Routledge
Taylor & Francis Group

Journal of Science Teacher Education

ISSN: 1046-560X (Print) 1573-1847 (Online) Journal homepage: http://www.tandfonline.com/loi/uste20

A Practice-Based Professional Development Program to Support Scientific Argumentation From Evidence in the Elementary Classroom

Evan J. Fishman, Hilda Borko, Jonathan Osborne, Florencia Gomez, Stephanie Rafanelli, Emily Reigh, Anita Tseng, Susan Million & Eric Berson

JOURNAL OF SCIENCE TEACHER EDUCATION
http://dx.doi.org/10.1080/1046560X.2017.1302727

Routledge
Taylor & Francis Group

A Practice-Based Professional Development Program to Support Scientific Argumentation From Evidence in the Elementary Classroom

Evan J. Fishman, Hilda Borko, Jonathan Osborne, Florencia Gomez, Stephanie Rafanelli, Emily Reigh, Anita Tseng, Susan Million, and Eric Berson

Stanford Graduate School of Education, Stanford University, Stanford, California, USA

ABSTRACT

Considerable evidence suggests that dialogical interaction in the classroom promotes students' scientific knowledge building and reasoning. Hence, scientific argumentation is recognized as a central component of the Next Generation Science Standards. A focus on argumentation, however, requires teachers to adopt instructional practices that facilitate this type of discourse. Therefore, the study reported here examined the impact of a practice-based professional development (PD) program on the scientific discourse practices of teachers and their students. Two cohorts of elementary school teachers attended different versions of the PD: 1 cohort attended the full program (institute, practicum, and follow-up sessions), whereas the other took part in the institute and follow-up days but did not take part in the practicum. We found that all teachers and their students, regardless of the cohort, made statistically significant improvements in their science discourse practices after attending the PD. An unexpected finding was that teachers who attended the full PD (with practicum) did not outperform the teachers who did not attend the practicum. However, students of teachers who did attend the practicum made improvements that approached statistical significance compared to students of teachers who did not attend the practicum. Thus, we report research that provides evidence of an effective model of practice-based PD that helps teachers to address the goals of the Next Generation Science Standards and equips them with instructional practices that promote student learning in science.

KEYWORDS

elementary science education; Next Generation Science Standards; scientific argumentation; teacher practice-based professional development

A growing body of evidence suggests that classroom learning is most effective when students and teachers engage in a dialogic interaction to explore ideas, constructs, and concepts (Cazden, 2001; M. Chi, 2009; Johnson & Johnson, 2009; Resnick, Asterhan, & Clarke, 2015; Wells, 2001). This form of dialogue is especially relevant in school science, in which argument and debate are essential for enhancing students' conceptual understanding and ability to reason scientifically (Alverman, Qian, & Hynd, 1995; Kuhn, 2010; Mercer, Dawes, Wegerif, & Sams, 2004; Osborne, 2010). Moreover, engaging students in argument from evidence is one of the core practices in the Next Generation Science Standards (NGSS), which will be implemented in many U.S. states over the next 3 years. However, although engaging in argument from evidence is a core epistemic practice of

CONTACT Evan J. Fishman dr.evanfishman@gmail.com Stanford Graduate School of Education, Stanford University, 485 Lasuen Mall, Stanford, CA 94305, USA.

these standards, it is a rare feature of science lessons (Henderson, Osborne, MacPherson, & Wild, 2015; D. P. Newton & Newton, 2000). To meet the challenge of transforming pedagogy and curriculum, therefore, many states are in the process of beginning an extended program of professional development (PD) to prepare and support teachers to implement the NGSS. Hence, building knowledge of how to best prepare teachers to meet the demands of the NGSS—the essential focus of our research—is of considerable interest to those working in the fields of science education, PD, and teacher education.

In the present study, we examined the efficacy of the Practicum Academy for Improving Science Education (PRACTISE) PD program, an innovative approach to science PD for elementary (Grades 3–5) classroom teachers. The overarching goal of the PRACTISE project is to develop a model of PD that will enable elementary teachers to engage and support their students in discursive scientific reasoning using evidence-based argument. Given the importance of this type of classroom discourse, we believe that this study offers important findings that inform knowledge and understanding of the kind of PD needed to support teachers in achieving the goals of the NGSS.

Scientific argumentation from evidence

One of the primary goals of education is to build knowledge and understanding of the major ideas that have shaped the cultural milieu. Yet formal education still relies heavily on the notion that classroom learning is a process of imparting or transmitting knowledge, despite evidence that suggests that this type of learning is ineffective for building knowledge (P. Newton, Driver, & Osborne, 1999; Nystrand, Gamoran, Kachur, & Prendegarst, 1997; Reddy, 1979; Weiss, Pasley, Smith, Banilower, & Heck, 2003). Discourse in the classroom is commonly dominated by a form of dialogue known as initiation–response–evaluation, in which the teacher asks a question, the student provides a brief phrase-like answer, and the teacher responds by offering an evaluation of whether it is appropriate (Cazden, 2001; Edwards & Mercer, 1987; Lemke, 1990). Although this form of discourse is effective at imparting knowledge, it suffers from three major weaknesses. First, the cognitive demand is very limited, often requiring little more than recall from the student. Second, the nature of the interaction is restricted solely to one between the teacher and a single student, which minimizes the discursive demands. Third, evidence would suggest that it is not an effective way of supporting student learning (M. Chi, 2009; Hattie, 2008).

Ohlsson (1996) has argued that constructing knowledge is an epistemic process that requires students to engage in the discursive activities of describing, explaining, predicting, arguing, and critiquing. Empirical support for this claim can be found in M. Chi's (2009) meta-analysis of the learning outcomes of instructional activities defined by three distinct forms: active, constructive, and interactive—active in that they are doing something, constructive in that they produce a product, or interactive in that they engage in dialogue and produce a product. Her review showed that interactive approaches were more effective than constructive approaches, which in turn were more effective than active approaches. Likewise, other researchers have shown similar results (e.g., Cazden, 2001; Johnson & Johnson, 2009; Wells, 2001). For science specifically, a summary of the value of interactive dialogue was provided by Osborne (2010).

In the case of science, the need to engage in such dialogue is even more pressing, as the teacher of science is fundamentally a teacher of unnatural (Wolpert, 1992), uncommon sense (Cromer, 1993) or what might be considered crazy ideas—for instance, the idea that a spinning Earth rather than a moving sun causes day and night when the obvious explanation is that it is the sun that moves, the idea that air has mass, or the idea that plants get the majority of their mass from the air. Such ideas have to be argued for, particularly, as in many cases, consensual scientific belief was the product of many years of argument. Thus, the notion of arguing from evidence is a central feature of *A Framework for K–12 Science Education* (National Research Council, 2012), a document that forms the basis of the U.S. NGSS (NGSS Lead States, 2013).

Likewise, we argue that the effective learning of science is dependent on opportunities to engage in discursive interactions in which students engage in critique and evaluation. Although constructing explanations or argument is an activity that undoubtedly has value, deeper understanding requires students to identify flaws in both their own reasoning and that of others. As Mercier and Sperber (2011) showed, human reasoning suffers from confirmation bias and group interaction, and discussion is essential to sustaining epistemic vigilance. Thus, reasoning often produces its best results when conducted in group contexts (M. Chi, De Leeuw, Chiu, & Lavancher, 1994; Hatano & Inagaki, 1991; Mercier & Sperber, 2011).

Hence, if students are to engage in the eight practices of science suggested by *A Framework for K–12 Science Education* (National Research Council, 2012) and develop an understanding of scientific reasoning, then they will need to engage in dialogic interaction—that is, to argue from evidence. Although it is possible that such practice can take place on an individual basis through well-structured problems, like any performance the nature of practice is better understood when it is modeled and supported by an accomplished practitioner, in this case the teacher. The challenge for science teachers, though, is to build and scaffold their own argumentative competence and to provide opportunities for students to reason critically. In short, to change the common form of discursive interaction that dominates in science classrooms, the following is needed: a repetitive sequence of the teacher asking a question; the student responding with a short, phrase-like answer; followed by an evaluation that is more discursive than dialectic (dialectic in the sense that the discourse consists of extended student responses in which differences are explored and ideas are progressively and cumulatively improved; Alexander, 2006).

