continued over a sufficient duration. Incorporating these characteristics in to a professional development program on a variety of scales can help promote the long-term adoption of reformed teaching strategies. Collaboration among geoscience professionals was one of the predominant drivers for change. Consequently, this research provides insight for the development of future professional development opportunities seeking to better prepare instructors to implement reformed instructional strategies in their classrooms.

Keywords: Professional development, Mixed methods, Rasch analysis, InTeGrate, Teacher Belief Interview, Geoscience

* Correspondence: mapelch@ncsu.edu
North Carolina State University, 2800 Faucette Drive, Jordan Hall Campus,
Box 8208, Raleigh, NC 27695, USA

Background

A principal goal of STEM educational reform is to encourage a shift from instructor-centered classrooms where students are largely passive to student-centered environments where learning is an active process (Singer et al. 2012; Singer and Smith 2013). This transition has been supported by the development of research-validated teaching strategies that have been shown to improve student learning (Freeman et al. 2014; Handelsman et al. 2004; Kober 2015; Singer and Smith 2013) and reduce the achievement gap among student populations (Eddy and Hogan 2014; Haak et al. 2011). The effective implementation of research-validated instructional strategies will be referred to as reformed instruction (see MacIsaac and Falconer 2002). Despite the evidence in favor of the adoption of reformed teaching methods, many instructors in K-12 and higher education institutions have been reluctant to adopt these new strategies (Barak and Shakhman 2008; Henderson and Dancy 2007) and participation in professional development may be insufficient to promote the transition of these alternative approaches to instruction (e.g., Ebert-May et al. 2011). University faculty members report several barriers to instructional reform including limited training, insufficient time, and lack of instructional and peer support (Dancy and Henderson 2010; Fairweather 2010; Henderson and Dancy 2007; Sunal et al. 2001; Wieman et al. 2010).

Classroom change requires that teachers reconsider how they conceptualize the learning environment (Luft and Roehrig 2007). Instructor professional development has become the major focus of many systematic reform initiatives (Corcoran 1995; Corcoran et al. 1998; Garet 2001; Guskey 2002). Much of the research on this subject has occurred in K-12 or pre-service settings and examined instructors' beliefs regarding their roles in the classroom, how students learn best, and the most effective types of student-instructor interactions (Barak and Shakhman 2008; Fang 1996; Fishbein and Ajzen 1975; Hake 1998; Jones and Carter 2007; Kagan 1992; Luft and Roehrig 2007; Richardson 1996). Beliefs affect action (Guskey 1986; Hashweh 1996; Kang and Wallace 2005) and any steps seeking to promote lasting classroom change must consider a teacher's pedagogical beliefs (Barak and Shakhman 2008; Guskey 1986; Luft and Roehrig 2007). Successfully shifting the pedagogical beliefs of instructors toward a reformed mindset is essential if professional development is to positively influence teaching in STEM classrooms (Keys and Bryan 2001).

Professional development is intended to be a primary driver for the adoption of research-validated teaching strategies (Henderson et al. 2011; Wieman et al. 2010). Instructor professional development is a complex process which requires cognitive and emotional involvement of teachers both individually and collectively

(Avalos 2011). Professional development has been defined as either formal or informal learning opportunities designed to enhance teachers' professional competence, including knowledge, beliefs, motivations, and self-regulatory skills (Richter et al. 2011; Veenman 1984). An instructor's preference for either formal or informal professional development is not static (Richter et al. 2011) and may vary throughout their careers (Dall'Alba and Sandberg 2006; Gregorc 1973; Henderson et al. 2012; Huberman 1989; Richter et al. 2011; Sikes et al. 1985; Unruh and Turner 1970). The most common types of formal professional development opportunities often feature curricula designed to be large-scale national, state-wide, district-wide, or intra-institutional programs (e.g., Choy et al. (2006); Feiman-Nemser 2001). Examples of formal professional development are workshops, retreats, and courses where experts disseminate information (Feiman-Nemser 2001). Informal opportunities do not follow a specific curricula and are often smaller-scale opportunities that happen within a teachers own school setting (Desimone 2009). These opportunities tend to be less common than formal opportunities (Choy et al. 2006) and often consist of individual learning, collaborative and mentoring activities, and teacher networks (Desimone 2009; Parise and Spillane 2010). Informal professional development is often embedded in the classroom or school context allowing instructors to reflect on their experiences and share ideas among colleagues (Putnam and Borko 2000) and because of this is often a more authentic experience than formal professional development opportunities.

Professional development opportunities have had an inconsistent impact on classroom practice (Ebert-May et al. 2011; Feiman-Nemser 2001; Garet 2001; Henderson and Dancy 2007). Professional development has more effectively served as a medium for disseminating information on reformed teaching strategies (Henderson et al. 2012). Instructors may be more likely to use reformed teaching strategies if they have attended a professional development opportunity (Henderson et al. 2012), but the degree of change toward alternative instructional strategies is relatively modest. Ebert-May et al. (2011) studied the effectiveness of two national professional development programs in biology. They found that 89 % of participants reported implementing reformed, active learning strategies in their classrooms. However, videotaped observations of the participants revealed that 75 % of them still relied on traditional or instructor-centered methods (Ebert-May et al. 2011). Professional development programs may not provide instructors with the tools necessary to overcome situational barriers to adoption that are often unique to an instructor's teaching environment (Ebert-May et al. 2011; Garet 2001; Henderson et al. 2012; Henderson and Dancy 2007).

Situational factors can vary but commonly include class size and room layout, perceived student resistance to change, expectations of content coverage, and preparation time (Henderson and Dancy 2007). Implementing institutional changes that promote reformed instruction is culturally and logistically difficult due to the “norms” of practice and organizational structures at most universities (Hora 2012). Such reform is further complicated in STEM fields since different disciplines have their own set of standards and unique curricular requirements (Hora et al. 2013; Singer et al. 2012).

Effective professional development is characterized by some combination of the following factors: (1) it occurs on an extended timescale beyond a single workshop; (2) it provides guidance and feedback on the design of reformed lessons; (3) it places emphasis on collaboration; and (4) it is comprised of challenging and authentic experiences (Hawley and Valli 1999; Loucks-Horsley and Matsumoto 1999; Garet 2001; Henderson and Dancy, 2007; Penuel et al. 2007). Changing beliefs in adulthood is rare, and belief systems are often developed or realigned over prolonged time (Nespor 1987). Consequently, successful professional development needs to be of a sufficient duration, allowing for multiple cycles of presentation to provide participants with several opportunities to assimilate new knowledge and practices (Brown 2004) and to engage multiple cycles of feedback and reflection (Penuel et al. 2007). An in-depth analysis of an instructor’s own teaching strategies may provide the challenge to pre-existing teaching beliefs (Gess-Newsome et al. 2003) that can facilitate a “change in conversation” or gestalt shift that is critical to modifying those beliefs (Pajares 1992). Collaboration, especially among participants in the same field, has a positive impact on the effectiveness of a professional development program (D’Avanzo 2013; Garet 2001; Penuel et al. 2007). Collaboration during professional development is most successful when the participants, rather than outside experts, drive and determine their interaction (Hargreaves and Dawe 1990). For example, one-on-one collaboration between pre-service and in-service teachers has been shown to be an effective model of professional development by actively engaging teachers in their own research projects (Burbank and Kauchak 2003). Some forms of professional development may be too passive to adequately challenge participants to rigorously explore and reflect on new concepts (Penuel et al. 2007). Active participation during professional development is also important for the opportunity to be effective (D’Avanzo 2013; Loucks-Horsley and Matsumoto 1999; Penuel et al. 2007; Putnam and Borko 2000). For example, participants should be involved in the examination of effective teaching strategies and should be encouraged to investigate why and how a particular strategy could be effective in their classes. Finally, professional development

that models the authentic application of teaching strategies to lessons that are relevant to participants’ classes is more likely to support instructional change (Burbank and Kauchak 2003; D’Avanzo 2013; Smylie 1995).

We sought to examine changes in the pedagogical beliefs of instructors participating in a national curriculum development program that featured many of the elements of effective professional development discussed above. This study focuses on the degree to which participation in creating resources for the Interdisciplinary Teaching about the Earth for a Sustainable Future (InTeGrate) project (<http://serc.carleton.edu/integrate/>) contributed to changes in the way instructors thought about and described their teaching. The InTeGrate project is a multi-disciplinary effort to promote geoscience literacy in the context of societal issues, and to increase the numbers of geoscience-related undergraduate majors. While the overall InTeGrate project is multifaceted, this study focuses on college faculty involved in the development of a series of introductory geoscience modules. Each module was designed around a topic by teams of three geoscience faculty from a variety of institution types. The curriculum development aspects of the InTeGrate project were designed around the theory of participatory design where participants are a central component of the development process (Mao et al. 2005; Schuler and Namioka 1993). While certain aspects of the InTeGrate project mirror formal professional development (e.g., workshops for participants), other aspects more closely resemble those of informal professional development. For example, participants are required to pilot developed materials in their own classrooms. Additionally, participants spent a significant amount of time discussing pedagogy and instructional strategies with peers within their development teams. InTeGrate material development lasted approximately four semesters (sufficient duration) which included the design of original materials, pilot testing of those materials in classes, and subsequent modifications based on participant’s reflections and feedback from other team members and trained observers. Authors worked as a development team (collaborative) but also interacted with a larger project team that included material reviewers, classroom observers, assessment experts, and leadership and evaluation personnel. Activities designed by the teams were rooted in their own experiences and applicable to students in the courses they taught (authentic). Finally, module authors were provided multiple opportunities (e.g., workshops, webinars, technical reviews) to formally and informally place their work in context of discipline-based education research (Singer et al. 2012) and reflect on their experiences under the guidance of an extensive multi-part materials development rubric and expert reviews (rigorous). Workshops,

webinars, and technical reviews were conducted under the direction and guidance of the project's leadership and covered a variety of topics such as how to design measurable learning objectives, metacognition, and creating student-centered classroom activities.

