

Assessing Dimensions of Inquiry Practice by Middle School Science Teachers Engaged in a Professional Development Program

Joni M. Lakin · Carolyn S. Wallace

Published online: 14 January 2015

© The Association for Science Teacher Education, USA 2015

Abstract Inquiry-based teaching promotes students' engagement in problem-solving and investigation as they learn science concepts. Current practice in science teacher education promotes the use of inquiry in the teaching of science. However, the literature suggests that many science teachers hold incomplete or incorrect conceptions of inquiry. Teachers, therefore, may believe they are providing more inquiry experiences than they are, reducing the positive impact of inquiry on science interest and skills. Given the prominence of inquiry in professional development experiences, educational evaluators need strong tools to detect intended use in the classroom. The current study focuses on the validity of assessments developed for evaluating teachers' use of inquiry strategies and classroom orientations. We explored the relationships between self-reported inquiry strategy use, preferences for inquiry, knowledge of inquiry practices, and related pedagogical content knowledge. Finally, we contrasted students' and teachers' reports of the levels of inquiry-based teaching in the classroom. Self-reports of inquiry use, especially one specific to the 5E instructional model, were useful, but should be interpreted with caution. Teachers tended to self-report higher levels of inquiry strategy use than their students perceived. Further, there were no significant correlations between either knowledge of inquiry practices or PCK and self-reported inquiry strategy use.

Keywords Scientific practices · Inquiry practices · Science teacher preparation · Program evaluation

J. M. Lakin (✉)

Department of Educational Foundations, Leadership, and Technology, 4036 Haley Center,
Auburn University, Auburn, AL 36849, USA
e-mail: jml0035@auburn.edu; joni.lakin@auburn.edu

C. S. Wallace

Department of Elementary and Early Childhood Education, Kennesaw State University,
Kennesaw, GA, USA

Introduction

Inquiry-based science teaching was originally conceived of by Schwab and Brandwein (1966) in the 1960s so that students might have opportunities to learn how scientific knowledge is generated and to participate in the practices of science. Schwab noted that traditional science teaching practices limited learning to the acquisition of a predetermined set of facts or “rhetoric of conclusions” (Schwab & Brandwein, 1966, p. 24). Possibly no other topic in science education has been discussed as much as inquiry-based teaching and learning. In general, inquiry-based science refers to students using knowledge, reasoning, and skills to conduct investigations much like scientists would. Teaching science through inquiry was a focus of two widely cited publications in the 1990s, *The National Science Standards* (National Research Council [NRC], 1998) and *Benchmarks for Scientific Literacy* (American Association for the Advancement of Science [AAAS], 1993). Based on publications by the NRC (1998), many researchers and practitioners have defined inquiry to include the following elements of student engagement: (a) exhibit curiosity and ask questions; (b) propose preliminary explanations or hypotheses; (c) conduct simple investigations; (d) gather evidence based on observations; (e) explain based on evidence; (f) consider other explanations; and (g) communicate explanations.

Research on the theory and practice of inquiry-based science has experienced a resurgence of interest in the twenty-first century (Capps, Crawford, & Constat, 2012; Jeanpierre, Oberhauser, & Freeman, 2005; Johnson, 2009, 2010; Supovitz & Turner, 2000). One aspect of the ongoing research is to identify characteristics of effective professional development programs that foster in-service science teachers’ use of inquiry-based strategies in the classroom (Supovitz, Mayer, & Kahle, 2000; Supovitz & Turner, 2000). Researchers have indicated that the following factors were the most germane for sustaining inquiry-based practices: (a) the duration of professional development activity (at least 80 h) and continuance of follow-up support (Jeanpierre et al., 2005; Johnson, 2007, 2009; Kimble, Yager, & Yager, 2006; Supovitz & Turner, 2000); (b) an increase in the teachers’ science process skill and content knowledge (Jeanpierre et al., 2005; Kimble et al., 2006; Supovitz & Turner, 2000); (c) administrative support (Johnson, 2009, 2010; Supovitz & Turner, 2000); (d) allowing teachers a role in creating the curriculum materials (Huffman, Thomas, & Lawrenz, 2003); (e) implementing professional development activities directly in the classroom context (Johnson, 2010); and (f) the establishment of a collaborative professional development community (Butler, Novak Lauscher, Jervis-Selinger, & Beckingham, 2004; Jeanpierre et al., 2005; Johnson, 2007).

Determining the impact of professional development ideally requires an array of observation and assessment data to explore the degree and quality of inquiry practices that occur in the classroom. Supovitz and Turner (2000) developed the idea of using self-report survey data as a means of quantifying inquiry teaching strategies. Although every form of assessment has limitations, self-reporting by questionnaire is a convenient way for researchers to obtain information on teachers’ use of inquiry in the classroom. The current study continues this vein of research

through its investigation of a battery of research instruments used in the context of a professional development program to promote inquiry. The purpose of this study was to investigate the validity of two self-report measures for teachers' frequency of the use of inquiry-based strategies. We investigated validity, first, by comparing and contrasting the results obtained by each of these two instruments with each other. We further inquired into the validity of the two instruments by studying the relationships between these instruments and other related scales that could provide evidence that the teachers were in fact using inquiry as they reported and creating high-quality inquiry experiences. These additional scales included teachers' preferences for using inquiry, students' reports of inquiry use, teachers' knowledge of inquiry practices, and teachers' pedagogical content knowledge for using inquiry.