Practice-based models of PD

Many scholars have argued for the importance of practice-based PD, which centers teacher learning on the core tasks and activities of teaching rather than knowledge or theory (e.g., Ball & Cohen, 1999; Ball & Forzani, 2009; Borko, 2004; Borko, Jacobs, & Koellner, 2010; Lampert, 2010). One common component of practice-based PD programs is the use of records of practice such as classroom video as a tool for bringing the central activities of teaching into the PD setting (Ball & Cohen, 1999; Gaudin & Chaliès, 2015; Koellner & Jacobs, 2015; Seago, Jacobs, Heck, Nelson, & Malzahn, 2014; van Es & Sherin, 2010). Like other records of practice, such as examples of student work and instructional materials, video provides an opportunity for teachers to study their practice collaboratively without being physically present in the classroom (Borko, Jacobs, Seago, & Mangram, 2014). Clips from videotaped classroom episodes can be viewed repeatedly and from multiple perspectives, enabling teachers to closely examine classroom interactions as well as the content addressed in the lessons and to discuss ideas for improvement.

Changing teaching practices also requires opportunities to test what works or does not work in classrooms, opportunities for reflection, and a community to share experiences (Jennings & Mills, 2009; Martin & Hand, 2009).During the school year, however, external constraints such as time, state standards, testing requirements, and instructional resources can limit the opportunities teachers have to practice new instructional moves or reflect on practices collaboratively with peers. The structure and constraints of schools can limit teachers' implementation of new strategies regardless of changes to their knowledge or beliefs. Although uncommon in PD for veteran teachers, practicum experiences—courses designed to provide supervised practical applications of previously or concurrently studied theory and methods—are a hallmark of professional preparation in teaching as well as in fields of study such as medicine, nursing, and social work (Grossman et al., 2009; Ryan, Toohey, & Hughes, 1996).From an educational perspective, a practicum experience provides teachers with an opportunity to put their newly acquired skills and knowledge to use with real students in a highly supportive classroom environment. This allows teachers to focus on changing their practice without the typical constraints or outside pressures that occur during the school year. In an evaluation of previous practicum-based academies, evidence from teacher surveys and interviews indicated that teachers who attended practicum-based PD programs made significant shifts in their knowledge and beliefs and that they were comfortable with implementing new practices they had learned (B. Chi, Visintainer, Chung, de Nova, & Sacco, 2011).

Practice-based PD programs also afford teachers the opportunity to refine their practices through sustained professional discourse and support from the collaborative PD community of colleagues and PD leaders (Ball & Cohen, 1999; Borko et al., 2010). One way this is achieved is through engagement in communities of practice, in which teachers meet on a regular basis to reflect on their classroom teaching throughout the school year.

The present study

The PRACTISE project was designed to study the impact of an innovative science PD program for upper elementary classroom teachers on the instructional practices of classroom teachers. The PD model (also known as "Academy") consisted of three components: a 1-week institute, a 2-week practicum, and four 1-day follow-up sessions. The overarching goal of this PD was to help teachers learn to facilitate collaborative, critical scientific discourse and specifically argumentation from evidence.

This study examined the impact of a practice-based PD designed to help teachers facilitate scientific argumentation in their classroom. More specifically, we sought to examine the efficacy of the Academy PD model in enhancing teachers' instructional practice using whole-classroom discourse. The PD facilitators enacted two versions of the PD: the full Academy (institute, practicum, and follow-up sessions) and the Academy minus the practicum (institute and follow-up sessions only). The difference between these approaches allowed us to assess whether the practicum component of the PD contributed uniquely to the nature and quality of teachers' instructional practices. Hereafter we refer to the teachers who attended the full Academy as Cohort 1 (practicum group) and the teachers who attended the Academy minus the practicum as Cohort 2 (no-practicum group).

This study focused on the first phase of a multiyear research project and examined changes in teachers' and students' discourse practices. Specifically, the following research questions were addressed:

(1) To what extent, if any, does teachers' participation in a PD program focused on discourse and argumentation influence classroom discourse practices?
(2) What differences in discourse practices, if any, are associated with teachers' participation in the practicum and no-practicum versions of the PD program?

Given that the PRACTISE PD was designed to help teachers facilitate scientific argumentation in the classroom, we hypothesized that teachers' use of these instructional practices would improve after they attended the PD for both cohorts. Hence, with respect to the first research question, we expected teachers to make statistically significant gains on measures of their use of classroom discourse to support scientific argumentation after attending the PD program. Because the practicum was an intensive component of the PD program, we also hypothesized that participation in the practicum would provide a statistically significant benefit for teachers' use and quality of discourse practices compared with Cohort 2, the teachers who did not receive the practicum.

Instructional practices to support scientific discourse

To measure the quality of classroom discourse, we drew from the extant literature to focus on six classroom practices that are seen as instrumental to supporting scientific discourse, three of which are teacher related and three of which are student related. In what follows, we offer a rationale for the inclusion of each practice. The following three practices are teacher related:

(1) *Ask*: teachers' use of open-ended questioning in an effort to generate productive science discourse. The typical closed question asked within the standard triadic initiation–response–evaluation dialogue seeks to elicit the right answer in order to make knowledge public (Cazden, 2001; Edwards & Mercer, 1987; Lemke, 1990). In contrast, engaging students in an argumentative discourse requires a dialogue in which there are potentially several valid responses. Argumentation is then required to resolve the difference and evaluate the validity of different responses. This type of discourse can be initiated only by using open questions when there is no self-evident singular better answer. Such questions are generative or productive (Chin & Osborne, 2010; Dillon, 1982; King, 1990)in that their function is to foreground difference and to generate cognitive conflict. Thus, productive questions ask students to elaborate or explain their reasoning.
(2) *Press*: the extent to which teachers press students to elaborate their reasoning. Rather than simply evaluating a student response as good or not good, a teacher can develop a student's thinking by applying a technique such as teacher press (Kazemi & Stipek, 2001), in which he or she asks productive follow-up questions such as "How do you know?" or "Can you tell me more about that?" or even simply "Why?" (Mercer & Howe, 2012). The subsequent questions are productive if they encourage "students to formulate hypotheses, predict outcomes, brainstorm ideas,

generate explanations, make inferences and conclusions, as well as to self-evaluate and reflect on their own thinking" (Chin, 2006, p. 1336).

(3) *Link*: the extent to which teachers relate one student contribution to another. One of the key features of dialogic teaching is that it is cumulative, whereby "teachers and children build on their own and each other's ideas and chain them into coherent lines of thinking and enquiry" (Alexander, 2006, p. 26). Whereas discussion is characterized as sharing information, dialogue's main goal is to "achieve a common understanding" (p. 29). In order for students to achieve a common understanding, the teacher can use discourse moves that connect or compare student responses in an effort to provide a continuous and cumulative conceptual thread throughout the entire discourse segment in which flawed thinking is exposed and good ideas are built on.

The following three practices are student related:

(1) *Explain/Claim*: the extent to which students offer extended explanations and support claims with evidence. Empirical evidence suggests that the discursive act of offering an explanation enhances student understanding (M. Chi, 2009; M. Chi et al., 1994; Franke et al., 2015; Hatano & Inagaki, 1991; Webb, 1989). In a review of 19 published studies on learning mathematics and computer science, Webb (1989) found that the level of elaboration of students' interaction with other students was related to achievement: Those students who gave high-level elaboration to other members of the group achieved more highly, whereas for those offering only low-level explanations the experience had no effect.

(2) *Coconstruct*: the extent to which students build on one another's ideas, ask others to clarify, elaborate, or extend ideas. For such discourse to be effective, students need to participate as a community of learners (Rogoff, 1990). In such a community, all class members, including adults, take responsibility for their contributions to their own learning and the whole class's functioning by listening to mutual contributions in a conversational way and exploring the materials and ideas. Identifiable features of student discourse that adhere to such values would be children receiving, building, and elaborating one another's answers (Alexander, 2006).

(3) *Critique*: the extent to which students challenge or critique one another's ideas. The construction of knowledge is dialogue between construction and critique (Ford, 2008). One indication that students are attempting to engage in the cognitive effort required to coconstruct a deeper understanding of the concept at hand is any discourse in which students ask questions either of one another or alternatively of the teacher. There is substantive empirical evidence that the asking of questions improves student learning (Chin & Osborne, 2008, 2010; Dillon, 1982; King, 1990, 1994). Thus, supporting students to ask questions about the phenomenon under discussion and of one another has the potential to stimulate more extended and elaborated arguments. More fundamentally, a discursive space that encourages the asking of questions helps communicate an epistemic stance toward knowledge that sees all claims to know as requiring justification and the maintenance of epistemic vigilance.

For the reasons outlined here, we consider all of these discursive practices foundational to fostering and facilitating scientific argumentation in the classroom. Hence, our method of assessing growth in teacher and student scientific argumentation focused on these specific practices (see the "Science Discourse Instrument (SDI)" section).