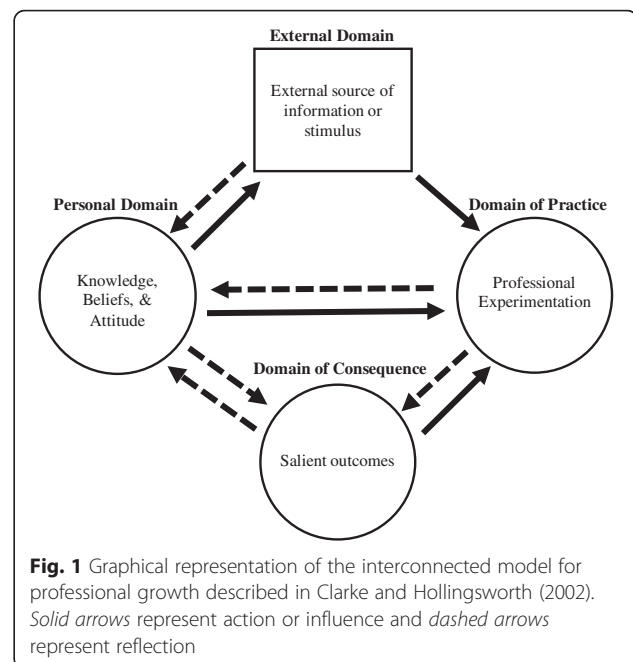
Theoretical framework

Instructors are attracted to professional development because they feel that it will expand their knowledge and skills (Guskey 2002). However, research on the long-term impacts of professional development suggest that this model is often not sufficient to promote long-term changes in the classroom (e.g., Ebert-May et al. 2011). In a modification to the traditional model of teacher change, Guskey (2002) proposed that instructors would modify their existing beliefs and attitudes if they experienced benefits from a new teaching strategy, such as increased student engagement or improvements in learning outcomes. Participants in the InTeGrate project had the opportunity to assess the effectiveness of new instructional strategies as they pilot the materials they developed in their classes. This practical experience could help secure the evolution of the instructors' pedagogical beliefs if it resulted in improvements in student engagement and learning.

Guskey's linear model for instructional change was expanded by Clarke and Hollingsworth (2002) who proposed a critical distinction between change sequences and growth networks. Clark and Hollingsworth noted that change in one domain can lead to change in another domain (change sequence), such as a teacher learning a new strategy at a workshop (external domain) and then experimenting with it in their classroom (domain of practice). However, they suggested that the concept of growth networks that involve numerous connections of enactment and reflection between multiple domains (Fig. 1) would be a more effective model for representing permanent changes in instructional practice and teaching beliefs. The goal of professional development programs may be to change teacher practice, but without fostering continued enactment and reflection across multiple domains, it may limit professional growth and may not result in the long-term change. The InTeGrate project sought to influence multiple domains within an interconnected model of professional growth.

This study assessed the effectiveness of the material development process employed in the InTeGrate project as a potential model for professional development capable of influencing instructors' pedagogical beliefs. We investigated the following questions:

1. How are the pedagogical beliefs of instructors altered through the scaffolded development of new instructional materials?



2. How did the material development process affect different aspects of an instructor's pedagogical beliefs?
3. To what extent did the material development process affect commonly cited barriers to the long-term adoption of research-validated teaching strategies?

Methods

We employed a convergent parallel mixed methods design for this study. Employing a mixed methods design to characterize beliefs is important because previous research has shown that multiple types of data are necessary to accurately describe beliefs (Pajares 1992; Richardson 1996). Mixed methods research combines qualitative and quantitative methods to reach a comprehensive understanding of the phenomenon under investigation, thus providing greater confidence in the study conclusions (Johnson and Onwuegbuzie 2004). A convergent parallel mixed methods study implements the quantitative and qualitative instruments during the same phase of the research and integrates the results (Creswell and Plano Clark 2011). We administered quantitative and qualitative instruments in a two-phase pre-post protocol to analyze how the development of course materials affected participant's pedagogical beliefs. Collecting both qualitative and quantitative data provides an opportunity to determine the degree to which an instructor's description of teaching beliefs (qualitative) converges or diverges with their conceptualizations of those same beliefs as revealed in self-report surveys (quantitative; Creswell and Plano Clark 2011). A panel

of STEM educators selected the study subjects who would participate in the InTeGrate project. All participants were post-secondary instructors from a variety of higher education institutions across the USA. Table 1 lists each participants' pseudonym and institutional type.

Qualitative instrument

Qualitative data were generated by application of the Teacher Belief Interview (TBI, Luft and Roehrig, 2007), a semi-structured instrument consisting of seven questions (Table 2) designed to assess how instructors characterize various aspects of science teaching and learning. We used the protocol outlined by Luft and Roehrig (2007) to transcribe and code the interview recordings. The coding protocol employs a thematic content analysis by guiding the researcher to assign individual responses to one of five categories: traditional, instructive, transitional, responsive, and reform-based instruction; therefore, any change, in even a single category, is likely practically significant. The traditional and instructive categories encompass participant's statements that often describe science as a series of rules and facts, and represent the most instructor-centered pedagogical beliefs. Responses that fall in the transitional category

Table 1 Study participants

| Pseudonym | Post-secondary teaching experience (years) | Institution type |
|-----------|--|------------------------------|
| Beth | 16 | Private, bachelors |
| Amy | 13 | Community college |
| Jackie | 13 | Public, masters |
| Rachel | 7 | Community college |
| Ellen | 10 | Private, bachelors |
| Karen | 9 | Public university, doctoral |
| Hilary | 13 | Community college |
| Mark | 13 | Community college |
| Sandy | 25 | Private, bachelors |
| Tyler | 5 | Public, bachelors |
| Mandy | 12 | Public, bachelors |
| Linda | 13 | Private, doctoral |
| John | 11 | Public university, doctoral |
| Becky | 24 | Private university, doctoral |
| Tammy | 11 | Public, masters |
| Lori | 2 | Private, bachelors |
| Nicole | 6 | Community college |
| Anne | 3 | Public university, doctoral |
| Adam | 8 | Community college |
| Lauren | 32 | Community college |
| Henry | 10 | Community college |

List of all 21 participants in the study, their post-secondary teaching experience in years, and the type of institution

Table 2 Seven semi-structured questions from the Teacher Belief Interview

| Number | TBI question |
|--------|---|
| 1 | How do you maximize student learning in your classroom? |
| 2 | How do you describe your role as a teacher? |
| 3 | How do you know when your students understand? |
| 4 | How do you know what to teach and what not to teach? |
| 5 | How do you decide when to move on to a new topic in your class? |
| 6 | How do you know when learning is occurring in your classroom? |
| 7 | How do your students learn science best? |

view science as consistent, connected, and objective, and represent pedagogies that typically include a mix of traditional and reformed strategies. The responsive and reform-based categories include answers dominated by student-centered pedagogical beliefs that characterize science as dynamic and integrated within a social and cultural construct. For comparative purposes, each of the five TBI codes can be assigned a numeric value of one to five (1 = traditional; 5 = reform-based) to generate a total TBI score for each participant. Scores can range from 7 to 35 representing the end-members for instructor-centered and student-centered pedagogies, respectively.

Pre-development interviews were conducted shortly after participants were selected to work on the project, but before they had any formal interaction with their teams or with project leadership. Post-development interviews occurred when participants completed the development, piloting, and final revision of their team's materials. Consequently, the time between participants pre- and post-development interview typically fell between 1 and 2 years. Interviews of participants took place in person or over the phone and consisted of two phases: (1) a preliminary round of informal questioning, and (2) the semi-structured format of the TBI. The semi-structured portion typically lasted between 20 and 45 min with an average time of 30 min. The preliminary phase began with a short description of the interview process. We asked interviewees to comment on their progress with the development of their instructional materials. Most discussions then addressed the nature of their current classes. We chose this initial questioning strategy to ease participants into the interview and to put them in the frame of mind to think about their teaching. Subsequently, we asked the seven TBI questions in the same order with the only exceptions representing occasions when an interviewee's responses dictated a deviation from the pre-determined sequence. The interviewer took the stance of a passive participant in the second phase of the interview to limit any

potential for biasing the interviewees' comments and to preclude leading the participants. Probing questions were asked after the instructor had introduced a topic and most of these supplementary questions guided interviewees to elaborate on the meaning of commonly used pedagogical terms such as "hands-on activities", "engaged", and "facilitating" or to describe what a certain activity would look like in their class. These clarifications were necessary to ensure that we could accurately code each TBI question. Pre- and post-interviews were conducted in the same way with the exception that the post-development interview also included the additional final question, "Do you think that working on the InTeGrate project has affected your teaching in any way?"

Qualitative reliability

Interviews in the current study were co-coded by two researchers that resulted in a Cohen's Kappa value of inter-rater agreement of 0.84 which qualifies as "good" co-coder agreement (Cicchetti and Sparrow, 1981). One coder was the primary investigator while the other was completely independent of the research and the study's participants. It was not possible to assess changes in teaching practice for comparison with changes in teaching beliefs due to the broad geographic distribution of participants and the lack of control observations. However, others have reported that the TBI shows a strong significant correlation with classroom observations of teaching practice (Ryker et al. 2013). Ryker et al. (2013) compared TBI scores and observations of geoscience instructors' classroom practices using the Reformed Teaching and Observational Protocol (RTOP). RTOP (Sawada et al., 2002) describes the degree of reformed teaching in a classroom based on five sub-scales (Lesson Design and Implementation, Content: Propositional Knowledge, Content: Procedural Knowledge, Student/Student relations, and the Instructor/Student relationship) and has well documented validity (Piburn et al. 2000; Sawada et al. 2002) and reliability (Amrein-Beardsley and Popp, 2012; Sawada et al. 2002). TBI scores had a strong positive correlation ($R^2 = 0.60$) with RTOP scores (Ryker et al. 2013). This suggests that teaching beliefs measured by the TBI have complementary classroom practices.