Inquiry and Scientific Practices

Schwab's original concern that students be engaged in inquiry-based practices continues to be a paramount goal for science educators and policy makers as evidenced by recent US national curriculum standards documents, including the *Next Generation Science Standards* (NGSS; NRC, 2013) and the *Framework for K-12 Science Education* (NRC, 2012). The Framework document, which outlines theory and research that buttresses the standards, includes *scientific and engineering practices* as one of its three key dimensions for science learning. Scientific practices include the skills, reasoning abilities, and content knowledge that are necessary for students to engage in investigations about the natural world. The new term, *scientific practices*, was chosen, in part, to help clarify what is meant by inquiry-based science. Authors of the Framework (NRC, 2012) assert that:

Engaging in the practices of science helps students understand how scientific knowledge develops; such direct involvement gives them an appreciation of the wide range of approaches that are used to investigate, model, and explain the world... The actual doing of science or engineering can also pique students' curiosity, capture their interest, and motivate their continued study; the insights thus gained help them recognize that the work of scientists and engineers is a creative endeavor; one that has deeply affected the world they live in. (pp. 41–42)

Given the continued priority for students to engage in inquiry-based practices, research on how teachers think about and implement these practices continues to be of significance to the science education community. Unfortunately, studies show that many in-service and pre-service teachers have developed naïve or incorrect conceptions of inquiry-based teaching and these remain pervasive in US school systems (Capps et al., 2012; Pruitt & Wallace, 2012; Seung, Park, & Jung, 2014; Supovitz & Turner, 2000). For example, Seung et al. (2014) found that both pre-service teachers and their mentors had difficulty connecting appropriate inquiry features to video-taped teaching episodes and tended to define inquiry too broadly. Pruitt and Wallace (2012) found that a state department intervention featuring inquiry-based coaching did little to improve students' science achievement scores. In this study, even when working one-on-one with experienced teacher mentors,

teachers in high-needs schools were not willing or able to use inquiry to a degree that resulted in student science achievement as measured on inquiry-oriented end-of-course examination. Thus, more studies of conditions surrounding successful inquiry implementation and its measurement are warranted.

The impetus for our research was our role as external evaluators for the “NanoBio Partnership for the Alabama Black Belt” (or “the NanoBio MSP”), a program funded as an NSF Math and Science Partnership Grant. The focus of the program is to provide teacher professional development in inquiry-based pedagogical skills to promote greater student achievement and motivation in science. A major focus of the professional development activities is the introduction of inquiry-based instructional modules (a series of classroom activities) that have been developed by university science faculty to introduce cutting-edge areas of science, specifically science at the nanometer scale, as a means of increasing science interest and achievement in middle school students. The partnership includes university science faculty, education faculty, and administrators and teachers from nine low-performing school districts in low-income areas of Alabama. In Alabama, students study earth and space science in sixth grade, life science in seventh grade, and physical science in eighth grade. Each module relates to one or more grade-level standards.

The instructional modules were intended to relate to state standards at the relevant grade levels and to provide enrichment about current, exciting developments in science. Importantly, the modules are also designed to promote the use of inquiry in the classroom and include hands-on activities with guided exploration of the concepts. The modules, along with the professional development training for the modules, were expected to increase the amount of time that teachers used inquiry-related teaching practices in their classroom. Thus, to evaluate the impact that the program had on inquiry use in the classroom, we adapted or developed several instruments for this project that measured teachers’ use of inquiry strategies.

Rationale for the Study

Methods of measuring teachers’ beliefs about, knowledge of, plans to implement, and actual practices of inquiry have received much attention in the literature (Bodzin & Beerer, 2003; Lumpe, Haney, & Czereniak, 2000; Trumbell, Scarano, & Bonney, 2006). Assessing the impact of professional development on teachers’ beliefs, knowledge, and practices presents a number of methodological issues because measures can vary widely based on the personal interpretation of the observers. Possibly the most accurate means of integrating several sources of evidence is the use of extended case study observations. Extended case study observations allow researchers to use a variety of in-depth qualitative methods, such as interviews and classroom observations, to compare stated beliefs, knowledge, and self-reports of practice with observed classroom practices (Lotter, Harwood, & Bonner, 2007). However, the number of teachers that can be studied is very limited with an in-depth case study method.

The use of classroom observation protocols, in which researchers rate inquiry-based implementation, is also commonly used (Bodzin & Beerer, 2003; Luft et al.,

2011). This method can increase the number of teachers whose practices are studied, but it lacks the dimension of teachers' own intentions or beliefs about their actions. Classroom observation protocols, such as the Science Teacher Inquiry Rubric (Bodzin & Beerer, 2003), also may be singling out a specific aspect of inquiry-based practice, such as whether instruction is more teacher-centered or more student-centered, and as such yield little information about other aspects, such as using class discussion as a basis for forming understandings.

Due to these constraints, several researchers have identified self-report as a valuable tactic for gathering a wide variety of data about the impact of professional development on inquiry-based teaching (Lee, Hart, Cuevas, & Enders, 2004; Mullens et al., 1999; Supovitz & Turner, 2000). The main issue faced with self-report, however, is the problem of response biases. These biases include social desirability, the desire to impress others favorably, and impression management—the urge to present impressions that are congruent with what one wishes to convey to the public, whether conscious or subconscious (Podsakoff, McKenzie, Lee, & Podsakoff, 2003; Randall & Fernandes, 1991). Naive concepts of what inquiry entails may also contribute to over-reporting. These biases could result in self-report measures of inquiry strategy use that are higher and less variable than what someone else might observe in the classroom.

The purpose of this investigation was to further refine and validate two measures of teachers' self-report of the frequency with which they use inquiry-based strategies. To investigate the validity of the two self-report instruments, we studied the relationships between them, as well as their relationships to other dimensions of the teachers' knowledge and the use of inquiry. These additional dimensions included: (a) teachers' self-reported preferences for inquiry strategy use; (b) teachers' knowledge of inquiry practices; (c) agreement between teachers' and students' reports of inquiry use; and (d) measures of teachers' pedagogical content knowledge (PCK; Shulman, 1987) for using inquiry strategies.