Methods

Participants

A total of 44 elementary school teachers participated in the study. These participants were recruited from a large urban school district in Northern California. All teachers taught science at either the third-, fourth-, or fifth-grade level. Seven teachers left the project because of circumstances that prevented their participation (e.g., health issues). This resulted in 18 teachers in Cohort 1 (practicum group) and 19 teachers in Cohort 2 (no-practicum group; see Table 1). For Cohort 1, the majority of teachers were female (78%), and the average years of teaching experience was 10.9. For Cohort 2, the majority of teachers were also female (79%), and the average years of teaching experience was more than 8.4. Cohort assignment was carried out at the school level; that is, after stratifying for socioeconomic status, we randomly assigned schools to either cohort. This was done so that teachers from the same school could participate in identical PD programs. After we accounted for attrition, Cohort 1 consisted of eight schools; in six of these schools the majority of students (75%–95%) were eligible for free or reduced price meals. Cohort 2 consisted of 10 schools; in nine of these schools the majority of students (70%–93%) were eligible for free or reduced price meals.

The practicum-based PD program

The institute was a weeklong summer program in which the PD leaders introduced a variety of instructional practices intended to support scientific discourse and argumentation among students. The PD leaders decided on this focus based on research, outlined previously, showing that opportunities for students to engage in collaborative discourse—to advance claims, support their ideas, be challenged, and challenge others—lead to improvements in students' conceptual understanding and scientific reasoning. In this institute, the PD leaders used an inquiry-based curriculum about ocean science to help demonstrate the type of lessons and investigations that promote argumentation. Through

Table 1. Teacher and school demographic information.

Variable	Cohort 1	Cohort 2
No. of schools	8	6
Average percentage of students eligible for FRPM	59.85%	80.35%
No. of teachers	18	19
Sex	78% female	79% female
Grade taught		
Grade 3	6	6
Grade 4	6	6
Grade 5	6	7
Average years of experience	10.9	8.4

Note. FRPM = free or reduced price meals.

explicit and implicit means, the teacher participants learned to create and establish classroom norms, develop instructional moves, and use activities that support student discourse (see Table 2 for a summary of institute activities and content; see also Berson, Borko, Million, Khachatryan, & Glennon [in press] for additional information).

Following the institute, approximately half of the teachers attended the practicum, an intensive 2-week program in which the participants spent the morning teaching science to local summer school students. This allowed the teachers to use and try out the instructional practices they learned in the institute. These lessons were supervised by the PD leaders and videotaped by the research team. Later in the afternoon, teachers participated in video discussion groups and one-on-one coaching sessions (see Table 2 for details). Thus, the practicum allowed teachers to practice using instructional practices that facilitated scientific discourse in a low-stakes and highly supportive environment and to refine these practices using modeling, reflection, and feedback.

The institute and practicum took place during the summer of 2013. During the next academic year, the teachers in both cohorts attended four follow-up sessions. These follow-up sessions were 1-day programs in which the PD leaders continued to model instructional practices and discuss strategies for facilitating scientific argumentation. In these follow-up days, teachers also shared and discussed their own experiences of implementing these practices in their classrooms. As was the case in the practicum, teachers engaged in video discussion groups, this time using actual classroom video that had been

Table 2. Brief overview of the PD sessions.

Time	Session	Days	Activities and content
Summer 2013	Institute (Cohorts 1 and 2)	Five	Presentations and practical activities on argumentation in the science classroom demonstrating its value and application Teachers participated in discussion and modeling of argumentation strategies for promoting discussions (e.g., norms to create a culture of talk, asking productive questions, writing to support talk) and activities to engage students in providing claims and arguing from evidence (e.g., idea line up, concept cartoon). Introduction to an ocean science curriculum that features lessons that promote argumentation Small- and whole-group discussion about applying argumentation practices in the classroom Reflection and lesson planning for the AY
Summer 2013	Practicum (Cohort 1 only)	10	Teachers led science lessons in summer school for students. These lessons were observed by PD leaders and videotaped for later review. After each day of teaching video discussion groups were held in which small groups of participants would watch a teacher's lesson and provide feedback. These groups were moderated by PD leaders. Teachers participated in one-on-one coaching sessions in which a PD leader would closely review the teacher's lesson and discuss. Reflection and lesson planning for the AY
AY 2013–2014	Follow-up sessions (Cohorts 1 and 2)	Four sessions (1 day each)	Teachers participated in small-group activities that involved hypothetical situations of scientific argumentation in the classroom and strategies for overcoming barriers. This involved modeling and discussions about the strategies and activities. Whole-group discussions about the teachers' experience with scientific argumentation during the year Video discussion groups were held using teachers' classroom video from the current AY.

Note. PD = professional development; AY = academic year.

captured during the school year. These videos provided examples of authentic teaching and learning episodes that served as records of practice to situate discussion and reflection in the teachers' classrooms (see Table 2).

Procedure

We captured classroom video of the teachers' science lessons in the school year before the PD (baseline; Time 0) and in the school year following the PD (Time 1). Each teacher was videotaped at least twice (Observations 1 and 2) and some were videoed three times in the winter and/or spring of each school year. Because we were not able to capture many of the teachers for a third observation, this observation was dropped from the analysis. Before videotaping began, teachers were told, "We are interested in videotaping a science lesson in which students have opportunities to talk with the teacher or with each other. Please choose a typical science lesson in which talk happens." The science lessons that we videotaped were between 30 and 90 min in length.

Once the classroom video data were collected, we chose to analyze the discourse by identifying two 15-min segments (segment A and segment B) containing a minimum of 5 min of whole-classroom discussion or talk about science (see the Appendix). Segment A generally occurred in the beginning of the lesson, whereas segment B generally occurred toward the end of the lesson. This may reflect how the teachers typically structured their science lessons (i.e., with an opening discussion and concluding discussion). There were some instances in which segment B occurred earlier in the lesson. This occurred when teachers held the second whole-group discussion earlier in the lesson to make time for a closing, nondiscussion activity or if the opening discussion exceeded 30 min and there was no further whole-group discussion.

Science Discourse Instrument (SDI)

To measure the teacher and student scientific argumentation practices previously discussed, we developed an SDI based on the six practices (see the Appendix for the complete instrument). Thus, the SDI focused on the three teacher-related discourse practices (Ask, Press, and Link) and the three student-related practices (Explain, Coconstruct, and Critique). Much of our instrument was an adaptation from other scales that have been extensively tested and validated; to have conducted a full validation study would have exceeded the time and resources we had available. Specifically, the Ask practice was an adaptation of a scale in the Dialogic Inquiry Tool developed by Reznitskaya (2012). Our teacher Press and Link practices were adapted from the Press and Link scales in the Instructional Quality Assessment (Matsumura et al., 2006). Likewise, the Explain/Claim and Coconstruct scales were adapted from the Instructional Quality Assessment scale for student linking and providing. The major adjustment made to the SDI scales was to change the language to reflect the discourse of argumentation in science more accurately rather than that of mathematics or language arts. In particular, neither the Dialogic Inquiry Tool nor the Instructional Quality Assessment identifies explicitly the process of critiquing scientific claims with the use of evidence-based reasoning, a practice that we consider essential for productive scientific argumentation. Each of these practices was described in the SDI, as were instructions for rating. The rating for each practice was on a scale from 0 to 4 (see the Appendix for more detail).

The SDI was also designed to capture specific features of the science lesson (e.g., purpose of the discourse, use of think time, references to discussion norms; Osborne et al., 2015). Here we only report the findings from the six instructional practices, our principal focus.

Rating

Our research team rated video segments with respect to the teachers' and students' science discourse practices. Prior to rating, six researchers were trained to rate the video segments using the SDI. As a group, raters began by watching a single segment and discussed what ratings it should be given. Following these discussions, the researchers rated videos independently. Then raters joined in pairs and rated several videos. This training continued until the researchers achieved reliability among pairs. After training concluded, the same process was used in the rating of all video observations. Each video segment was rated by a pair of raters. Each rater in the pair first rated the segment individually. Next the two raters discussed any disagreements in ratings and reached a consensus rating. Ultimately a consensus was reached for all ratings.

Rater reliability on the video data used to train raters was assessed using generalizability (G) theory (GENOVA Software; Brennan, 2000). We chose a G theory approach because it is commonly used to evaluate the dependability of behavioral measures (Shavelson & Webb, 1991). It is particularly well suited for investigating the reliability of measures of instruction because it can assess the relative importance of multiple sources of variance simultaneously (e.g., teachers, raters, occasions; see Moss, Cyr, & Dubois-Comtois, 2004). Hence, we ran a Teacher × Rater G study treating segment as a fixed facet (analyzing data for segments A and B separately). The interrater reliabilities (generalizability coefficients) are reported in Table 3. These coefficients were based on two raters and four teachers (indicated by "Raters = 2" in Table 3), and on six raters and four teachers (indicated by "Raters = 6"). Overall, Link and Press yielded the lowest reliability. The results suggested that the ratings were more reliable with six raters as opposed to two. For Press segment B, the estimated universe (true) score variance was zero. The teachers simply did not vary in their Press (segment B) practices.