Quantitative instrument

The Beliefs about Reformed Science Teaching and Learning (BARSTL) survey (Sampson and Grooms 2013) measures an instructor's construct of reformed pedagogical beliefs. The BARSTL survey consists of 32 statements evenly divided among four sub-categories: (1) how people learn about science, (2) lesson design and implementation, (3) characteristics of the teacher and

learning environment, and, (4) nature of the science curriculum. Instructors respond to each statement with one of four options on a Likert scale: strongly agree (SA), agree (A), disagree (D), or strongly disagree (SD). Sixteen of the thirty-two items were traditionally phrased statements and were reverse scored (i.e., the more traditional the instructors teaching beliefs, the lower their score). Traditionally phrased items represent a post-positivist perspective of knowledge emphasizing the transmission of knowledge from instructor to student. The other sixteen statements were reform-phrased items that represented a constructivist view that the individual creates knowledge, and it can be unique to each student. The higher the total score on the BARSTL survey, the more aligned an instructor's pedagogical beliefs were with reform ideologies (Sampson and Grooms 2013). The BARSTL survey requires participants to respond to items that generate ordinal data. This form of data is commonly used in the social sciences to create overall scores by assigning the ordered responses a numerical value (Boone et al. 2011) and these scores are used for various parametric statistical tests. However, ordinal data often do not meet the fundamental assumptions necessary to apply parametric statistical tests (Bond and Fox 2007). Ordinal data are not linear, equal interval, and often are not normally distributed (Bond and Fox 2007; Boone et al. 2011). Additionally, ordinal data are not additive and simply changing the numeric values assigned to categories can affect the overall scores and interpretation of respondents' performances. We did not use raw scores from the BARSTL surveys that could have been obtained from assigning numeric values to the four-item Likert scale. Instead, we report BARSTL scores as scaled values, referred to from this point as *person measures* that were calculated using Rasch analysis. Winsteps Rasch analysis converted raw survey scores into person measures in an effort to better characterize results from the BARSTL survey. Person measures are expressed in logits (log-odds units), which are equal-interval units that can be applied to parametric statistical tests because they avoid many of the issues of the non-linearity of rating scales as well as the non-linearity of raw survey data (Boone et al. 2014).

George Rasch (1960) developed Rasch analysis which represents a one-way probabilistic approach based on Item Response Theory (IRT). The application of Rasch measurement in the social sciences has been most notably discussed in the 1967 Invitational ETS conference and in a wide-variety of subsequent publications (Andrich 1978; Boone et al., 2011; Choppin 1985; Libarkin and Anderson 2005; Linacre 1998, 2006, 2010; Siegel and Ranney, 2003; Smith 1991; Tong 2012; Wilson and Adams 1995; Wright and Stone 1979; Wright 1977, 1984). Rasch modeling is both norm-referenced

(comparing individuals to the group) and criterion-referenced (measured according to specific standards; Siegel and Ranney 2003). This is accomplished by considering the difficulty of the item on a continuum, the participant's response to that item (ability), and the probability that a participant will choose a response (Boone et al. 2014). Winsteps software (Linacre and Wright 2000) was used to compute person measures for the BARSTL survey results for all participants in this study, and to transform the original logit range to a more comparable, but still linear, scale ranging from 0 to 100. Just as with the raw survey data, a higher person measure indicates more reformed pedagogical beliefs while a low person measure indicates beliefs that are more traditional.

Wright Maps (also known as person-item maps) were constructed using logit values for person measures and items (Wilson and Draney 2000). Wright Maps are a method of displaying complex rating scale and test data (Boone et al. 2014) and place items and person measures on the same linear scale, much like a ruler. Wright maps can simultaneously display both items and person measures because Winstep Rasch analysis transforms both data types into logits (Boone et al. 2014; Linacre and Wright 2000).

BARSTL survey reliability and validity

In their validation of the BARSTL survey, Sampson and Grooms (2013) showed that the instrument was reliable ($\alpha = 0.77$, $p = 0.001$). Item review panels, sub-scale correlations, and a confirmatory factor analysis suggested that it had reasonable content and construct validity (Sampson and Grooms, 2013). In addition to calculating standard reliability values, Winsteps software allows the researcher to determine both the person and item reliabilities using the scaled logit values as opposed to the non-linear values commonly used in a Cronbach alpha calculation. Any measurement of reliability is limited to considering how items are related to one another (Cortina, 1993; Cronbach, 1951). Rasch analysis simultaneously considers the reliability of items and an individual's responses to those items to generate a reliability value for persons (similar to the Cronbach alpha) and a value for item reliability (similar to a Cronbach alpha but pertains to the inter-item relationship). Winsteps Rasch analysis yielded acceptable reliabilities and separation indices using participants' responses to the BARSTL survey for persons and items (Table 3) indicating that the instrument displays good internal reliability. Separation indices greater than one indicate that the sample size is sufficient to locate items along the trait of interest (Linacre and Wright, 2000; Table 3); reformed pedagogical beliefs. Winsteps Rasch analysis also allows the researcher to quantify pre- and post-survey bias by

Table 3 BARSTL person and item reliability values

| | Pre-development | | Post-development | |
|------------------|-----------------|------|------------------|------|
| | Person | Item | Person | Item |
| Separation index | 1.90 | 3.11 | 5.03 | 3.46 |
| Reliability | 0.78 | 0.91 | 0.86 | 0.92 |

Values of reliability and separation indices calculated using Winsteps software for the pre- and post-development BARSTL surveys

calculating differential item function (DIF) values for all survey items. DIF is a measure of how consistently an item or survey measures a group of respondents or if a particular group of respondents preferentially performs better or worse on a particular item (Boone et al., 2011). None of the thirty-two items in the BARSTL survey display significant DIF values ($p < 0.05$). Reliability values, separation indices, and DIF values calculated in Winsteps suggest that the BARSTL survey has reasonable construct validity as a measurement instrument (Boone et al., 2011; Linacre and Wright, 2000; Linacre, 2010).

Results

Teacher belief interview results

A majority (71 %) of the participants show gains on their TBI scores from pre- to post-development interviews (Fig. 2). TBI scores are approximately normally distributed from the both phases of the interviews based on visual analysis of Q-Q plots. A pre-development kurtosis value of 0.548 (standard error of 0.972) and a skewness value -0.999 (standard error of 0.501); and a post-development kurtosis value of -0.545 (standard error of 0.972) and skewness value of -0.415 (standard error of 0.501) also support the assumption of a normal distribution. No outlier values were detected using the outlier labeling rule and g-factor described in Hoaglin and Iglewicz (1987). Pre-development TBI scores range from 17 to 32 with a mean score of 25.8. Post-development TBI scores range from 23 to 32 and have a mean score of 28.2. There is a statistically significant ($t(20) = 3.43$, $p = 0.003$, $d = 0.70$) change toward more student-centered responses with a moderate to high effect size. Participants with the lowest initial TBI score exhibited the greatest normalized gains in the post-development TBI score (Fig. 3). Fifteen of the twenty-one participants improved their TBI scores representing a shift toward more student-centered pedagogical beliefs, five instructors showed no change in their beliefs, and one instructor's TBI score decreased in their post-development interview. Responses to four TBI questions (How do you maximize learning in your class? How do you describe your role as a teacher? How do you know when learning is occurring in your classroom? How do your students learn science best?) do not show significant shifts between the pre- and post-development interviews (Fig. 4). Responses to the remaining three questions (How do

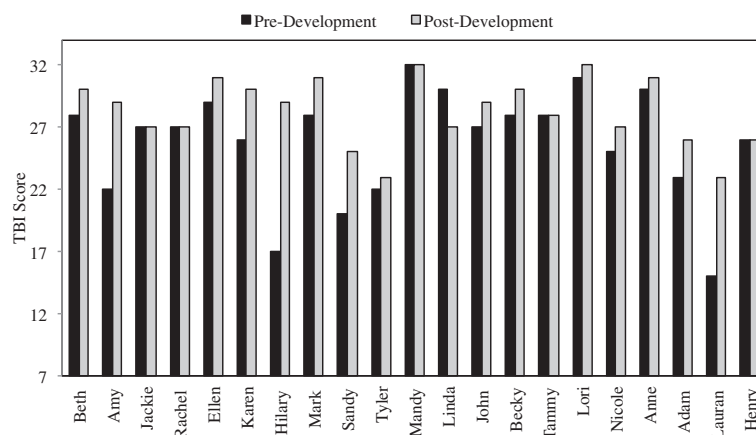


Fig. 2 Graph displaying participants' pre- and post-development TBI scores

you know when to move on to a new topic in class? How do you decide what to teach and what not to teach? How do you know when your students understand a concept in class?) showed the greatest divergence among the participants and all demonstrate a significant shift toward more student-centered codes for the post-development interviews (Fig. 4).

Description of variability in TBI responses

When asked how they maximized learning in their classrooms, responses ranged from transitional where Sandy stated: "I try to make them feel very open and I hope to convey to them that I welcome questioning and commenting at any time..." to more student-centered responses where Anne noted: "I think setting up the environment where they can take responsibility for their own learning and really process things on their own..." Sandy focused on creating positive environment (transitional) while Anne focused on creating an environment where students can take charge on their own learning (reformed).

When questioned about how they view their roles as teachers, some participants described a transitional role, for example, Amy reported "I've got to make sure I tell them that I struggled with certain things when I was a student...I share things like that with them so they realize that they are at a normal place..." Nicole's response provides a more reformed, student-centered, conceptualization of the role of a teacher "I customize my classes to what I think students will most relate to. As long as they met the student learning outcomes I'll still customize it so that I feel that they can relate to the material and carry it kind of away and use it in their real life".