To our knowledge, the relationship between teachers' self-reports of inquiry use and pedagogical content knowledge for using inquiry strategies has not been investigated previously. We chose to investigate the construct of pedagogical content knowledge because, in addition to measuring the teachers' own knowledge of inquiry strategies (hypothesizing, analyzing data, etc.), we also wanted to know whether they could translate these practices into teachable forms. Further, we wanted to investigate whether there was a correlation between the teachers' knowledge of teaching with inquiry and their self-reported frequency of the use of inquiry. If teachers have strong preferences for using inquiry and report frequent usage of strategies, but show low pedagogical content knowledge for inquiry, this would indicate the need for more professional development on topic-specific inquiry-based science. Further, it may signal a mismatch between the reported frequency of strategy use and effective inquiry-based teaching. If, however, frequency and/or preferences for using inquiry strategies are low and pedagogical content knowledge is high, then a different type of professional development on the benefits of inquiry-based learning would be warranted.

Theoretical Frameworks: Pedagogical Content Knowledge and the Five E Learning Cycle

We chose to draw on Shulman's (1987) seminal model of the types of knowledge needed for teaching. Shulman described what he viewed as the diverse and necessary knowledge bases for teaching. One of these knowledge bases, pedagogical content knowledge (PCK) is viewed as Shulman's unique contribution to the theory of teacher thinking and has garnered much attention in the literature. Shulman defined PCK as "the particular form of content knowledge that embodies the aspects of content most germane to its teachability" (p. 9) or the specialized knowledge that teachers develop for the effective teaching of particular topics in school science (Shulman, 1987). Developing PCK in regard to inquiry-based science teaching would mean that teachers would activate and grow their knowledge bases for using inquiry-based pedagogical methods for a range of science topics that they normally teach. A PCK framework continues to be widely recognized and researched by science teacher educators studying teaching (Alonzo, Kobarg, & Seidel, 2012; Avraamidou, 2013).

The pedagogical model for the program is based on the well-known *Five E Learning Cycle* (Bybee et al., 1989). It was originally conceived of by Karplus (1979), has undergone several revisions, is used in multiple forms, and is highly supported by research studies of its effectiveness (Goldston, Dantzler, Day, & Webb, 2013). The 5E model is strongly aligned with the NRC's (1998) framework of inquiry, but is only one example of many applications of inquiry practice. During the *engage* phase, the teacher motivates the students for the investigation and discerns their prior knowledge. In the *explore* phase, students explore the scientific phenomena to clarify their prior knowledge and become open to new learning. In the *explain* phase, students make sense of their data from the explore phase and work with guidance from the teacher to build the scientifically correct concept. During the *elaborate* phase, students have the opportunity to solve a new problem or apply the newly learned concept. Students are assessed in the *evaluate* phase (Bybee et al., 1989; Olson & Loucks-Horsley, 2000). Space precludes an in-depth explanation of the various theories underlying this model, but it has roots in the individual constructivist-based theories of Piaget (Karplus, 1979) and in the social constructivist theories of learning developed by Vygotsky (Goldston et al., 2013).

Research on Measuring Inquiry or Scientific Practices

Attempting to quantify and qualify teachers' use of inquiry-based teaching has a long research history (Bodzin & Beerer, 2003; Lumpe et al., 2000; Minner, Jurist Levy, & Century, 2010; Welch, Klopfer, Aikenhead, & Robinson, 1981). Supovitz and Turner (2000) reported that much was learned about the validity and reliability of self-report survey data regarding inquiry in the 1990s. Considering several studies in math and science education, they concluded that in general, survey data had moderate to strong validity with regard to aligning with observations of classroom practice, teacher logs, and student work samples. Bodzin and Beerer

(2003), however, found that classroom observation and teacher self-report using the Science Teacher Inquiry Rubric (STIR) yielded a modest correlation of 0.58.

Mullens et al. (1999) analyzed teacher surveys on classroom practice relative to classroom observation and teacher logs. Overall, they found remarkable consistency between the measures, indicating that self-report is a valid means of assessing classroom practice. However, one key discrepancy was found between teacher surveys and researcher logs for some classrooms: Activities that were highly emphasized by professional standards at the time (including the use of graphics to demonstrate concepts and cooperative learning activities) tended to be over-reported in teacher surveys. They interpreted this as evidence of social desirability bias, which impacts general ratings of frequency more than relatively objective logs of specific occurrences (Podsakoff et al., 2003). Items clearly related to modern/reform or "old-fashioned" teaching practices were especially prone to these biasing effects in Mullens et al. study. Thus, validity of self-report for reform-based practices such as inquiry may be lower than some of the early 1990s studies of self-report might indicate.

A series of studies by Lee et al. (Lee, Deaktor, Hart, Cuevas, & Enders, 2005; Lee et al., 2004; Lee, Penfield, & Maerten-Rivera, 2009) utilized several measures of elementary teachers' knowledge of and beliefs about science teaching, including inquiry-based pedagogy. In Lee et al. (2004), researchers used a combination of self-report survey data, classroom observations, and focus group interviews to measure change in teachers' opinions on the importance of science content, discourse and inquiry-based knowledge, and their practices of inquiry and discourse through classroom observations. The researchers were investigating the impact of a large-scale professional development program on teacher change in diverse classrooms. Self-reports of the importance of inquiry, knowledge of inquiry, and more objective measures of the teachers' actual classroom practice indicated that, both before and at the end of the intervention, teachers showed high self-report measures (often above 4.0 on a 5-point scale) for the importance and knowledge of inquiry. In contrast, objective reports of their classroom practice were lower (usually between 2.5 and 2.9 on the same scale). Teachers seem to rate their own knowledge and beliefs in inquiry more highly than they were rated in practice. Lee et al. did not report correlations between the scales, so it is unknown whether teachers' self-ratings were associated with objective ratings, even though the absolute ratings differed.