Results from the G study also provided information about which variables should be aggregated in subsequent analysis. For example, we found that there were substantive

Table 3. Generalizability coefficients for each scale.

	Segment A		Segment B	
Practice	Raters = 2	Raters = 6	Raters = 2	Raters = 6
Ask	.50	.75	.81	.93
Press	.80	.92	.00	.00
Link	.23	.48	.59	.81
Teacher aggregate	.76	.90	.69	.87
Explain	.81	.93	.48	.73
Coconstruct	.83	.94	.72	.89
Critique	.98	.99	.96	.98
Student aggregate	.94	.98	.91	.97

differences in ratings of segments A and B because these segments essentially represented different phases of the lesson and that reliability was improved if they were treated separately. Thus, we did not combine the segments in the subsequent analyses; rather, we treated segment as a fixed effect. Conversely, the results showed that Observation Occasions 1 and 2 could be combined, which allowed us to average ratings across observations so that all segments had a single Time 0 (baseline) and Time 1 (Year 1) rating on each of the six practices. In addition, our exploratory investigation suggested that aggregating the teacher and student practices into a teacher practice measure and a student practice measure may have benefited the analyses structurally; that is, the measures were more reliable when combined as demonstrated in Table 3. Thus, we conducted both an analysis with these teacher and student measures aggregated and an analysis with the practices not aggregated.

Results

Equivalence of cohorts

We examined the equivalence of cohorts at Time 0 to assess whether randomization was successful. A one-way analysis of variance (ANOVA) assessed the between-subjects effects of Time 0 ratings for all practices. With the exception of Critique segment B, there were no significant differences in Time 0 ratings between cohorts, which suggests that randomization was successful, $F(1, 33) = 4.38$, $p = .044$.

Descriptive statistics

Table 4 displays the means and standard deviations for all of the teacher and student practices, including the aggregated teacher and student measures. At Time 0, the mean ratings for each teacher and student practice were on the lower end of the scale. Overall, the ratings improved at Time 1 for both cohorts. This suggests that the PD had a positive impact on the teachers' and students' argumentation practices.

It is notable that teacher asking (segments A and B) yielded a mean rating above 3.0, which is equivalent to the occasional use of that practice at a high level (see the Appendix). That is, teachers occasionally asked open-ended questions that elicited discussion (e.g., "How do you know our planet is mostly water?"). Student explaining also yielded ratings above 3.0 on average, which indicates that students occasionally backed up their claims with relevant evidence (e.g., "The earth is mostly water because I see the globe is mostly blue").

Ratings for teacher Press improved noticeably at Time 1, which indicates that the PD may have helped teachers to elicit deeper student thinking (e.g., "Why do you say that? What's your evidence?"). Similarly, student Coconstruct was markedly improved at Time 1, which indicates that students more frequently built on one another's claims (e.g., "I agree with Matt because I read in a book that the earth was mostly water").

Conversely, ratings for teacher linking remained relatively low at Time 1. This demonstrates that teachers continued to struggle with linking students' ideas together or summarizing the discussion to move forward (e.g., "Sammy has an idea, Dana has an idea. Which idea is stronger or more compelling, and why?"). Students' use of Critique also remained low at Time 1. That is, students rarely expressed a counterargument or example

Table 4. Descriptive statistics by group.

Measure	Practicum			No practicum		
	N	*M*	*SD*	*N*	*M*	*SD*
			Segment A			
Teacher practices						
Ask						
Time 0	14	1.93	0.81	17	1.85	0.90
Time 1	14	3.21	0.83	17	3.00	0.97
Press						
Time 0	14	1.89	0.79	17	1.56	0.68
Time 1	14	2.86	0.86	17	2.77	0.66
Link						
Time 0	15	1.67	0.56	17	1.35	0.61
Time 1	15	1.80	0.49	17	1.91	0.64
Aggregated measure						
Time 0	13	1.78	0.45	16	1.60	0.66
Time 1	13	2.63	0.62	16	2.36	0.70
Student practices						
Explain						
Time 0	18	1.97	0.78	17	1.76	0.56
Time 1	18	3.08	0.73	17	2.35	0.81
Coconstruct						
Time 0	18	1.33	0.57	17	1.24	0.56
Time 1	18	2.42	1.06	17	1.65	0.66
Critique						
Time 0	18	1.14	0.61	17	1.03	0.45
Time 1	18	2.08	1.10	17	1.38	0.63
Aggregated measure						
Time 0	18	1.48	0.62	17	1.34	0.47
Time 1	18	2.53	0.91	17	1.79	0.61
			Segment B			
Teacher practices						
Ask						
Time 0	14	2.36	0.84	16	1.94	0.91
Time 1	14	3.32	0.72	16	3.25	0.86
Press						
Time 0	14	2.21	0.58	16	1.72	0.84
Time 1	14	2.90	0.76	16	2.28	0.75
Link						
Time 0	14	1.61	0.59	16	1.34	0.65
Time 1	14	2.18	0.69	16	2.18	0.76
Aggregated measure						
Time 0	13	2.18	0.43	16	1.80	0.64
Time 1	13	2.83	0.70	16	2.48	0.67
Student practices						
Explain						
Time 0	18	2.67	0.80	17	1.71	0.66
Time 1	18	3.67	0.79	17	2.68	0.88
Coconstruct						
Time 0	18	1.42	0.65	17	1.12	0.49
Time 1	18	2.94	0.84	17	2.29	0.90
Critique						
Time 0	18	1.39	0.81	17	0.94	0.35
Time 1	18	2.42	0.93	17	1.50	0.79
Aggregated measure						
Time 0	18	1.66	0.70	16	1.25	0.46
Time 1	18	2.84	0.79	17	2.51	0.78

(e.g., "I disagree with what Javier that crabs are arachnids because they live in the ocean so they are crustaceans"). These results suggest that Link and Critique may be the most difficult practices for teachers and students to adopt.

The impact of PD models

To address our research questions we compared the Time 0 and Time 1 ratings and the ratings of Cohort 1 (practicum) and Cohort 2 (no practicum). In all analyses, missing ratings were deleted pairwise. First we conducted a set of analyses in which we combined the three teacher practices (Ask, Press, Link) and the three student practices (Explain, Coconstruct, Critique). These analyses were run separately for teacher practices and student practices, each with a 2 × 2 × 2 (Cohort × Segment × Time) split-plot ANOVA design with cohort a between-subjects variable.

Research Question 1: To what extent, if any, does teachers' participation in a PD program focused on discourse and argumentation influence classroom discourse practices?

As expected, we found a significant effect of time for teacher practices and student practices: teacher practices, $F(1, 27) = 48.76$, $p < .001$, $\eta_p^2 = .64$; student practices, $F(1, 33) = 55.83$, $p < .001$, $\eta_p^2 = .63$. This indicates that when we controlled for segment and cohort, the teacher and student ratings were significantly higher at Time 1 than at Time 0.

Research Question 2: What differences in discourse practices, if any, are associated with teachers' participation in the practicum and no-practicum versions of the PD program?

Of particular note is the finding that teachers who attended the more extensive version of the PD (Cohort 1) did not make statistically significant gains over their counterparts who attended the less extensive version of the PD (Cohort 2); that is, the Cohort × Time interaction was not significant ($p = .883$). However, the difference in gains between Cohort 1 and Cohort 2 approached significance for the student practices measure, with the difference between ratings at Time 0 and Time 1 being greater for Cohort 1 than Cohort 2, $F(1, 33) = 3.35$, $p = .076$, $\eta_p^2 = .09$.

Secondary analysis

We then conducted an analysis that included the six practices as separate variables in a single model, using a 2 × 2 × 2 × 6 (Cohort × Segment × Time × Practice) split-plot ANOVA and also treating practice as a vector of outcomes in a multivariate ANOVA. This set of analyses (nonaggregated model, in which the six practices were examined separately) allowed us to further explore how the use of specific argumentation practices changed after the first year of the PD. Again we found a significant main effect for time: The average rating was higher at Time 1 than at Time 0 (for Cohort 1 and Cohort 2 combined and all six practices combined). However, again the Cohort × Time interaction was not significant. That is, Cohort 1 did not make significantly greater gains than Cohort 2. When we looked at the practices separately, the only practice for which the gains were significantly greater for Cohort 1 than Cohort 2 was the student discourse move of Coconstruct for segment A ($p = .014$), but this might have been a chance effect reflecting multiple significance testing.