The question "How do you know when learning is occurring in your classroom?" seeks to clarify an instructors conceptualization of assessment (Luft and Roehrig,

2007). When discussing how to assess learning, Hilary specified that she emphasized "The write-ups for the labs that they turn in" and with "bigger assessments [exams] at the end of three or four chapters..." This is an instructor-centered response where she primarily focuses on formal summative assessments. Conversely, Mandy stated "[In class] when students are starting to ask the level of detail of questions showing that they are really noticing the nuances of the data..." This description is more student-centered, focuses on the student's and their knowledge, and demonstrates that she is looking for students to initiate a significant interaction with either her or possibly asking these questions with their peers.

Participants responses to "How do your students learn science best?" were mostly student-centered. However, the responses ranged from transitional, such as Tyler who stated that his students learn science best "by actually working problems" and "if they sort of get the

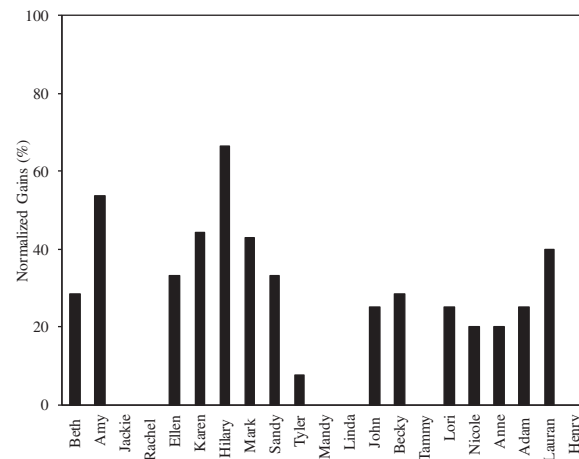
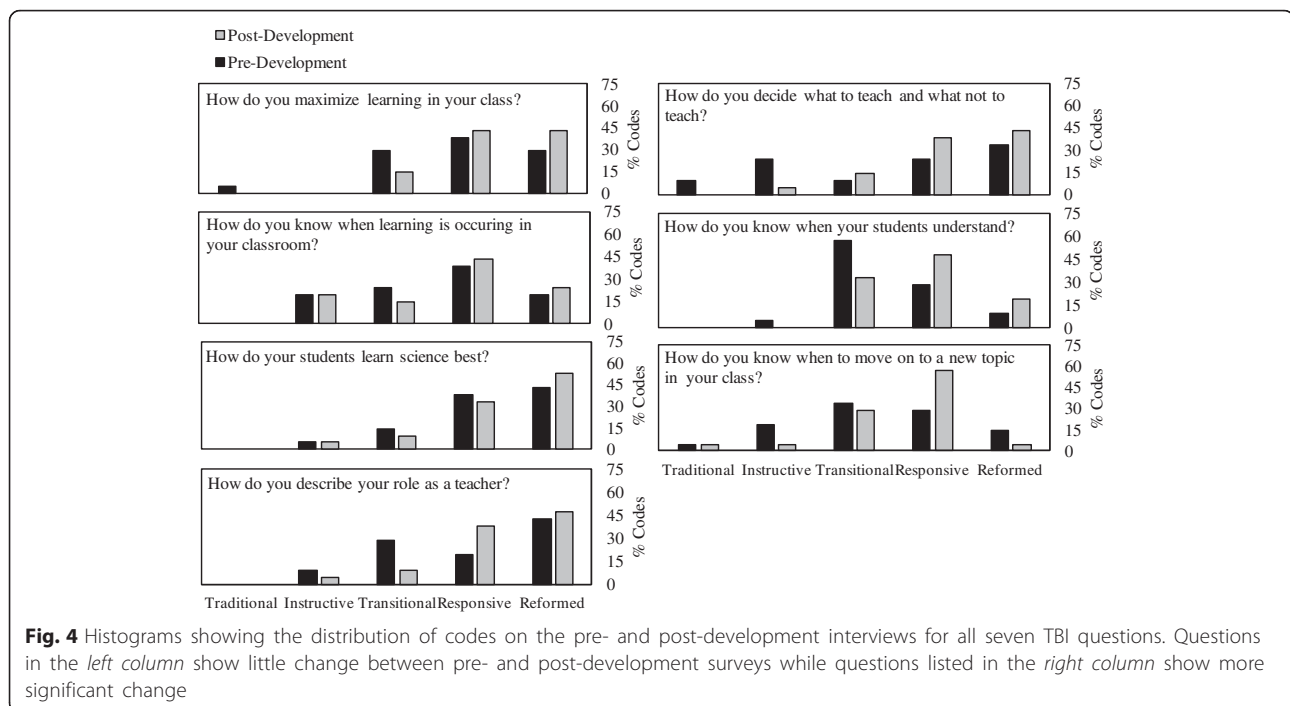


Fig. 3 Normalized gains for the pre- to post-development interviews



chance to apply what they have learned” to more student-centered responses from Jackie asserted that her students learn best “when they are doing something that has no right or wrong answer. When things that they do have multiple answers, you [students] can ask multiple questions” and she elaborates further by stating that when “The answer is not the end. An answer can be questioned...” is an experience when where students often best learn science.

Responses to three questions (How do you know when to move on to a new topic in class? How do you decide what to teach and what not to teach? How do you know when your students understand a concept in class?) were the more likely to be coded as “traditional” in the pre-development interviews. When asked about when to move on to a new topic, the responses ranged from an instructor-centered response from Hilary during her pre-development interview that focused on schedule and time constraints “I really can’t spend more than the...on whatever topic we are scheduled before there is just so much material that needs to be covered”. Moving along the coding spectrum, we can examine a transitional response from Tyler’s pre-development interview that included the students in the decision about when to move on “I pause when they are producing some results...So there are sort of intermediate steps along the way that I can see if they identify this rock sample correctly...” Unlike Hilary’s response, Tyler based his decision to move on partly on the students. An example of a reformed, student-centered reply can be read in Ellen’s post-development interview response “In that case I might

run the lecture over into the next time or have a couple of slides, because that’s the point is to get it and the that obviously helps me next time to redesign the lecture”. Ellen is not only considering whether or not students “get” the information but also verbalized the option of revisiting concepts and redesigning future lessons based on her current students’ experiences with the content.

The responses to the question “How do you decide what to teach and what not to teach?” show a similar variability. For example, a traditional response from the pre-development interviews comes from Hilary who stated “Well of course time is always a constraint...and then if there is extra time then I have more that I can teach”. She also noted, “I have topics that I want to make sure that I cover over the course of the semester”. This indicated a traditional adherence to syllabus and time constraints. A transitional response, such as John’s, from his pre-development interview where he declared “I think of what skills and knowledge in geology contribute to understanding...that’s what focuses my teaching, because majors make up 90 % of my class”. John’s statement described a distinct shift away from instructor-centered decisions and situational factors to thinking more about his students when he chooses class topics. A reformed response comes from Anne’s post-development interview when she described something she had done recently in one of her courses. She said “I kind of modified the class the second time around to give them [students] a lot more chances to practice peer teaching so that they felt more comfortable before teaching their own classes”. In this response, Anne not only

considered the needs of her students, much like John, she also used that student-generated feedback to modify future iterations of her class.

Participant responses were again variable ranging from instructor-centered to student-centered when asked about assessing students' understanding of concepts. An instructor-centered response was evident in Sandy's pre-development description of this process "I guess, you know, I give them [students] quizzes, pop quizzes, but it is not so much to gauge what they understand as to keep them, you know, on track and constantly engaging in the material". Again, in a similar phrasing for instructor-centered responses to learning, she focused on formal assessment delivered by the instructor. Reformed responses to this questions are defined by an instructors ability to determine when students have shown the application of knowledge outside of class or in a novel setting (Luft and Roehrig, 2007). An example of this can be seen in Mark's post-development response "I know that they've [students] understood a concept if they can relate something in the real world, based on that concept..." Tammy's post-development response also places an emphasis on the application of knowledge outside of classroom constraints: "I guess when they tell me something that I don't know, or something that I haven't thought about or they have an inspiration or something that makes them want to go further than the original constraints of the assignment".

Responses from participants showing the highest gains

Instructors with the greatest normalized gains between pre- and post-development interviews display the most significant changes in their responses to interview questions. Amy, Karen, Hilary, Mark, and Lauren all show a 40 % or greater normalized gain between their pre- and post-development TBI interviews (Fig. 2). We can gain some insight into individual instructors changing beliefs by examining Amy and Hilary's responses to TBI questions. These two instructors exhibited the broadest change between the pre- and post-development interviews. For this study, a pedagogically significant change is one in which the participants overall code for a response shifts among the three broad categories of pedagogical beliefs (instructor-centered, transitional, and student-centered) defined in Luft and Roehrig (2007). An example of a significant shift would be if an instructor initially responded to a question in a manner that received a traditional code, while their post-interview response was transitional. Amy showed significant shifts toward more student-centered beliefs in her responses to several questions. Amy described her role as a teacher in the pre-development interview stating "I've got to make sure I tell them that I struggled with certain things when I was a student" and "I try to share

things with them so they realize that they are at a normal place". Her words illustrate transitional pedagogical beliefs by focusing on establishing a welcoming and positive classroom environment. Her response shifted from transitional to student-centered in the post-development interview where she described her role as "fostering questions, having them ask questions and be really curious about the world" and she elaborates on this point stating, "I want them to be curious and I want to foster that curiosity". Amy had instructor-centered ideas about how to decide when to move on in class during the pre-development interview. Initially she stated "I focus mostly on time constraints" and "I assign a schedule, it's not incredibly rigid, but it is structured". This instructor-centered focus became more student-centered in post-development because Amy incorporated feedback from students with the possibility of revisiting topics. For example, she stated "when they tell me they are still not understanding then I will build something into the future classes" and "If they [students] need an extra day to review or synthesize everything then I am okay with adding an extra day to do that". This student-centered theme also carried over in her response to assessing learning in the classroom. In the pre-development interview, she described the process of assessing learning very transitionally "I try to arrange the course material so that things sort of repeat themselves" and she does not continually assess students, "I don't necessarily assess students along the way..." During the post-development interview, Amy describes more student-centered practices stating "I listen to how they are communicating with their group mates" and she elaborates on this theme stating that she knew they were learning when "they were questioning each other and working through it".