In a later, related study (Lee et al., 2009), Lee's research group found that teachers' self-report of inquiry beliefs and knowledge had leveled out around 3.0 (out of 5), and there was greater alignment between self-report and observation data, perhaps indicating that greater knowledge of inquiry led to more accurate self-reports. Although student science achievement had increased overall throughout the years of this same professional development program, there was no significant relationship between fidelity of implementation (the degree to which the teachers adhered to the professional development materials) and student achievement. Lee et al. (2009) suggested that this finding indicates the difficulty of accurately measuring relationships among beliefs, knowledge, degree of implementation, and actual practice.

The Current Study

An important short-term outcome of the NanoBio MSP project was evidence that teachers were increasing their use of inquiry-related pedagogical strategies in their science classrooms. Similar to the studies discussed before, the NanoBio MSP project attempted to measure teacher change with regard to increases in beliefs in, use of, and knowledge about inquiry-based teaching. In order to meet our own needs to use self-report to collect data about beliefs in, uses of, and knowledge of inquiry, we created or adapted five scales that reflected contemporary thinking about the salient aspects of inquiry: two that measured the use of inquiry strategies, one that measured the knowledge of inquiry practices, one that measured teachers' preferences for using inquiry, and one that measured the inquiry-related PCK.

To assess changes in the use of inquiry-related strategies, we developed two self-report measures of inquiry use in the classroom to detect the impact of professional development and the instructional modules on classroom practice. One measure, the *Inquiry Strategies Scale* (IS), simply listed common instructional strategies with inquiry and non-inquiry strategies intermixed. The second measure, the *5Es Inquiry Scale*, explicitly reflected the 5Es model, asking about the frequency with which teachers used various specific 5E strategies in their classroom.

The study focused on the validity of the Inquiry Strategies Scale and the 5Es Inquiry Scale for the purposes of evaluating teachers' use of inquiry methods and orientation in the classroom. We also investigated correlations between self-reports of inquiry use and teachers' preferences, knowledge of inquiry practices, and PCK for inquiry. The following validity-related research questions were asked:

1. *Dimensionality* Do the instruments yield patterns of inquiry-based strategy use, and if so, how can these be described?
2. *Convergence* Are teachers who express greater use of 5E strategies more likely to use specific inquiry strategies on the other assessment?
3. *Predictive evidence* Do teachers' preferences for inquiry-focused science practices, their pedagogical content knowledge for inquiry (PCK), or knowledge of inquiry practices correlate with how frequently they report using inquiry strategies?
4. *Agreement of multiple informants* Do students' average ratings of a teacher's use of strategies (as measured by the IS scale) agree with the teacher's self-report?

Given that the purpose of these instruments was to reflect classroom practices and beliefs, questions one and two explored the dimensionality of the measures: whether either behaved according to expectations and whether the two measures are redundant. Question three focuses on the predictive value of the measures—are the reports of strategy usage correlated with knowledge about and attitudes toward inquiry? Clearly, a useful measure of strategies will correlate with other measures of knowledge about and attitudes toward those strategies, if these strategies are being implemented effectively and with the intention of promoting inquiry. Question four

provided another view of teachers' inquiry use from their students to provide evidence of possible biased reporting by teachers. The third question was of most interest to us from an evaluation perspective, because the sensitivity of the strategy use measures to differences in teacher preferences or knowledge of inquiry is key to its use for measuring the impact of professional development activities.

Methods

The population studied consisted of approximately 90 middle school science teachers in the Black Belt region of Alabama (a historically poor region with a traditionally agricultural economy) participating in an educational intervention targeted to sixth- through eighth-grade science teachers. The intervention consisted of professional development workshops during the year, a summer professional development conference (Summer Institute for the South Eastern Consortium for Minorities in Engineering [SECME], Jeffers, Safferman, & Safferman, 2004), and the availability of new curriculum materials in the form instructional modules. The themes of the workshops and modules were scientific research at the nanometer scale and inquiry-based science instruction.

Participants

Of the participants in the program, 60 completed all or some of the baseline assessments administered in spring of 2012 before the start of the intervention (12 completed assessments in the summer of 2013 when their school district joined the program). To understand the participation rate, it is important to note that decisions to participate in the program were (in theory) made at the district level, but principals and teachers varied substantially in their participation and compliance with the program and evaluation activities.

All teachers who responded taught some combination of sixth-, seventh-, and eighth-grade science. Sixth-grade teachers were mixed in terms of whether they specialized in science (in a departmentalized middle school) or whether they were general classroom teachers covering all subjects. All seventh- and eighth-grade teachers were in science departments in middle or high schools.

Pretests were administered online in spring and summer of 2012 before teachers began any portion of the professional development interventions. All participants in the sample were administered the Inquiry Strategies Scale and the 5Es Inquiry Scale. They also completed a short multiple-choice test of inquiry practices (using inquiry-based reasoning, such as identifying a hypothesis, the benefits of replication). Teachers were randomly assigned to complete either the POSITT-F measure of orientation toward science and inquiry or an objective test of inquiry-focused PCK. The measures are described in more detail in the section that follows.