Discussion and implications

The impact of the practicum

Our central findings demonstrate that the PRACTISE Academy, with its combination of a summer institute and follow-up sessions during the school year, is an effective model for helping teachers to develop the core practices of scientific discourse and to incorporate them into their ongoing classroom instruction. Overall, the teachers who participated in the study made significant improvements in their discourse practices, as did their students. Thus, the combination of a 1-week workshop and four follow-up days was successful at aligning teachers with the standards of the NGSS and specifically in this case developing their ability to engage their students in argument from evidence.

However, based on our findings, we cannot conclude that the practicum experience added significantly to the impact of the summer institute and follow-up sessions. Although students of teachers who attended the practicum made improvements that approached statistical significance, the practicum teachers' own practices did not show a significant improvement over those of their no-practicum counterparts, although on average their mean ratings were higher. One possible explanation for this unexpected outcome is that the practicum teachers may have established a culture of talk more rapidly at the beginning of the academic year, including discussion norms supportive of argumentative discourse. However, by spring (when the video data were collected) the no-practicum teachers had caught up with those who had attended the practicum so that there was no difference in their ratings. In contrast, if the practicum teachers established norms that supported students' discussions earlier in the year, their students may have been more skilled and able to engage in appropriate discursive acts in the spring, so that their ratings were higher than those of the no-practicum students, approaching significance (Resnick, Michaels, & O'Connor, 2010).

Another possibility is related to the fact that both cohorts attended the same follow-up sessions throughout the school year. These follow-up sessions may have had a stronger impact than the summer program given that they took place during the school year and enabled teachers to try out and refine the discourse practices in a continuing and cumulative manner. It is important to consider that both the practicum and no-practicum versions of the PD met all of the criteria for effective PD suggested by Desimone (2009), including the duration criterion of more than 20 hr of contact time. She also indicated that research "shows support for activities that are spread over a semester (or intense summer institute with follow-up during the semester)" (p. 184). In addition, the PD program design called for the PD leaders to take their observations of the teachers' strengths and needs into account in developing each follow-up session. Thus, it is possible that the follow-up sessions were the more focused and impactful component of the PD. This explanation receives support from the literature on practice-based PD in that the core activities of these follow-up sessions were centered on teacher learning (Ball & Cohen, 1999; Putnam & Borko, 2000). Moreover, these core activities were discussion groups based on teachers' classroom video, a valuable tool for bringing the central activities of teaching into the PD setting (Gaudin & Chaliès, 2015; Kiemer, Groschner, Pehmer, & Seidel, 2015; Koellner & Jacobs, 2015; Seago et al., 2014; van Es & Sherin, 2010). These two possible explanations suggest that a more revealing measure of the impact of the PD would be the change in teachers' and students' discourse practices over the course of the year, which is a hypothesis to be tested by future research.

In sum, elements of the PRACTISE PD were effective at helping teachers to implement pedagogy more consistent with the core principles of the NGSS. Using elements of practice-based PD, such as video reflection with peers, and sustained support from PD facilitators, teachers and students made significant improvements in their discourse practices after attending the PD (e.g., Ball & Cohen, 1999). This is a promising finding for science educators who are interested in teaching students to think critically about science, back up their claims with evidence, and interact with peers to build on and debate scientific concepts. The practicum element, however, in which teachers were able to practice their newly learned argumentation practices with students attending summer school, did not contribute significantly to the improvement of such practices.

Individual instructional practices

The data show that, on the Ask practice, which measures teachers' ability to ask open-ended questions, teachers' ratings improved over time, from asking such questions rarely to a little more than occasionally. Such questions provide opportunities for students to elaborate and explain their reasoning about the important scientific ideas in the lesson. Perhaps not surprising then is that the other practice that improved from being demonstrated rarely to at least occasionally was Explain, which represents the frequency with which students' responses are about claims/explanations and the extent to which those claims/explanations are supported by evidence and reasoning. Such discourse does not, however, necessarily facilitate dialogic interactions that require some form of comparison and contrast.

What is absent from the dialogue then, as the low Link and Critique ratings indicate, is the facilitation of comparison and contrasting that support the critical evaluation of student thinking and a purposeful epistemic goal. Rather, much of the talk exhibited in such lessons takes the form of being cumulative in that students (and teachers) share knowledge in an uncritical fashion, repeating and elaborating one another's ideas, accepting and agreeing with previously stated contributions (Mercer, 2010). This is distinct from the kind of discourse that Mercer (2000) referred to as "exploratory talk," which is characterized by a critical and constructive engagement with contrasting ideas advanced by other students and in which individuals are asked to justify their claims with reasons and the goal is to reach agreement on a common understanding. Hence, although the discourse observed in these classrooms might be said to be discursive, it was not dialectic in the sense that differences in the ideas students held were elicited, arguments were critiqued, and differences were resolved.

Future directions

This article reported findings from the first 2 years of our 4-year project. These findings provide evidence of the kinds of changes that can be effected in teachers' use of a classroom discourse that supports argumentation from evidence and the challenges such discourse poses. In the next phase of our project we will examine what happens with an additional year of ongoing PD, whether changes are sustained in the final year during which there will be no support, and what differences there might be between the two cohorts.

The literature suggests that these scientific discourse practices facilitate student learning; thus, it is important to examine whether PD programs, such as the one featured in this study, are effective at promoting students' science knowledge. As part of our multiyear project, we have designed grade-specific science assessments that students take at the beginning and end of the school year. We will use these student assessment scores to gather evidence of students' science knowledge in relationship to the specific PD models their teachers participated in and their changes in discourse practices.

In addition, we plan to conduct detailed case studies of a small number of teachers in each cohort. The rating scales have enabled us to document broad patterns of change in teachers' and students' discourse practices. However, they do not provide detailed information about the relationships between specific teacher discourse moves and student discourse moves. An analysis of these relationships will help us to better understand the unexpected patterns of change revealed in our analyses to date. This more in-depth analysis will also enable us to examine the relationship between specific activities and discourse practices emphasized in the PD and changes in the nature of discourse in the teachers' classrooms. Understanding these complex relationships will provide important insights into the type of PD that does and does not support teachers in engaging their students in argument from evidence.

Acknowledgments

We gratefully acknowledge Craig Strang, Emily Weiss, and Bernadette Chi at the Lawrence Hall of Science, University of California, Berkeley, for enacting the Practicum Academy for Improving Science Education professional development academy. We thank Rich Shavelson for his tremendous guidance with data analysis.

Funding

The research reported here was supported by the National Science Foundation through Grant No. 1220666 to Stanford University.

References

Alexander, R. (2006). *Towards dialogic teaching*. York, England: Dialogos.

Alverman, D. E., Qian, G., & Hynd, C. E. (1995). Effects of interactive discussion and text type on learning counterintuitive science concepts. *Journal of Educational Research*, *88*, 146–154.

Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as a learning profession* (pp. 3–32). San Francisco, CA: Jossey-Bass.

Ball, D., & Forzani, F. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, *60*, 497–511.

Berson, E., Borko, H., Million, S., Khachatryan, E., & Glennon, K. (in press). Practice what you teach: A video-based practicum model of professional development for elementary science teachers. *Orbis*.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (Vol. 7, pp. 548–556). Oxford, UK: Elsevier.
Borko, H., Jacobs, J., Seago, N., & Mangram, C. (2014). Facilitating video-based professional development: Planning and orchestrating productive discussions. In Y. Li, E. A. Silver, & S. Li (Eds.), Transforming mathematics instruction: Multiple approaches and practices. (pp. 259–281). Dordrecht, The Netherlands: Springer International Publishing.
Brennan, R. L. (2000). Performance assessment from the perspective of generalizability theory. *Applied Psychological Measurement*, 24(4), 339–353.
Cazden, C. B. (2001). *Classroom discourse* (2nd ed.). Portsmouth, NH: Heinemann.
Chi, B., Visintainer, T., Chung, J., de Nova, J., & Sacco, K. (2011). *¡Youth & the ocean! (¡YO!) Final evaluation report, 2007–2010*. Berkeley, CA.
Chi, M. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73–105.
Chi, M., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.
Chin, C. (2006, November). *Teacher questioning in science classrooms: What approaches stimulate productive thinking?* Presentation at the International Science Education Conference, Singapore.
Chin, C., & Osborne, J. F. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education*, *44*(1), 1–39.
Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *Journal of the Learning Sciences*, *19*(2), 230–284.
Cromer, A. (1993). *Uncommon sense: The heretical nature of science*. New York, NY: Oxford University Press.
Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.
Dillon, J. T. (1982). The effects of questions in education and other enterprises. *Journal of Curriculum Studies*, *14*(2), 127–152.
Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London, UK: Methuen.
Ford, M. J. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, *92*, 404–423.
Franke, M. L., Turrou, A. C., Webb, N. M., Ing, M., Wong, J., Shin, N., & Fernandez, C. (2015). Student engagement with others' mathematical ideas. *The Elementary School Journal*, *116*(1), 126–148. doi:10.1086/683174
Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, *16*, 41–67.
Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*, 2055–2100.
Hatano, G., & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 331–348). Washington, DC: American Psychological Association.
Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.
Henderson, B., Osborne, J., MacPherson, A., & Wild, A. (2015). Beyond construction: Five arguments for the role of critique in teaching science. *International Journal of Science Education*, *37*, 1668–1697.
Jennings, L., & Mills, H. (2009). Constructing a discourse of inquiry: Findings from a five-year ethnography at one elementary school. *Teachers College Record*, *111*, 1583–1618.
Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher*, *38*(1), 37–51. doi:10.3102/0013189X08330540
Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *The Elementary School Journal*, *102*(1), 59–80.