Hilary also showed significant shifts toward more student-centered responses on several questions (Figs. 2 and 3). During the initial interview, Hilary described how she assessed student understanding, "There are always the assessments, and with assessments I do try to mix it up a little bit..." This suggests a focus on measures given by the instructor. She went on to say, "If they actually come to a...some kind of solution [correct] to this problem..." and here she was not only focusing on assessments but on the correct solutions to those assessments to gauge understanding. Together, these two responses align with a transitional pedagogical belief system. However, in the post-development interview Hilary's response shifted to a more student-centered view, stating "when they can come up with a very...an original idea or thought" and "when they [students] are getting the information and maybe making comments on how this is useful or important to them..." The question about determining what to teach in class received

the most traditional responses (Fig. 4). During her initial pre-development interview, Hilary responded in an instructor-centered way where she stated “time is always a constraint...I do have topics that I want to make sure I cover over the course of the semester. Her response on the post-development interview shifted toward student-centered where she stated that she chooses topics “that they can internalize and see the importance of those things, whether you are a geologist, a teacher, or construction worker things that are going to affect their everyday life”. This response showed a student-centered focus on science literacy and she elaborated further when asked why she chooses topics in this way “It’s not because they are students, but because they are humans”. When asked during the pre-development interview how she assesses learning in the class, Hilary stated “I kind of have these bigger assessments at the end of three or four chapters”. Initially, she has an instructor-centered focus on summative assessments given after she covered topics in class. This view changed during the post-development interview where she responded in a more student-centered manner by stating that she assessed learning by “the discussions really, again they [students] give me a very good idea of what’s going on during a week to week basis”. She was focusing on interactions among students to determine if learning occurred and she was continually doing this as opposed to focusing on summative assessments given after the fact.

BARSTL survey results

Figure 5 graphs person measures calculated by Winsteps Rasch analysis and Fig. 6 shows each participant’s normalized gains. BARSTL scores, as represented by person measures from the pre- and post-development phases, are normally distributed based on the visual analysis of

Q-Q plots. Pre-development kurtosis value of -0.325 (standard error of 0.972) with a skewness value of -0.157 (standard error of 0.501), and a post-development kurtosis value of 0.262 (standard error of 0.972) with a skewness value of 0.562 (standard error of 0.501) also support the assumption of a normal distribution. No outlier values were detected using the outlier labeling rule and g-factor described in Hoaglin and Iglewicz (1987). There was a statistically significant ($t(20) = 2.74$, $p = 0.013$, $d = 0.50$) improvement toward more reformed responses on the survey with a moderate effect size. Fifteen of the twenty-one participants’ BARSTL survey scores increased on the post-development survey. Pre-development person measures range from 47.17 to 64.21 with a mean of 56.70. Post-development person measures range from 47.39 to 69.89 with a mean of 59.13. In contrast with the TBI scores, participants with the lowest initial person measures do not show the most significant gains (Figs. 5 and 6). Many of the participants do show improvement, but there is little correlation between the degree of improvement and their pre-development survey score.

Wright Maps in Fig. 7 display the person measures of each participant and the item measures for the pre- and post-development surveys. The right side of each Wright map displays item measures and item position. Higher item measures (located toward the top of the map) represent the most difficult items and items at the bottom of the map represent the least difficult items. Only the higher scoring participants agreed or strongly agreed with difficult items while both high- and low-scoring participants agreed with less difficult items. The BARSTL survey contains equal numbers of reform-phrased and traditionally phrased items and the latter are scored in reverse. Therefore, participants who stated that they disagreed or strongly disagreed with traditionally phrased items

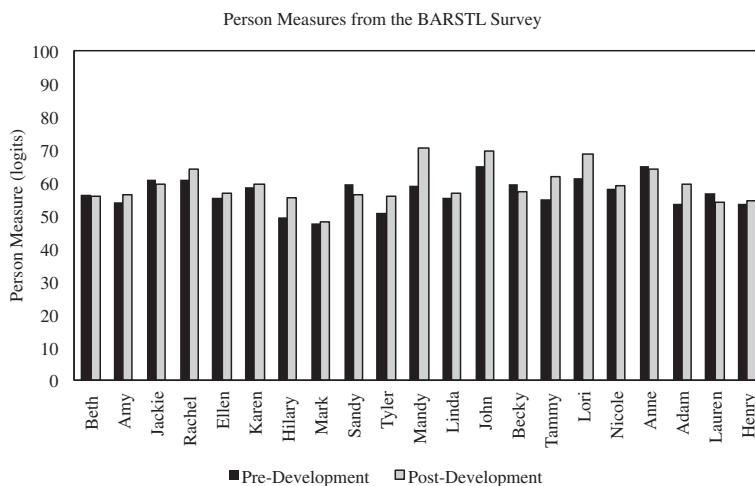
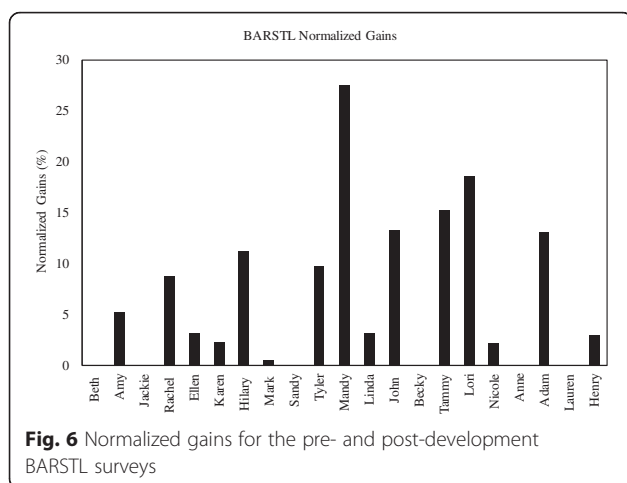
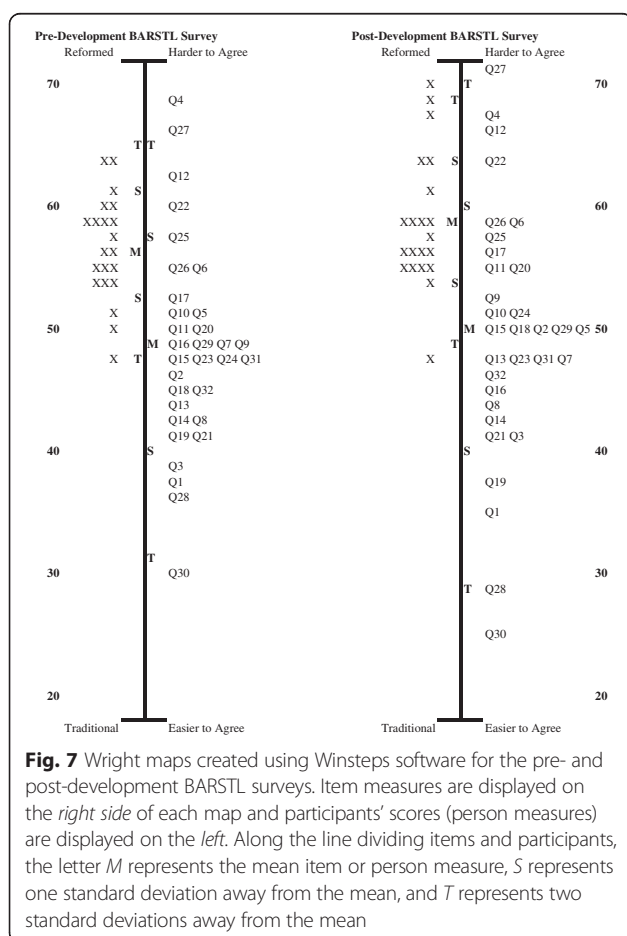


Fig. 5 Graph displaying participants’ pre- and post-development BARSTL survey measures



would obtain a higher total point value and be placed higher on the Wright map (see *X* values on left side of each Wright map in Fig. 6). The left side of the Wright Map places all of the participant's person measures (shown in logits) on the same linear scale as the items. More reformed instructors will have a higher score and be located toward the top and instructors with more



traditional teaching beliefs will have lower scores and be located toward the bottom. Participants person measures calculated from the post-development surveys show a significant shift toward more reformed pedagogical beliefs (positioned higher on scale along left side of figure in post-development result; Fig. 7).

The most difficult item on the pre-development survey was the traditionally phrased Q4: "Students are more likely to understand a scientific concept if the teacher explains the concept in a way that is clear and easy to understand". Q4 is located in the "How people learn about science" subcategory and represents a post-positivist view of knowledge. The least difficult item was the reformed-phrased Q30: "The science curriculum should help students develop the reasoning skills and habits of mind necessary to do science". Q30 remained the item with the most agreement on the post-development survey, while Q4 is replaced by another traditionally phrased item, Q27: "Students should know that scientific knowledge is discovered using the scientific method". Q27 is located in the "Nature of the Science Curriculum" subcategory, and similar to Q4, represents a strongly post-positivist view of knowledge in which the only way to create scientific knowledge is through a prescribed common method.