We analyzed pretest data because it would offer an ideal opportunity to explore the psychometric qualities of the scales. In particular, we were interested in the presence of a scale *ceiling*, meaning we were concerned that at pretest teachers might give uniformly high ratings of their skills and teaching methods. High scores

at pretest might reflect a response bias or a lack of variability that would inhibit our ability to detect program effects at posttest.

In addition to teacher surveys, some information from students was available as well. Teachers were asked to administer the survey to one section of each grade level they taught (at most three classes for a teacher who has sixth-, seventh-, and eighth-grade sections). This resulted in 10–47 student surveys for 46 of the teachers or a total of 1,085 students. Although the student survey consisted of a wide range of measures, the only scale relevant to this study was the student version of the IS scale which asked students to rate the use of strategies by their science teacher (the format was identical to the teachers' survey). Students' average rating for each teacher was used in the analyses.

Measures

With the exception of the POSITT-F, all assessments were developed or adapted from prior projects to be appropriate for the focus of the current project. Experts on the research and evaluation team contributed to the development of these instruments.

Measures of Inquiry Use

A scale measuring the use of inquiry strategies (Inquiry Strategies Scale, IS) in the classroom was developed that asked teachers about the frequency with which they used a range of inquiry- and non-inquiry-related practices. The original scale was developed and researched by the NSF MSP-Motivation Assessment Program (MSP-MAP; Karabenik & Maehr, 2007) for mathematics teachers and adapted for our purposes by rephrasing items to refer to science instruction and removing a few items that were specific to mathematics. The Likert-type scale had five anchor points: never, rarely (a few times a year), sometimes (once or twice a month), often (once or twice a week), all or almost all science lessons. Tables 2 and 3 show all survey questions. In addition to the 11 inquiry items on the IS scale, another three items were included to measure non-inquiry strategies ("Listen to a lecture about science," "Copy notes or problems off the board," and "Memorize formulas and facts for a test or quiz").

For the purposes of the project, and in alignment with the professional development activities provided by the program, we developed a second scale based on the 5E learning cycle, a widely accepted constructivist framework for the inquiry-based science classroom (Bybee et al., 1989). We began with professional development materials for teachers that described best practices for inquiry instruction and developed and refined three items that exemplified the range of behaviors associated with each of the 5Es. After review by experts on the research and evaluation team, the result was 15 survey items that asked teachers how well the statements described their classroom (five-point Likert type from "Not at all true" to "Very true"). Total scores for the IS and 5Es scales were created by taking the average of teacher responses across each scale.

Inquiry Practices Assessment

The inquiry practices assessment addressed inquiry reasoning strategies such as formulating hypotheses, creating appropriate experimental designs, and data interpretation. Some items from the scientific practices section were drawn from released items from the National Assessment of Educational Progress (NAEP); others were developed by a science education faculty member (the second author). The 13 test items were reviewed by educational measurement specialists and refined to balance the answer key, minimize guessing, and check alignment to evaluation goals. The test centered on three vignettes about science problems (a child exploring soil erosion, a student's laboratory experiment about surface area and dissolving salt, and a commercial test of nanomaterials and water resistance) with three to four questions focused on identifying the hypotheses, understanding of the use of control conditions, and reporting of data. The test was scored with a single correct answer for each item. Science and education faculty reviewed the questions and best answers for accuracy and relevance to the target construct (knowledge of scientific inquiry).

Pedagogical Content Knowledge Measure

A measure of pedagogical content knowledge (PCK) was also developed for the project. This assessment specifically measured knowledge about the use of inquiry methods in a science classroom. This assessment was developed in collaboration between assessment and content experts on the research and evaluation teams. The PCK assessment consisted of seven multiple-choice items where teachers identified examples of using inquiry methods in the classroom, with distractors reflecting poor or non-examples of inquiry-focused pedagogy. This test was scored with a single correct inquiry answer from the four answer choices.

Pedagogy of Science Inquiry Teaching Test

The POSITT-F (Schuster et al., 2007) was developed to assess teachers' orientations toward inquiry versus directive pedagogical methods. The assessment consists of 16 vignettes describing the use of inquiry-related methods in a science classroom. They then pose a question ranging from asking how the teacher should proceed with a lesson or how the reader would conduct the class differently. Each item consists of four answer choices reflecting direct didactic, direct active, guided inquiry, and open discovery teaching approaches. The measure is intended to reflect attitudes or preferences and is not scored right/wrong. For our purposes, the two directed methods were classified as reflecting direct (teacher-centered) instructional orientation, while the latter two inquiry answer choices indicated a student-centered instructional orientation. We then created a total score for each teacher that represented the percent of discovery or inquiry-focused responses that were selected relative to the number of direct instruction responses.

Data Analysis

All analyses were conducted by SPSS. Research question one (What are the patterns of strategy use by teachers?) was addressed using exploratory factor analysis (EFA). In SPSS, we used the maximum likelihood estimator and varimax (orthogonal) rotation to estimate factors. We explored multiple criteria (eigenvalues, scree plots, and theory) for selecting the optimal number of factors to retain. Questions two through four were addressed with analyses of correlations and independent samples *t* tests.

The sample size in this study is small for a factor analysis approach. However, the factor structures we explored were relatively simple (2–3 factors per scale) and with relatively few items. Costello and Osborne (2005) discussed current and best practices in EFA. Our sample-size-to-item ratios are 4:1 for the IS and 3:1 for the 5Es scales, which is consistent with common practice in the validity literature (more than 40 % of the surveyed literature had ratios similar to or smaller than this), though it is below the standard 10:1 ratio that is preferred (Costello & Osborne, 2005). Future researchers should consider replicating our factor analysis to explore whether our results generalize to other samples. We followed Costello and Osborne's guidelines on maximizing the replicability of EFA results (e.g., only factor loadings above 0.4 were considered important).