Kiemer, K., Groschner, A., Pehmer, A., & Seidel, T. (2015). Effects of a classroom discourse intervention on teachers' practice and students' motivation to learn mathematics and science. *Learning and Instruction*, *35*, 94–103.
King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal interaction. *American Education Research Journal*, *27*, 664–687.
King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, *31*(2), 338–368.
Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, *66*, 51–67.
Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*, 810–824.
Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education*, *61*(1–2), 21–34.
Lemke, J. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex.
Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, *39*(1), 17–38.
Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment* (CSE Technical Report No. 681). Retrieved from http://eric.ed.gov/?id=ED492885
Mercer, N. (2000). *Words and minds: How we use language to think together*. New York, NY: Routledge.
Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology*, *80*(1), 1–14.
Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Education Research Journal*, *30*(3), 359–377.
Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, *1*(1), 12–21.
Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74.
Moss, E., Cyr, C., & Dubois-Comtois, K. (2004). Attachment at early school age and developmental risk: Examining family contexts and behavior problems of controlling-caregiving, controlling-punitive, and behaviorally disorganized children. *Developmental Psychology*, *40*(4), 519–531.
National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: Division of Behavioral and Social Sciences and Education, Board on Science Education, Committee on a Conceptual Framework for New K–12 Science Education Standards.
Newton, D. P., & Newton, L. D. (2000). Do teachers support causal understanding through their discourse when teaching primary science? *British Educational Research Journal*, *26*, 599–613.
Newton, P., Driver, R., & Osborne, J. F. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, *21*, 553–576.
NGSS Lead States. (2013). *Next generation science standards: For states, by state*s. Washington, DC: National Academies Press.
Nystrand, M., Gamoran, A., Kachur, R., & Prendegarst, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York, NY: Teachers College Press.
Ohlsson, S. (1996). Learning to do and learning to understand? A lesson and a challenge for cognitive modelling. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines* (pp. 37–62). Oxford, UK: Elsevier.
Osborne, J. F. (2010, April 23). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*, 463–466.

Osborne, J., Berson, E., Borko, H., Busch, K., Zaccarelli, F., Million, S. & Tseng, A. (2015, August). *Assessing the quality of classroom discourse in science classrooms*. Paper presented at the European Association for Research in Learning and Instruction, Limassol, Cyprus.
Putnam, R., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *21*(1), 4–15.
Reddy, M. (1979). The conduit metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 284–324). New York, NY: Cambridge University Press.
Resnick, L., Asterhan, C., & Clarke, S. N. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.
Resnick, L., Michaels, S., & O'Connor, C. (2010). How (well-structured) talk builds the mind. In J. Sternberg (Ed.), *From genes to context: New discoveries about learning from educational research and their applications* (pp. 163–194). New York, NY: Springer.
Reznitskaya, A. (2012). Dialogic teaching: Rethinking language use during literature discussions. *The Reading Teacher*, *65*, 446–456.
Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford, UK: Oxford University Press.
Ryan, G., Toohey, S., & Hughes, C. (1996). The purpose, value and structure of the practicum in higher education: A literature review. *Higher Education*, *31*(3), 355–377.
Seago, N. M., Jacobs, J. K., Heck, D. J., Nelson, C. L., & Malzahn, K. A. (2014). Impacting teachers' understanding of geometric similarity: Results from field testing of the learning and teaching geometry professional development materials. *Professional Development in Education*, *40*, 627–653.
Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer* (Vol. 1). Thousand Oaks, CA: Sage.
van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, *13*(2), 155–176.
Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Education Research*, *13*, 21–39.
Weiss, I. R., Pasley, J. D., Smith, P., Banilower, E. R., & Heck, D. J. (2003). *A study of K–12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.
Wells, G. (Ed.). (2001). *Action, talk and text: Learning and teaching through inquiry*. New York, NY: Teachers College Press.
Wolpert, L. (1992). *The unnatural nature of science*. London, UK: Faber and Faber.

Appendix

Science Discourse Instrument

Stanford University

Introduction

This instrument is based in a view that one of the most effective means of supporting the development of student understanding of scientific concepts and reasoning is the opportunity for students to engage in dialogic discourse. Discourse may be about conceptual, procedural or epistemic aspects of science—that is what we know, why it happens or how we know. To that end, classroom discussions addressed by this instrument *may* focus on:

- elicitation of student ideas about a phenomenon
- organization and interpretation of data to find patterns prior to building explanations
- models or content representations that student or their teachers have created
- critiques/comparisons of others' models/content representations (including those of their peers, those in textbooks or other resources)
- attempts to use and apply science ideas and evidence previously explored to explain new scenarios/phenomenon
- efforts to make connections between concepts by synthesizing and summarizing key ideas
- investigations that students have designed or in the process of designing and their strengths and weaknesses.

A common feature of all such discussions is the potential to engage in argument. Argument, in turn, depends on there being a difference of view—either in the ideas that students present as explanations for the phenomenon of interest, their differing interpretations of the data they have collected, their differing experimental designs, or differences over the classification of living things/materials. This instrument applies to any instance that has the potential to generate a dialogic whole class or small group discussion. Whole class discussion can either be between the teacher and the whole class or may use a structure, such as think-pair-share, which involves the whole class in discussion. Small group discussions with sufficient audio can also be rated using this instrument.

This instrument is *not* designed to evaluate the importance or accuracy of particular science content that is the subject of discussion. While it is ultimately important that discussions converge on canonical/normative scientific ideas, addressing the extent to which those ideas are developed in the discussion is beyond the scope of this instrument. Evaluations of the scientific ideas developed in the discussion are necessarily specific to the particular scientific topic

Discussion Features

Feature A: Discussion Purpose
Feature B: Discourse Structures
Feature C: Referencing Discussion Norms
Feature D1: Participation
Feature D2: Participation (Gestures)
Feature E: Linking Across Relevant Contexts
Feature F: Think Time

Teacher Discursive Forms

Rubric 1: ASK (Nature of Teachers' Questioning)
Rubric 2: PRESS (Teacher Press)
Rubric 3: LINK (Teacher Linking Contributions)

Student Discursive Forms

Rubric 4: EXPLAIN/CLAIM (Nature of Students' Responses)
Rubric 5: COCONSTRUCT (Student Coconstructing)
Rubric 6: CRITIQUE (Student Challenge & Critique)

The levels within most teacher and student rubrics are organized in the following way:
Level 4: There are consistent, high quality examples of the target practice.
Level 3: There are occasional, high quality examples of the target practice.
Level 2: There was one high quality example OR multiple examples of lower quality.
Level 1: There is no evidence of the target practice.
Level 0: There is no science discussion happening in the segment.
There is a qualitative shift between low quality (Level 2) and high quality practice (Level 3).
There is a quantitative/frequency shift between occasional (Level 3) and consistent (Level 4).
Protocol Notes

- ***"Consistently"*** means that the teacher/student engages in the target practice during the majority of possible opportunities in the segment. The practice is regularly and substantially reflected in segment.
- ***"Occasionally"*** means that the teacher/student engages in the target practice during some possible opportunities in the segment, but does not do so consistently. There were some missed opportunities to engage in the practice.
- ***"Rarely"*** means that the teacher/student engages in the target practice but the majority of opportunities in the segment are missed. The practice is hardly reflected in the segment.

Segmentation Instructions

This protocol is designed for rating two 15-minute segments of video selected from a single class period of science instruction.

When selecting segments, prioritize discussion about science content rather than discussion about classroom norms or students reflecting on a prior discussion. If entire discussion is about norms or reflecting on the quality of a prior discussion, indicate purpose in Feature A and rate using 0's in the other rubrics.