Analysis of Wright maps allows the identification of items that represent the statistically significant difference between groups by comparing the overlap of mean person measures from the pre- and post-development surveys (Boone et al., 2014). Responses to four statements comprise most of the difference between the pre- and post-development BARSTL surveys (Fig. 6). These four statements are: Q6 (traditionally phrased), "Learning is an orderly process; students learn by gradually accumulating more information about a topic over time"; Q17 (reform-phrased), "Students should do most of the talking in science classrooms"; Q25 (reform-phrased), "A good science curriculum should focus on only a few scientific concepts a year, but in great detail; and Q26 (traditionally phrased), "The science curriculum should focus on the basic facts and skill of science that students will need to know later". Items Q25 and Q26 explicitly deal with the design of the science curriculum, which is influenced by situational factors, and items Q6 and Q17 can be influenced by situational factors but are also shaped by other aspects of an instructor's pedagogical beliefs.

Discussion

Twenty of twenty-one participants showed increases between their pre- and post-development scores on either

the TBI and/or BARSTL survey (Table 4). Nearly half (10) of the participants showed congruent increases between their scores on each instrument and another four exhibited an increased score on one instrument but no change on the other (Table 4). Participants who entered the InTeGrate project with the most traditionally aligned beliefs recorded a shift toward more reformed pedagogical beliefs. Positive gains in scores were also recorded for participants who already had student-centered beliefs when they began the project. This suggests that the scaffolded material development process in the InTeGrate project is an effective method of faculty change for instructors who exhibit a range of pre-existing pedagogical beliefs. A key component of the InTeGrate project is the collaborative development of new instructional materials. The fact that the project supported collaboration of peers with similar teaching backgrounds and experiences sets InTeGrate apart from many traditional professional development experiences. Rogers (1995) posits that such teams are efficient at passing information among group members. This is important for the InTeGrate project because much of the

communication occurs at the team level (e.g., developing materials, discussing pilot experiences, and sharing information). Dissemination among peers is also important because team members started the project with different pedagogical beliefs. Therefore, participants with more reformed beliefs, or who had more experience using student-centered activities in their classes, were able to efficiently share their “lessons learned” from those experiences.

Six of the participants showed inconsistencies between their pre- and post-development phases of the BARSTL survey and TBI. Five participants had a decrease in BARSTL survey score but showed an increase in their TBI score. This discrepancy between the qualitative and quantitative data could be the result of limitations in the design of the BARSTL survey or an unavoidable limitation of Likert instruments designed to measure complex psychometric constructs as any form of personal belief is often difficult to measure (e.g., Fang 1996; Fishbein and Ajzen 1975; Jones and Carter, 2007; Rokeach, 1968). However, one of the reasons for employing Rasch analysis was to minimize the negative effects of quantitatively analyzing ordinal surveys. The validity and reliability of the BARSTL survey has not been rigorously confirmed outside of the initial design and validation study by Sampson and Grooms (2013). Consequently, differences in the results may reflect limitations in the BARSTL survey used in the context of this study. We interpret the more consistent changes in TBI scores to be due to the more effective manner in which pedagogical beliefs are characterized using qualitative methodology. While interviews can be considered a form of self-report, interview data tend to be a more authentic representation of real-world teaching practices (Luft and Roehrig 2007). Providing instructors with the opportunity to elaborate and describe their views on teaching has the potential to more accurately capture their beliefs (Ambrose and Clement 2004; Munby 1982). An example illustrating the importance of allowing instructors to elaborate on their practices was evident when asked, “How do you describe your role as a teacher?” Many instructors described their role as a facilitator. Facilitating learning is a characteristic of reformed instruction (Luft and Roehrig, 2007), however, when instructors were probed to elaborate on what they meant by the term “facilitator”, stark contrasts with that definition were often revealed. For example, during the pre-development interview Lauren described her role as “I give them the material or I set up the demo, or maybe I will have to do some lecturing as well”. This description is strongly traditional and not aligned with reformed pedagogical beliefs despite using a descriptor often associated with reformed teaching strategies. Instructor’s teaching practices and beliefs are better reflected by a continuum where the

Table 4 Merged results for study participants

| Participant | BARSTL | TBI |
|-------------|--------|-----|
| Beth | – | + |
| Amy | + | + |
| Jackie | – | |
| Rachel | + | |
| Ellen | + | + |
| Karen | + | + |
| Hilary | + | + |
| Mark | + | + |
| Sandy | – | + |
| Tyler | + | + |
| Mandy | + | |
| Linda | + | – |
| John | + | + |
| Becky | – | + |
| Tammy | + | |
| Lori | + | + |
| Nicole | + | + |
| Anne | – | + |
| Andrew | + | + |
| Lauren | – | + |
| Henry | + | |

Shifts toward more reformed pedagogical beliefs between pre- and post-development phases for the BARSTL survey and TBI are indicated with a positive sign; shifts toward more traditional pedagogical beliefs are noted with a minus sign; the absence of a symbol indicates no change between pre- and post-development phases. Participants highlighted in *italics* show positive improvements on both instruments

categories of traditional and reformed represent end-members rather than strict dichotomous groups (Smith et al. 2014). The nuances of a continuum may be more accurately represented by qualitative coding methods like those used for the TBI than that in a Likert-type instrument such as the BARSTL survey.

Qualitative and quantitative results both suggest that the greatest change in participants' pedagogical beliefs originate from questions and items partially or directly influenced by situational classroom factors. Situational factors are broadly defined as anything unique to an instructors' academic environment that can inhibit the adoption of new teaching strategies. Common examples include expectations of content coverage, lack of time, departmental teaching norms, student resistance, class size and room layout, and allotted class schedule (Henderson and Dancy, 2007). Comparing mean person measures and the positions of items on Wright maps (Fig. 7) revealed four items related to situational factors that comprised the majority of the statistical significance calculated between the pre- and post-development BARSTL surveys. Additionally, during many of the pre-development interviews, instructors often referred to dealing with situational factors, or cited such factors as having an impact on their teaching decisions when discussing TBI questions about moving on in class and choosing what to teach (Fig. 4). References to situational factors were less common in the post-development interviews during which instructors were more likely to discuss student-centered factors as influences toward choosing topics and deciding when to move on to new material.

The InTeGrate project challenges instructors to situate the design of course materials first within their own classrooms, and then secondly within the classrooms of their collaborators and others who may use the materials. This authentic and practical application may help explain why the pedagogical change shown in the qualitative and quantitative data focused on items and questions influenced by situational factors. Participants' responses to regularly scheduled written reflections completed throughout their development experience provide support for this interpretation. For example, Rachel stated that "Our course structures all being different has been challenging as we develop materials" and Linda (who worked with a different group of instructors), noted "Our different experiences in the classroom, and also teaching outside the classroom, helped us think about multiple experiences". The value of thinking about the implementation of materials in different academic environments was also summarized in Beth's reflection where she acknowledged that her collaborators "helped me see different types of assignments that I could write and ways I could think about my own classroom and

students". She went on to discuss the value of thinking about applying teaching materials to different environments "I like that I was forced to think about how an exercise would work in a classroom different from my own with students that were also different". Mandy reflected that her group had to struggle and adapt to "disparate approaches to assessment in our classes due to class size, student populations, etc". Developing materials for a diverse audience can also be beneficial after instructors have had the opportunity to pilot their activities in their classrooms. For example, Amy expressed the view that "I liked classroom testing, that this was done in three different institutions, and the results shared, and how this sparked revisions..." The diverse application of teaching strategies required by participants in the InTeGrate project may better equip instructors with the pedagogical content knowledge and experience necessary to overcome situational barriers. Additionally, information and experiences gained from the piloting of their new materials were readily transferred within each team because group members all had similar teaching backgrounds.

The InTeGrate project fulfilled many of the requirements of effective professional development (sufficient duration, collaboration, rigorous and authentic, and provides guidance) and the advantages of collaboration, authenticity, and rigor were confirmed through participants' statements during interviews and in their written reflections. Participants frequently discussed the value of collaboration stating "Hearing different ways of doing things; expanding the way I think about things", "The best part was sharing knowledge and classroom experiences with faculty from different institutions with diverse students", and "Working with my colleagues to develop the module I learned different approaches that I didn't know or had not tried before". Almost every participant referred to the advantages of working collaboratively and only one participant commented on the team-focused aspect of this project as a negative. Many short-term professional development opportunities are limited in scope and do not have the opportunity to effectively create an environment where instructors are supported to work intensively and to think about and reflect upon their teaching beliefs and practices. For example, Linda stated that "InTeGrate has been really helpful because most other workshops that I have been to are less involved and you don't get to spend much time with people. You just grab at stuff and leave. It is a lot less intensive". Beth presented another important perspective on the value of a rigorous and authentic professional development experience "Instead of saying, nope I am not interested in that activity or I don't know how to do that activity, the project certainly forced me out of my comfort zone and that was a good thing". The structured