Results

Initial analyses focused on the basic psychometric properties of the measures. We were concerned about the presence of ceiling effects and low variability in our measures, because of the potential for social desirability bias, especially for reporting inquiry strategies. One criterion of ceiling effects is that there should be at least two standard deviations between the mean score and the maximum score (Bracken, 2007). Table 1 shows the descriptive statistics for each of the scales in

Table 1 Descriptive statistics

Scale	<i>N</i>	Items	α	Max	<i>M</i>	SD	Ceiling ^a
5Es total	44	15	0.91	5	4.06	0.66	1.4
Inquiry strategies	44	11	0.93	5	3.5	0.82	1.8
Non-inquiry strategies	44	3	0.68	5	3.36	0.74	2.2
POSITT-F inquiry answers ^b	48	16	– ^c	16	5.94	6.22	1.6
POSITT-F didactic answers ^b	48	16	– ^c	16	1.96	2.54	5.5
Knowledge of scientific practices	45	13	0.95	13	4.33	4.89	1.8
Pedagogical content knowledge ^b	21	7	0.85	7	2.67	2.46	1.8

^a Ceiling effect calculated as maximum score on the scale minus mean score, divided by SD. Value \geq 2, no ceiling effect

^b Forms were spiraled so that teachers completed either the POSITT-F or the PCK test

^c Because of the manner of scoring, internal consistency estimates are not appropriate

Table 2 Factor solutions for inquiry strategies

	One-factor solution		Two-factor rotated solution	
	1	2	1	2
Participate in small groups to make sense of science	0.515		0.429	0.264
Give an explanation about how they solved a science problem or designed an experiment	0.715		0.429	0.612
Use technology (e.g., probes, cameras) to learn science	0.626		0.266	0.728
Explain concepts to me or to another student	0.806		0.618	0.509
Design and conduct their own experiment or scientific investigation	0.639		0.314	0.697
Do science projects or investigations that take several days to complete	0.693		0.321	0.778
Apply science situations to life outside of school	0.811		0.826	0.260
Discuss alternative explanations for a question or problem	0.864		0.880	0.284
Explore questions created by students	0.860		0.773	0.410
Analyze data using charts, tables, or graphs	0.907		0.734	0.510
Write about the results of scientific investigations	0.821		0.658	0.473

Factor loadings above 0.4 are bold-faced

Table 3 Factor solutions for 5Es survey

	One-factor solution		Two-factor solution	
	1		1	2
My goal in class is to encourage students to get involved in science problems like a scientist would	0.510		0.200	0.517
When we start a new instructional unit, it is important to understand students' prior knowledge when planning my instruction	0.505		0.033	0.754
When we start a new instructional unit, I try to engage my students in the topic by using a demonstration activity or example	0.799		0.418	0.774
When a student brings up an interesting question or problem, I sometimes change the class plan to allow time to explore that question	0.697		0.717	0.263
I encourage my students to follow up an interesting question or problem by performing their own investigation	0.618		0.262	0.623
I often let students have a hands-on experience with scientific phenomena before introducing vocabulary words	0.772		0.723	0.356
I use questions to probe students' understandings of a concept and to adapt my instruction	0.636		0.179	0.768
I use discussions in class to promote learning about concepts	0.704		0.474	0.497
I let students develop explanations from prior knowledge and/or hands-on activities before I fill in the gaps myself	0.776		0.844	0.265
I choose problems to solve in class that can have many solutions	0.710		0.793	0.198
I use real-life examples of scientific principles in my class	0.628		0.300	0.585
After science concepts are explained, I allow students time to develop their understanding through additional examples and problems	0.664		0.499	0.395
I use activities such as concept maps, drawings or problems to further develop students' scientific concepts	0.561		0.338	0.438
When I assess my students' knowledge, I look for common misconceptions about science concepts	0.632		0.562	0.319
A large part of my students' grades are based on products or performance activities	0.286		0.358	0.038

Factor loadings above 0.4 are bold-faced

addition to this ceiling effect index. The 5Es scale showed a somewhat low score ceiling, meaning that teachers tended to give high ratings to their use of these strategies that approached the maximum scale point. The Inquiry Strategies Scale did not show as much of a ceiling effect. The measure of inquiry practices knowledge fell just short of the 2SD criteria, while the POSITT-F score showed some ceiling.

Cronbach's α estimates of internal consistency indicated that responses were consistent for the 5Es and Inquiry Strategies scales (α s above 0.9) as well as the other scales. The scale reflecting non-inquiry strategies was less consistent ($\alpha = 0.68$). This low consistency may reflect the fact that the scale only has three items that are quite distinct instructional practices.

Patterns of Strategy Use

How many factors does the Inquiry Strategies Scale yield? Exploratory factor analysis was used to uncover the relationships between items on the two scales. Although the IS and 5Es scales are correlated (see later section), they did not load together in a combined factor analysis. As a result, the two scales were factor-analyzed independently. Likewise, the non-inquiry strategies correlated with the IS scale but did not load with inquiry items and were analyzed separately as a single total score.

For the IS scale, we explored multiple criteria (eigen values, scree plots, and theory) for selecting the optimal number of factors to retain. There were two factors with eigen values >1 (the Kaiser criterion) that accounted for 71 % of item variance. This solution (shown in Table 2) is interesting because it includes a factor (factor 1) that reflects practices related to exploring results as well as interpreting and applying findings ("Discuss alternative explanations for a question or problem" and "Apply science situations to life outside of school"). The second factor reflected the use of more formal scientific practices, like multiday investigations and the use of technology.