To identify segments for rating, follow these guidelines:

(1) Scan the classroom video to identify instances of whole group discussion.
(2) If one, contiguous whole group discussion is identified:
    (a) *If whole group discussion is less than 15 minutes long*, end the segment at the end of the discussion. The beginning of the segment will start before the whole group discussion begins. The second segment should include other parts of the lesson during which students are talking in small groups or with the teacher.
    (b) *If whole group discussion is longer than 15 minutes and less than 30 minutes*, end the second segment at the end of the discussion and end the first segment at the beginning of the second segment.
    (c) *If whole group discussion is longer than 30 minutes*, begin the first segment at the beginning of the discussion and end the second segment at the end of the discussion.
(3) If more than one whole group discussion is identified:
    (a) *If both segments are less than 15 minutes each*, end each segment at the end of each whole group discussion. The beginning of the segment will start before the whole group discussion begins.
    (b) *If one segment is longer than 15 minutes*, decide which discussion is more productive. If the longer (>15 minute) segment is more productive, see 2b above. If the second (<15) discussion is productive, use 2a to segment both discussions.
(4) If there is not enough whole group discussion to constitute two video segments, select segments that include small group discussion with good audio of the teacher and students talking.
(5) If the entire lesson is less than 30 minutes, select one 15-minute segment from the end of the discussion moving forward.

Assign one set of ratings (one rating per rubric) to each 15 minute segment. Do not rate video segments with less than 5 minutes of discussion or classroom talk about science.

For 90 minute videos, rate 3 segments.

NA: The NA code can be used for either *all OR none* of the teacher rubrics. The NA cannot be applied selectively to some rubrics and not others. Assign NA to **all** teacher rubrics if teacher's involvement is so minimal that there is no basis for assigning rubric ratings.

Discourse Features

(The items scored below pertain only to the 15-minute segments)

Feature A: Discourse Purpose

What is the main purpose(s) of the discussion?

	Select One or Two Purposes
Share background knowledge	
Review/recap prior lessons	
Share observations	
Interpret data to find patterns	
Build explanations	
Design investigations	
Compare/evaluate claims	
Non-science reflection on discussion/norms	
Non-science topic (e.g., unrelated mathematics)	
Other ___________________	

Feature B: Discourse Structures

Which discourse structures are used by the teacher to support productive science talk?

Whole group discussion	Idea line up
Small-group discussion	Four corners
Pair discussion (including Think-Pair-Share)	

Feature C: Reference of Discussion Norms

Do the teacher and/or students make reference to classroom discussion norms?

Teacher makes meta-level comments that invoke, reference or reinforce classroom discussion norms including discussion guidelines, sentence frames, turn-taking protocols etc.

Students makes meta-level comments that invoke, reference or reinforce classroom discussion norms including discussion guidelines, sentence frames, turn-taking protocols etc. **(student references rules that guide discussion)**

Feature D1: Participation

Is there widespread student participation in teacher-facilitated discussions?

Broad participation is critical for productive scientific discussions. Students need to be given the opportunity to make regular contributions to the discussion. This rubric characterizes the degree to which there is broad individual student elaborated contributions to whole class science discussion. One-word responses constitute a contribution to discussion.

4	**More than 10** students make an individual contribution to whole group discussion.
3	**More than a few (5<x $\leq$ 10)** students make an individual contribution to whole group discussion.
2	**A few (2<x $\leq$ 5)** students make an individual contribution to whole group discussion.
1	**A couple (1 or 2)** students make an individual contribution to whole group discussion.
0	No class discussion OR class discussion not related to science.
Notes	This rubric pertains to vocalized student contributions to whole class science discussion. The use of hand gestures is noted below. If segment is predominantly small group discussion, mark as N/A.

Feature D2: Participation (Gestures)

Do students use physical gestures as a substantial form of participation in whole class discussion? (This refers to physical gestures that indicate communication; it does not refer to simply raising a hand.)

Substantial use of gestures as a form of communication Note: Gestures need to be a clear part of the classroom norms initiated by students as a form of communication. A single student or a couple of students using gestures does not constitute substantial use.	Y/N

Feature E: Linking Across Relevant Contexts

Do the teacher and/or students attempt to make contributions that link across relevant contexts?

Teacher makes an attempt to relate the discussion to relevant academic, personal, social or cultural contexts. Academic contexts refer to other school subjects or science topics other than the one being discussed, or a previous lesson.
Students make an attempt to relate the discussion to relevant academic, personal, social or cultural contexts. Academic contexts refer to other school subjects or science topics other than the one being discussed, or a previous lesson.
Note: For Students the instance must be related to personal experience or background experience

Feature F: Think Time

Does the instructional pace provide students the opportunity to think and process their ideas?

Productive science discourse depends on giving students the opportunity to think and process their ideas. By providing students with the opportunity to process their ideas, the teacher increases the potential for more productive and thoughtful student contributions. Think time can be an opportunity for students to process their ideas *before* sharing or *through* sharing with their classmates. Explicit examples of think time include:

- *explicit request by teacher for students to think before they talk*
- *explicit request by teacher for students to write ideas before they talk*
- *explicit request by teacher for students to think before they decide what corner to go to*
- *explicit comment by teacher that he/she is waiting for more students to raise their hands before anyone is called on to respond*
- *explicit request by teacher for students to talk in pairs or small groups to develop ideas before sharing to the whole group*
- *explicit comment by teacher that he/she will give the student more time to think about a question before the teacher asks the student to respond*

More than once	The teacher provides an **explicit** opportunity for think time **more than once** during the lesson.
Once	The teacher provides an **explicit** opportunity for think time **only once** during the lesson.
Not at all	The teacher **does not provide** any explicit opportunities for think time during the lesson.
Notes	If no evidence that small group talk leads to sharing in a larger group (whole group) then small group talk does not count as think time

Teacher Discursive Forms

Rubric 1: ASK (Nature of Teachers' Questioning)

Do the teachers ask open-ended questions intended to elicit diverse student responses?

This rubric focuses on the nature of the questions teachers pose in an effort to generate productive science discourse. Productive questions are open-ended and sufficiently puzzling that they cognitive engage students and lead to diverse student responses. While productive questions are often "why" or "how" questions, they can be any question about important scientific ideas in the lesson that is not "locally certain" (i.e., the answer is not obvious to students in the classroom). This rubric characterizes the degree to which teachers ask science questions with the intent to generate diverse views. Teachers can ask these questions throughout a discussion.

4	**Consistently** during the segment, the teacher asks **productive questions** that provide opportunities for students to elaborate and explain their reasoning about the important scientific ideas in the lesson. These questions are open-ended (typically "why" or "how") and can be answered in more than one way. Questions of this nature enable puzzlement, wonder and lead to cognitive engagement by the students. Alternatively, **a single open-ended productive question** sustains prolonged discussion or the teacher **consistently** asks the same or similar question again (or asking for more ideas about the same question) in order to elicit additional student ideas.
3	**Occasionally** during the segment, the teacher asks productive questions that provide opportunities for diverse student responses in which the students elaborate and explain their reasoning (generating discussion). OR The teacher **occasionally** asks the same or similar question again in order to elicit additional student ideas.
2	**Rarely** during the segment, the teacher asks questions that provide opportunities for diverse student responses in which the students elaborate and explain their reasoning (generating discussion). The teacher **rarely** asks the same or similar question again in order to elicit additional student ideas.
1	The teacher asks **closed or factual questions** that either elicit short, phrase-like responses or target a single expected correct answer. There are no efforts to ask students open-ended questions.
0	Class discussion was not related to science OR no class discussion.
NA	Assign NA to **all** teacher rubrics if teacher's involvement is so minimal that there is no basis for assigning rubric ratings.
Notes	Assign 3 if teacher asks question again in order to elicit more ideas even if the question is not a "why" or "how" question. Assign 3 if the questions never elicit a response that is more than descriptive If the teacher asks at least one question (or prompt) that initiates a prolonged student discussion, it should be assigned 4.

Rubric 2: PRESS (Teacher Press)

Do teachers press students to support their contributions with evidence and/or reasoning?

Teachers can elicit deeper thinking and reasoning from students by asking students clarify, elaborate, build on, provide evidence for and explain the reasoning behind their statements. Press refers to any question a teacher asks to encourage students to elaborate an idea or comment. Teachers can either press the student who made the initial comment or they can press the other students in the class to elaborate, clarify or build on the prior comment. In this way, students' ideas are pressed and explored more deeply. This rubric characterizes the degree that teachers press students to elaborate and provide evidence for their ideas.