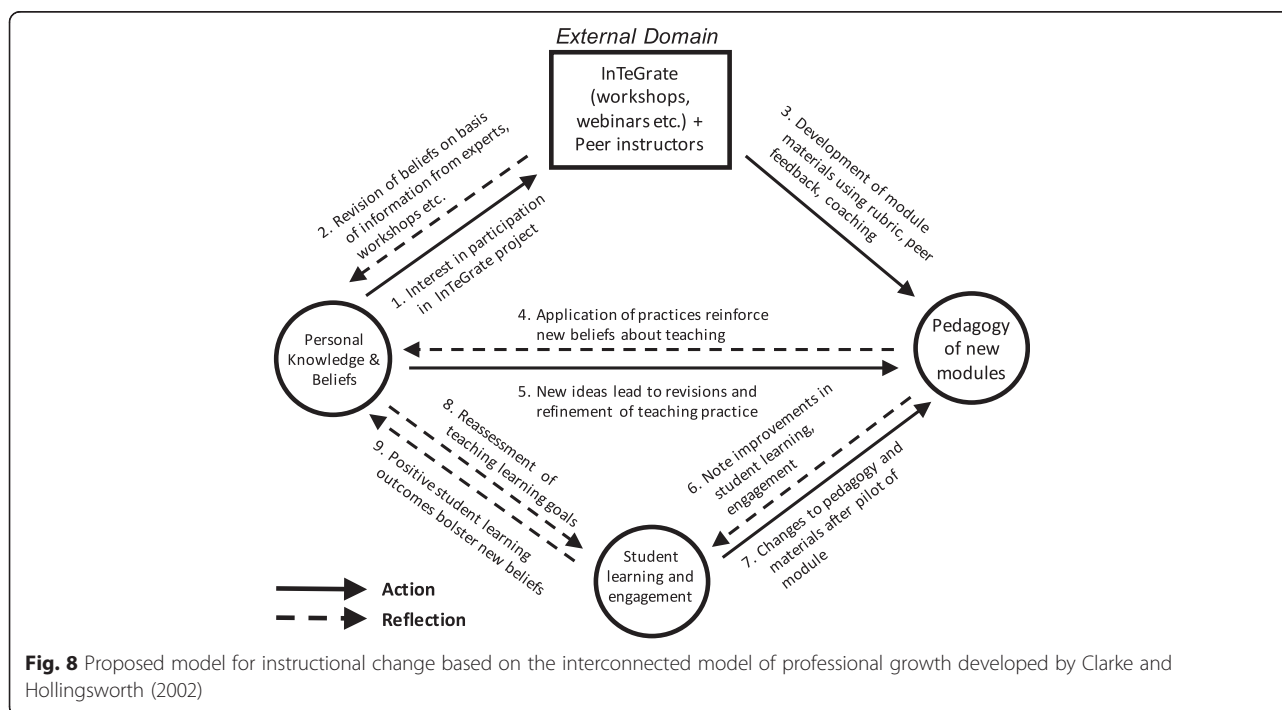
and supportive design of InTeGrate also facilitated multiple cycles of reflection and revision. This is a crucial aspect to the InTeGrate project because these processes drive instructional change within interconnected model (Clarke and Hollingsworth, 2002).

Data from this study allows us to propose a model for how the InTeGrate project promoted instructional change framed by the interconnected model of professional growth (Fig. 8; Clarke and Hollingsworth, 2002). Participants' external domain was being strongly influenced by key components of the InTeGrate project (e.g., workshops, peers; Fig. 8). Participants' desire and interest to work on InTeGrate project was initially driven by their pre-development pedagogical beliefs (Arrow #1, Fig. 8). InTeGrate participants were recruited to join a project seeking to create student-centered teaching materials. This characterization served to attract instructors with an interest in reformed instruction, and this explains why many of the instructors who applied already had transitional to student-centered pedagogical beliefs. Their initial participation in the project's workshops (both online and in person), and their interactions with team members, assessment experts, and project leaders represented the earliest opportunities for reflection and revision of their pedagogical beliefs (#2, Fig. 8). The InTeGrate project's rubric requirements and structured feedback from leadership and assessment experts influenced the development of their lessons (#3, Fig. 8). Once instructors met the project-wide and pedagogical requirements

outlined in the rubric, they piloted the new materials in their own classes. The experience of piloting the new materials not only reinforced and strengthened new-found pedagogical beliefs (#4, Fig. 8) in the teaching process, but also facilitated reflection leading to revision and refinement of activities and practices for future iterations of the lessons (#5, Fig. 8). Evidence of greater student engagement and student learning (#6, Fig. 8), as seen in daily formative assessment activities and summative exams, may also have contributed to the evolution of beliefs (#9, Fig. 8). Student reactions that showed greater engagement and/or success for specific types of activities undoubtedly also influenced decisions regarding revisions to the pedagogy (#7, Fig. 8). Similarly, the success of InTeGrate lessons that were specifically created to promote geoscience literacy in the context of societal issues may cause an instructor to reconsider their teaching and learning goals for other lessons (#8, Fig. 8). Different instructors may have been more influenced by some of these factors than others but the interconnected model provides us a means of illustrating the different drivers that lead to changes in teaching beliefs.

Study limitations

Results in this study are limited due to the non-random method participants that were chosen and the small sample size. The participants in the study submitted proposals to develop materials for the InTeGrate project,



and factors such as teaching experience and pedagogical content knowledge contributed to their selection for the project. All of the participants are post-secondary instructors with several years of teaching experience. Therefore, the results of this study may not be applicable to new instructors with little to no prior experience, or to instructors outside of higher education. While the study generated a sufficient quantity of qualitative information, the numbers of participants used in Winsteps Rasch analysis remains low. It is recommended that a range of 16 to 36 participants be used to achieve results applicable to ± 1 logits within a 95 % confidence interval (Wright and Stone, 1979). This minimum sample size is only applicable for exploratory purposes, such as was used in this study, and should not be applied to item design, evaluation, or high-stakes assessment (Wright and Stone, 1979). The distributions of person measures from this study were compared to a larger sample ($n = 103$) of BARSTL surveys completed by similar higher-education geoscience faculty. This larger sample of faculty was comprised of additional participants from the InTeGrate project working on either whole courses or upper-level classes, and from faculty not related to the InTeGrate project. The mean of this larger sample was similar to the means calculated in this study for the pre- and post-development surveys, and this increases the likelihood that the results of this study are representative of post-secondary geoscience instructors.

The BARSTL survey was originally designed and validated to measure the teaching beliefs of K-12 teachers. This study chose to employ the BARSTL survey because it was explicitly designed and validated around the construct of reformed pedagogical beliefs and thus should align with the TBI. All thirty-two items on the BARSTL survey were reviewed to ensure that they were applicable in some way to post-secondary instructors. After a sufficient amount of interview data was collected, the items in the BARSTL survey were re-evaluated to ensure that concepts instructors discussed in their interviews aligned with items from the survey. Preliminary explorations into the validity of the BARSTL survey revealed that several items show poor discrimination potentially attributed to ambiguous wording. Additionally, visual analysis of the Wright maps for the pre- and post-development surveys (Fig. 7) shows that many of the items group toward the mean and that several items plot on the same item measure. Items that plot on the same measure are theoretically redundant and could also contribute to a less desirable distribution of scores along the trait of interest (Boone et al., 2011). The narrow distribution of items revealed in the Rasch analysis could also be the likely cause for the correspondingly narrow range of person measures despite the broad range of pedagogical beliefs revealed in the participants' interviews.

Conclusions

The process of material development employed in the InTeGrate project incorporates many of the characteristics of effective professional development and influences multiple domains within the interconnected model of professional growth. Results suggest that incorporating several of the characteristics of effective professional development into a single opportunity can mitigate commonly cited barriers to the long-term adoption of reformed teaching strategies. The overarching design of the InTeGrate project created a scaffold structure where participants were encouraged to think differently about their teaching, teach differently, and to be continually reflective. Consequently, the InTeGrate project changed the way participating instructors thought about their teaching. Another powerful facet of the InTeGrate project was collaboration situated within the geoscience discipline and was exemplified in the following quote from one of the participants:

"It [InTeGrate] has helped because it is not just someone in an unrelated field talking about what they do, but it is colleagues in the geosciences teaching the same courses, using real tangible examples in different ways to create assignments and activities that you can use in your own classroom".

This model of professional development may be a more effective strategy for promoting the long-term adoption of reformed teaching strategies that promote student learning. We suggest that the project was successful in promoting instructional reform because it addressed the key domains (Personal, External, Practice, and Consequence) of an interconnected model of professional development. Creating well-supported, long-term professional development activities such as the InTeGrate project is beyond the capabilities or the needs of many organizations interested in faculty professional development. However, several of the characteristics of the project such as the authentic application of instructional strategies, collaboration, the application of a well-designed rubric, and structured feedback can be designed and implemented in a variety of scales. This research thus provides insight for the development of future professional development opportunities seeking to better prepare instructors to implement reformed instructional strategies in their classrooms.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MP was the principal investigator of this study. He carried out the research design, data analysis, data interpretation, review of the pertinent background literature, and wrote the majority of the manuscript. DMCC edited and

revised much of the manuscript. He also provided input and guidance on research design and the interpretation of these data. Both authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Science Foundation award 1125331. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation. We would like to thank colleagues throughout the InTeGrate project for their ideas and support for this research. Additionally, we would also like to thank Dr. Ellen Iverson for providing valuable feedback and insight into the revision of this manuscript and to Dr. Carol Baldassari for sharing data that supported the conclusions outlined in the manuscript. Lastly, we would like to thank Doug Czajka for helping to develop the theoretical model that supported the interpretation of these data.