Although the additional factors have theoretical interest, the most parsimonious and reliable solution was a single-factor solution, accounting for 61 % of the item variance and showing strong item-factor loadings for almost all items. Therefore, we conclude it is reasonable to treat the IS scale as unidimensional scale and use a single composite score for the remaining analyses. The Cronbach's α for this scale ($\alpha = 0.92$) also confirms that a single dimension is plausible.

How many factors does the 5Es Inquiry Scale yield? The 5Es Scale formed two factors with eigen values >1 (see Table 3). Once rotated, the two factors accounted for about 25 % of the item variance each (total 56 %). Larger numbers of factors were explored (including a five-factor solution which the scale design would predict), but did not yield meaningful factors. In the two-factor solution, our interpretation of the first factor was that the items indicate the use of assessments and teaching strategies that explore concepts, probe misconceptions, emphasize student performances, emphasize real-life applications, and allow for multiple solutions. Thus, we characterize this factor as practices reflecting *open-ended explorations and assessments*. The items for the second factor reflect activities in

Table 4 Independent samples t tests of differences between teachers expressing high and low agreement with 5E items

	Low ($N = 15$)		High ($N = 30$)		d	t ($df = 43$)	Sig. ^a
	M	SD	M	SD			
<i>Inquiry strategies</i>							
Participate in small groups to make sense of science	3.50	1.22	3.93	1.01	0.39	1.23	0.22
Give an explanation about how they solved a science problem or designed an experiment	3.27	1.10	3.73	1.01	0.44	1.41	0.16
Use technology (e.g., probes, cameras) to learn science	3.20	1.15	3.53	1.22	0.28	0.88	0.38
Explain concepts to me or to another student	3.60	0.99	3.97	0.96	0.38	1.19	0.24
Design and conduct their own experiment or scientific investigation	2.53	0.74	3.13	1.07	0.65	1.94	0.06
Do science projects or investigations that take several days to complete	2.67	0.98	3.20	1.00	0.54	1.70	0.10
Apply science situations to life outside of school	3.13	1.30	3.90	1.06	0.65	2.12	0.04
Discuss alternative explanations for a question or problem	3.20	0.86	3.87	1.01	0.71	2.19	0.03
Explore questions created by students	2.93	0.88	3.70	0.99	0.82	2.54	0.01*
Analyze data using charts, tables, or graphs	3.47	1.13	3.93	0.87	0.46	1.54	0.13
Write about the results of scientific investigations	2.67	0.82	3.73	1.01	1.16	3.53	<0.01*
<i>Non-inquiry strategies</i>							
Listen to a lecture about science	3.33	0.90	3.27	0.74	-0.08	-0.27	0.79
Copy notes or problems off the board	3.40	0.63	3.40	1.04	0.00	0.00	1.00
Memorize formulas and facts for a test or quiz	2.93	1.03	3.60	1.10	0.62	1.95	0.06

d Cohen's d mean effect size

* Significant based on Hochberg's step-up procedure (1988), which sets an adjusted α for each test and controls the family-wise type I error rate

^a One-tailed significance test used

class that allow students to engage in activities like a scientist would, activate prior knowledge, and follow up on their questions: in other words, activities *facilitating inquiry* by initiating learning and allowing flexibility in instruction.

Although the two-factor solution yielded two interpretable factors, many items showed relatively strong loadings on both factors. Thus, a single-factor solution was identified as the best solution because it accounts for nearly 42 % of the item variance, and all items (except the final item) show strong factor loadings. Either solution is reasonable. The Cronbach's α for this scale ($\alpha = 0.95$) also indicates that a single dimension is preferable.

Relationships Between the Scales

The IS and 5Es measures of strategy use correlated moderately ($r = 0.604$, $p < .01$). While this correlation is moderately high (reflecting a 36 % shared variance), it also indicates significant variation in how teachers respond to the different items on the two scales.

Another interesting way at looking at how the two scales relate to each other was to split the sample based on how much the teachers follow the 5Es approach to instruction and see whether this was associated with differences in their use of the strategies on the IS scale. Therefore, we performed a mean-split of the sample based on their use of 5Es strategies (e.g., creating groups of high and low users of the 5Es) and conducted t tests on each IS item. We found differences in the frequency with which teachers used several different specific strategies from the IS scale (see Table 4). The greatest differences in strategy use were writing about results [$t(43) = 3.53$, $p < .01$, $d = 1.2$, large effect] and exploring student questions [$t(43) = 2.54$, $p < .05$, $d = 0.8$, large effect]. These two strategies were much more commonly used among teachers with high scores on the 5Es scale than among teachers with low 5Es scores. This was also true, though not significant in this sample, for most other strategies surveyed. Among a list of instructional strategies that can be used for a variety of purposes (not just inquiry), these types of items may

Table 5 Correlations between scales

	5Es	IS	NIS	POSITT-F		Inquiry knowledge	PCK
				Inquiry	Didactic		
5Es Scale	1						
Inquiry strategies	0.604**	1					
Non-inquiry strategies	0.417*	0.164	1				
POSITT-F inquiry	0.361*	0.077	0.184				
POSITT-F didactic	0.138	-0.073	0.014	0.592**			
Knowledge of practices	-0.238	-0.076	0.191	—	—	1	
				0.354*	0.326*		
PCK	0.019	-0.272	0.157	- ^a	- ^a	0.880**	1

* $p < .05$; ** $p < .01$

^a Teachers were randomly administered the PCK test *or* the POSITT-F, so there is no overlap in samples

be the most sensitive to a teacher's inquiry orientation and the use of inquiry in the classroom. Differences were seen on other scales, but were not considered significant because of the use of a conservative α . Intriguingly, high-use teachers reported greater use of a number of strategies, including those not associated with inquiry (the NIS scale).