4	The teacher **consistently** asks students in the class to **provide evidence and reasoning** for their own comments (e.g., What's your evidence? Why do you think that?) or about the comments of another student (Can anyone build on what John just said?) There are few, if any, instances of missed press, where the teacher needed to press and did not.
3	The teacher **occasionally** asks students in the class to **provide evidence and reasoning** for their own comments OR **regularly asks for clarification or elaboration** (e.g., What do you mean?) about their own comments or about the comments of another student. There are some instances of missed press (where the teacher needed to press and did not), pressing for particular responses or providing the answer.
2	The teacher **rarely** asks students in the class to **provide evidence, reasoning, clarification or elaboration** about their own comments or about the comments of another student. There are some instances of missed press, where the teacher needed to press and did not. (The teacher sometimes presses for explanations, but there are instances of missed press.) OR the teacher revoices. The teacher may press students but the press is for **short phrase-like explanations** or **memorized knowledge** in response to a student comment OR the teacher may ask students to agree or disagree without asking why OR the teacher may **revoice** what a student has said.
1	There are **no effort**s to ask students to elaborate or justify their comments or the contributions of others AND there are no efforts to revoice.
0	Class discussion was not related to science OR no class discussion
NA	Assign NA to **all** teacher rubrics if teacher's involvement is so minimal that there is no basis for assigning rubric ratings.
Notes	*If teacher's practice resembles IRE, the max score for this rubric is 2.* *Revoicing means more than simple restating. Revoicing requires some degree of paraphrasing, rewording, synthesizing, summarizing a students' comments.* *Pressing can include statements such as:* *Critiques of a student's comment* *Playing devil's advocate* *"What do you mean?"* *"Why do you say that?"* *"Can you say more about that?"* *Other requests to elaborate …*

Rubric 3: LINK (Teacher Linking Contributions)

Does the teacher connect students' ideas and positions in a way that helps build and develop the discussion?

In productive science discussions, participants evaluate and critique different ideas based on reasoning and evidence to build deeper understanding. To facilitate these practices, it is helpful to first identify the particular ideas that have been expressed by participants, then link those ideas by comparing and contrasting them. Teachers can facilitate this linking by providing opportunities *for students* to compare, contrast and critique different (two or more) ideas that arise in discussion. Summarizing key points and linking ideas from the discussion helps move the discussion forward. This rubric characterizes the degree to which teachers summarize key points in the discussion, link (or provide opportunities for students to link) those ideas by comparing, contrasting or critiquing those ideas in order to move the discussion forward.

4	The teacher **consistently** connects/links (or provides opportunities for students to connect) multiple students' contributions to each other and shows (or provides opportunities for students to show) **how ideas/positions shared during the discussion relate to each other.**
3	The teacher **occasionally** connects/links (or provides opportunities for students to connect) multiple students' contributions to each other ***and*** shows (or provides opportunities for students to show) how ideas/positions shared during the discussion relate to each other OR **summarizes what has been said to move the discussion forward.**
2	The teacher **rarely** connects/links (or provides opportunities for students to connect) multiple students' contributions to each other and shows (or provides opportunities for students to show) how ideas/positions shared during the discussion relate to each other OR **rarely summarizes what has been said to move the discussion forward** (e.g., Johnny argued that plants are non-living because they don't move. Susie thinks they are living because they need sunlight. What other evidence might help us resolve this disagreement?). The teacher may **document student comments** in a visible space (e.g., white board, chart paper).
1	Teacher does **not make any effort** to link, document or summarize student contributions.
0	No class discussion OR class discussion was not related to science.
NA	Assign NA to **all** teacher rubrics if teacher's involvement is so minimal that there is no basis for assigning rubric ratings.
Notes	*Linking can include questions that the teacher specifically asks to encourage students to compare/contrast/critique each other's ideas (e.g., What's the difference between what Mary and Bobby said?)* *"Johnny, do you agree with that?"* *Just summarizing or revoicing a student's comment is insufficient for a 3*

Student Discursive Forms

Rubric 4: EXPLAIN/CLAIM (Nature of Students' Responses)

Do students offer explanations or claims/conjectures supported by evidence?

Science involves exploring the material world as well as developing causal and mechanistic explanations for how the material world works. The explanations that are best supported by evidence are those that become accepted in the scientific community until new evidence is introduced. Productive science discussions encourage students to move from discussion based on phenomenological observations to discussion involving ideas, explanations and conjectures supported by evidence. This rubric characterizes the degree to which students' responses are about claims/explanations and the extent to which those claims/explanations are supported by evidence and reasoning.

4	Students **consistently** offer extended explanations using science ideas and reasoning appropriate to the discipline OR make claims/conjectures while citing evidence for their claims/conjectures. Students' contributions consistently involve claims supported by evidence or extended explanations.
3	Students **occasionally** offer extended explanations using science ideas and reasoning appropriate to the discipline OR make claims/conjectures while citing evidence for their claims/conjectures. Students' contributions may offer extended responses based on observations.
2	Students **rarely** make claims/conjectures and students' explanations are typically **brief, vague, incomplete or irrelevant** OR students' contributions are predominantly observational accounts without elaboration.
1	Students **do not** provide claims, explanations or observations.
0	No class discussion OR class discussion was not related to science.
Notes	*Explanations and reasoning can be expressed in the form: "I think ______ because _______" or "This must be true because _______"* Students providing evidence from the textbook (or class material) should be considered as explanation.

Rubric 5: COCONSTRUCT (Student Coconstructing)

Do students' contributions link to and build on each other to coconstruct understanding?

This rubric characterizes the degree to which students coconstruct by building on each other's ideas and asking each other questions to clarify, elaborate or extend what has been said.

4	The students **consistently** make contributions that build on other students' comments or ask other students to clarify, elaborate or extend their comments. The contributions **show how the ideas/positions shared during the discussion relate to each other** by explicitly referencing another students' comment. Evidence of this can be in the form of "I agree, I want to add because ..." "I think ..." or "I want to add to what was said ..."
3	The students **occasionally** make contributions that build on other students' comments or ask other students to clarify, elaborate or extend their comments. The contributions **show how the ideas/positions shared during the discussion relate to each other** by explicitly referencing another students' comment. Evidence of this *can* be in the form of "I agree, I want to add because ..." "I think ..." or "I want to add to what was said ..."
2	Rarely during the discussion do students make contributions that build on other students' comments or ask other students to clarify, elaborate or extend their comments. When students do reference another student's contribution they **do not show how the ideas/positions relate to each other** (e.g., "I agree with Ana.") but do not explain *why* they agree.
1	Students **do not make any effort** to link or revoice students' contributions.
0	No class discussion OR class discussion was not related to science.
Notes	See Rubric 7 for student critique. For this rubric, do not consider how MANY students engage in the practice. Do count this discourse move even if initiated by the teacher. Do count if student asks a question that builds on a discussion idea.

Rubric 6: CRITIQUE (Student Critique)

Are students offering critiques of the contributions of other students or the teacher?

Scientific argumentation depends on there being disagreement about something that is "locally uncertain." Productive science discussion involves students critiquing each other's ideas. These challenges can take the form of asking questions or expressing disagreement. This rubric characterizes the degree to which students' critique each other's (or the teachers') ideas.

4	There are **multiple instances** in which a student(s) makes a comment that challenges or critiques the ideas raised by another student in the class or by the teacher **AND specifically makes an argument why that idea is wrong**. Simply making an alternative but unrelated argument is not sufficient. The critiques must be directed at the comments of another student (or teacher), not a critique of an idea in the lesson materials.
3	There is **one instance** in which a student(s) makes a comment that challenges or critiques the ideas raised by another student in the class or by the teacher **and specifically makes an argument why that idea is wrong. OR** there are **several instances** of counterexamples related to the prior statement but do explicitly state why their statement is correct. Simply making an alternative but unrelated argument is not sufficient. The critiques must be directed at the comments of another student (or teacher), not a critique of an idea in the lesson materials.
2	There are instances in which student(s) make a comment that challenges or critiques the ideas raised by another student, the teacher or the lesson materials. **However, the students typical say that they disagree and do not specifically argue why the other idea is wrong OR they offer a counterexample that challenges the original claim**. For example, "I disagree with Ana. I think that …" (what that student thinks, not what's wrong with Ana's argument).
1	Students **do not** make comments that challenge or critiques the ideas raised by other students, the teacher or the lesson materials.
0	No class discussion OR class discussion was not related to science.
Notes	Does not refer to questions about instructions (What do we do?) or other non-science related questions.
	A counterexample is a remark that illustrates a differing view but does not directly refute the original claim.
	For 3, it must be clear how the counterexample relates to the original claim.

Problematic Observations

What were some of the problems that you noticed in the lesson?

	Present
Question was problematic (confusing, lacked meaning, false decision etc.)	
Missed opportunities to clarify misunderstanding	
Talk is not dialogic/conversational.	
No attempt to resolve or drive to consensus.	
Evidence mainly derived from textbook.	
Poor classroom management.	
Teacher did not interject to keep discussion on track and moving forward.	
Student comments are primarily observational in nature.	
Other ___________________	