Received: 11 February 2016 Accepted: 10 March 2016

Published online: 22 March 2016

References

- Ambrose, R., & Clement, L. (2004). Assessing prospective elementary school teachers' beliefs about mathematics and mathematics learning: rationale and development of a constructed-response. ... *Mathematics*, 104, 56–69.
- Amrein-Beardsley, A., & Popp, S. (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, 5–24.
- Andrich, D. (1978). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*.
- Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education*, 27(1), 10–20.
- Barak, M., & Shakhman, L. (2008). Reform-based science teaching : teachers' instructional practices and conceptions. *Eurasia Journal of Mathematics, Science & Technology Education*, 4(1), 11–20.
- Bond, T. G., & Fox, C. M. (2007). Applying the Rasch Model: fundamental measurement in the human sciences. *Journal of Educational Measurement*, 2, 288.
- Boone, W. J., Townsend, S. J., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: an exemplar utilizing STEBI self-efficacy data. *Science Education*, 95, 258–280.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.
- Brown, J. L. (2004). *Making the most of understanding by design*. Washington DC: Association for Supervision and Curriculum Development.
- Burbank, M. D., & Kauchak, D. (2003). An alternative model for professional development: investigations into effective collaboration. *Teaching and Teacher Education*, 19, 499–514.
- Choppin, B. H. L. (1985). Lessons for psychometrics from thermometry. *Evaluation in Education*, 9(1), 9–12.
- Choy, S. P., Chen, X., Bugarin, R., & National Center for Educational Statistics, & Institute of Education Sciences (U.S.) (2006). Teacher professional development in 1999–2000 what teachers, principals, and district staff report. *Statistical Analysis Report*.
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86, 127–137.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947–967.
- Corcoran, T. B. (1995). *Transforming professional development for teachers: a guide for state policy makers*. Washington D. C: National Governor's Association.
- Corcoran, T. B., Shields, P. M., & Zucker, A. A. (1998). *Evaluation of NSF's statewide systematic initiatives (SSI) program: the SSIs and professional development for teachers*. Menlo Park, CA: SRI International.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Applied Linguistics (Vol. 2nd).
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- D'Avanzo, C. (2013). Post-vision and change: do we know how to change? *CBE Life Sciences Education*, 12(3), 373–382.
- Dall'Alba, G., & Sandberg, J. (2006). Unveiling professional development: a critical review of stage models. *Review of Educational Research*, 76(3), 383–412.
- Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78(10), 1056.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: toward better. *Educational Researcher*, 38, 181–199.
- 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.
- Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: how and for whom does increasing course structure work? *Cell Biology Education*, 13(3), 453–468.
- Fairweather, J. (2010). *Linking evidence and promising practices in STEM undergraduate education*.
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, 38(1), 46–65.
- Feiman-Nemser, S. (2001). From preparation to practice: designing a continuum to strengthen and sustain teaching. *Teachers College Record*, 103, 1013–1055.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: an introduction to theory and research*. Reading, MA: Addison-Wesley Publishing.
- 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 of the United States of America*, 111(23), 8410–5.
- Garet, M. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945.
- Gess-Newsome, J., Southerland, S. A., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: the anatomy of change in college science teaching. *American Educational Research Journal*, 40(3), 731–767.
- Grogan, A. F. (1973). Developing plans for professional growth. *NASSP Bulletin*, 57(377), 1–8.
- Guskey, T. (1986). Staff development and the process of teacher change. *Educational Researcher*, 15(5), 5–12.
- Guskey, T. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, 8, 3.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science (New York, N.Y.)*, 332(6034), 1213–1216.
- 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 Teachers*, 66, 64–74.
- Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., Dehaan, R., Wood, W. B. (2004). Scientific teaching. *Science*, 304, 521–522.
- Hargreaves, A., & Dawe, R. (1990). Paths of professional development: contrived collegiality, collaborative culture, and the case of peer coaching. *Teaching and Teacher Education*.
- Hashweh, M. Z. (1996). Effects of science teachers' epistemological beliefs in teaching. *Journal of Research in Science Teaching*, 33(1), 47–63.
- Hawley, W. D., & Valli, L. (1999). *The essentials of effective professional development: a new consensus. Teaching as the learning profession: Handbook of policy and practice* (pp. 127–150).
- Henderson, C., & Dancy, M. (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.
- Henderson, C., Dancy, M., & Niewiadomska-Bugaj, M. (2012). Use of research-based instructional strategies in introductory physics: where do faculty leave the innovation-decision process? *Physical Review Special Topics - Physics Education Research*, 8(2), 020104.
- Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 82(400), 1147–1149.
- Hora, M. T. (2012). Organizational factors and instructional decision-making: a cognitive perspective. *The Review of Higher Education*, 35(2), 207–235. 1.
- Hora, M. T., Oleson, A., & Ferrare, J. J. (2013). Teaching dimensions observation protocol (TDOP) user's manual. *Tdop.Wceruw.Org*, 28

- Huberman, M (1989). The professional life cycle of teachers. *The Teachers College Record*, 91, 31–57.
- Johnson, RB, & Onwuegbuzie, AJ (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Jones, MG, & Carter, G (2007). Science teacher attitudes and beliefs. In *Handbook of research on science education* (pp. 1067–1104).
- Kagan, D (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62(2), 129–169.
- Kang, NH, & Wallace, CS (2005). Secondary science teachers' use of laboratory activities: linking epistemological beliefs, goals, and practices. *Science Education*, 89(1), 140–165.
- Keys, C, & Bryan, L (2001). Co-constructing inquiry-based science with teachers: essential research for lasting reform. *Journal of Research in Science Teaching*, 38(6), 631–645.
- Kober, N (2015). *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering*. National Academies Press.
- Libarkin, J, & Anderson, S (2005). Assessment of learning in entry-level geoscience courses: results from the Geoscience Concept Inventory. *Journal of Geoscience Education*, 394–401.
- Linacre, JM (1998). Detecting multidimensionality: which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266–283.
- Linacre, JM (2006). Rasch analysis of rank-ordered data. *Journal of Applied Measurement*, 7, 129–139.
- Linacre, JM (2010). Predicting responses from Rasch measures. *Journal of Applied Measurement*.
- Linacre, JM, & Wright, BD (2000). *Winsteps [computer software]*. Chicago: Mesa Press.
- Loucks-Horsley, S, & Matsumoto, C (1999). Research on professional development for teachers of mathematics and science: the state of the scene. *School Science and ...*, 99, 258–271.
- Luft, J, & Roehrig, G (2007). Capturing science teachers' epistemological beliefs: the development of the teacher beliefs interview. *Electronic Journal of Science ...*, 11, 2.
- MacIsaac, D, & Falconer, K (2002). Reforming physics instruction via RTOP. *The Physics Teacher*, 40(November), 479–485.
- Mao, JY, Vredenburg, K, Smith, PW, & Carey, T (2005). The state of user-centered design practice. *Communications of the ACM*, 48, 105–109.
- Munby, H (1982). The place of teachers' beliefs in research on teacher thinking and decision making, and an alternative methodology. *Instructional Science*, 11(3), 201–225.
- Nespor, J (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19, 317–328.
- Pajares, MF (1992). Teachers' beliefs and educational research: cleaning up a messy construct. *Review of Educational Research*, 62, 307–332.
- Parise, LM, & Spillane, JP (2010). Teacher learning and instructional change: how formal and on-the-job learning opportunities predict change in elementary school teachers' practice. *The Elementary School Journal*.
- Penuel, WR, Fishman, BJ, Yamaguchi, R, & Gallagher, LP (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921–958.
- Piburn, M, Sawada, D, & Turley, J (2000). Reformed teaching observation protocol (RTOP) reference manual. ... of Teachers, 1–41.
- Putnam, RT, & Borko, H (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*.
- Rasch, G (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.
- Richardson, V (1996). The role of attitudes and beliefs in learning to teach. In *The handbook of Research in Teacher Education* (pp. 102–119). New York: Macmillan.
- Richter, D, Kunter, M, Klusmann, U, Lüdtke, O, & Baumert, J (2011). Professional development across the teaching career: teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education*, 27, 116–126.
- Rokeach, M (1968). *Beliefs, attitudes, and values: a theory of organization and change*.
- Ryker, K, McConnell, DA, Bruckner, MA, & Manduca, CA (2013). Teaching is believing: pedagogical beliefs, practices, and professional development. In *2013 GSA Annual Meeting in Denver*.
- Sampson, V, & Grooms, J (2013). Development and initial validation of the beliefs about reformed science teaching and learning (BARSTL) questionnaire. *School Science and Mathematics*, 113, 3–15.
- Sawada, D, Piburn, MD, Judson, E, Turley, J, Falconer, K, Benford, R, & Bloom, I (2002). *Measuring reform practices in science and mathematics classrooms: the reformed teaching observation protocol*, 102.
- Schuler, D, & Namioka, A (1993). Participatory design: principles and practices. *System*, 319. <http://doi.org/10.1016/j.janclin.2007.12.004>.
- Siegel, MA, & Ranney, MA (2003). Developing the changes in attitude about the relevance of science (CARS) questionnaire and assessing two high school science classes. *Journal of Research in Science Teaching*, 40(8), 757–775.
- Sikes, PJ, Measor, L, & Woods, P (1985). *Teacher careers: crisis and continuity*.
- Singer, S, & Smith, KA (2013). Discipline-based education research: understanding and improving learning in undergraduate science and engineering. *Journal of Engineering Education*, 102(4), 468–471.
- Singer, SR, Nielsen, NR, & Schweingruber, HA (2012). *Understanding and improving learning in*.
- Smith, RM (1991). The distributional properties of Rasch item fit statistics. *Educational and Psychological Measurement*.
- Smith, MK, Vinson, EL, Smith, JA, Lewin, JD, & Stetzer, MR (2014). A campus-wide study of STEM courses: new perspectives on teaching practices and perceptions. *Cell Biology Education*, 13(4), 624–635.
- Smylie, MA (1995). Teacher learning in the workplace: implications for school reform. In *Professional development in education: new paradigms and practices* (pp. 92–113).
- Sunal, DW, Hodges, J, Sunal, CS, & Whitaker, KW (2001). *Teaching science in higher education: faculty professional development and b ...*, (May).
- Tong, V (2012). *Geoscience research and education*.
- Unruh, A, & Turner, H (1970). *Supervision of change and innovation*. Boston: Houghton Mifflin.
- Veenman, S (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54(2), 143–178.
- Wieman, C, Perkins, K, & Gilbert, S (2010). Transforming science education at large research universities: a case study in progress. *Change: The Magazine of Higher Learning*, 42(2), 6–14.
- Wilson, M, & Adams, RJ (1995). Rasch models for item bundles. *Psychometrika*, 60, 181–198.
- Wilson, M, & Draney, K (2002). A technique for setting standards and maintaining them over time. *Measurement and Multivariate Analysis*. 325–332.
- Wright, BD (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97–116.
- Wright, BD (1984). Despair and hope for educational measurement. *Contemporary Education Review*, 3(1), 281–288.
- Wright, BD, & Stone, MH (1979). *Best test design*. Chicago: Mesa Press.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com