Relationship of IS and 5Es to Measures of Inquiry Knowledge, Preference, and PCK

Table 5 shows the correlations between the two measures of strategy use and other variables. The 5Es scale correlated moderately with POSITT-F Inquiry scores ($r = 0.361, p < .05$), indicating significant overlap between teachers who use more 5E methods and teachers who prefer more inquiry-oriented than didactic instructional strategies. The Inquiry Strategies Scale correlated weakly with POSITT-F scores, indicating it did not overlap with the POSITT-F scale as the 5Es scale did (see Table 5). This may indicate that the 5Es scale is more reflective of inquiry practice in the classroom than the IS scale, if indeed teachers with a stronger preference for inquiry strategies (as measured by POSITT-F) use more inquiry strategies.

The two measures of knowledge (inquiry practices and PCK) reveal more complex patterns. Recall that *knowledge of inquiry practices* requires identifying elements of inquiry in science content questions developed for middle school students (e.g., what is the hypothesis being tested?) and *PCK* requires identifying an inquiry-related method of addressing an instructional goal. Although knowledge of inquiry practices and PCK correlate very strongly with each other ($r = 0.88$), knowledge of inquiry practices correlated negatively and significantly with preferences for using inquiry, POSITT-F scores.¹ This seems to indicate that, although the knowledge scales measure some aspect of inquiry practices knowledge reliably, teachers with more knowledge of inquiry have more negative attitudes about using inquiry in this sample. Even more importantly, the PCK and inquiry practices test results correlated weakly or even negatively (though nonsignificantly) with the two focal scales measuring inquiry use. This indicates that a greater knowledge of inquiry pedagogy is not significantly related to its self-reported use in the classroom. One likely interpretation of these findings is that the teachers who lack understanding of inquiry strategies are more likely to overestimate their use of these strategies in the classroom. Another explanation is that social desirability strongly impacts self-reported use of inquiry strategies, especially when teachers have limited skills in using or identifying those strategies. Decades of emphasis on inquiry in science education programs may have instilled a basic preference for inquiry without the ability to enact such methods or critical content knowledge related to its use. Either explanation indicates that the use of inquiry in the classrooms is likely weak, although the teachers may not realize it and believe they are creating those experiences.

¹ PCK and POSITT-F tests were spiraled in separate forms and not administered to any of the same participants.

Contrast of Student Ratings on the IS Scale

One concern with self-report is the possibility of impression management or social desirability on the part of the teachers, who likely recognize the preference of student-centered instructional strategies by their peers or university-based colleagues. A point of comparison is student ratings of the frequency of the same strategies based on the IS scale. We believed students should be able to accurately recognize these strategies in their classroom and would not be biased toward over-reporting because they would likely not be aware of any social desirability of any of these practices. Comparing the classroom average of student responses to the teachers' responses, we found that there was a modest negative correlation ($r = -0.26$, $p = .08$), indicating disagreement between teachers and students on the frequency with which strategies are used. Consistent with this possible explanation, we found that the class on average tended to estimate the inquiry strategies as *less* frequently used (average differences of 0.5 rating points on a 5-point scale, $d = -1.04$) and non-inquiry strategies as *more* frequently used ($d = 1.28$) compared to teachers' reports. It should be noted that self-report by students is problematic in its own right, perhaps being influenced by availability bias or inaccuracy. It is also not known how much of their science teacher's range of strategies they would have encountered by early October when students were assessed. It is possible that teachers prefer more directive strategies early in the academic year.

Discussion

Assessing the impact of professional development on teacher practice is essential for identifying best practices in promoting teacher quality through in-service training. Our broader project has focused on professional development that helps teachers use inquiry in their science classrooms. The data reported in this study were gathered before the intervention and thus reflect teachers' initial attitudes and practices.

The Next Generation Science Standards (NGSS; NRC, 2013) continue to emphasize the importance of inquiry or scientific practices, a framework that was also supported by earlier standards (Olson & Loucks-Horsley, 2000), but has still not permeated science teaching (Pruitt & Wallace, 2012; Seung et al., 2014; Supovitz & Turner, 2000). Scientific practices in the classroom have been organized around the 5Es of inquiry by some researchers to promote high-quality lessons that support conceptual understandings (Goldston et al., 2013). The purpose of this study was to explore two measures of teachers' inquiry practices along with other scales that could shed light on the validity of the instruments.

In this study, we found that both the 5Es and Inquiry Strategies scales appear to be moderately useful for measuring change in teachers' instructional practice. The IS scale in particular showed sufficient ceiling effects that would allow for growth in the use of these methods as a result of training or interventions. The 5Es survey showed less room for growth than the Inquiry Strategies scales and could be revised to increase the ceiling. Both scales showed interesting patterns of teacher' use in an

exploratory factor analysis (e.g., 5Es showed factors reflecting open-ended exploration and facilitating inquiry), but can each be summarized as a unidimensional scale. Future research might explore whether the multidimensional factor structures reflect patterns of strategy use or approaches to inquiry in larger samples of teachers.

We suggest that the 5Es scale is preferable to the IS scale for science educators involved in similar research, if adjustments are made for raising the ceiling scores. A key difference in the two scales is that the IS scale is not specifically tailored to inquiry-focused instruction and reflects a mix of strategies that might be used by teachers not specifically trying to incorporate inquiry in their classroom. In other words, they may simply recognize the importance of student-centered instruction, but are not specifically incorporating inquiry strategies in the science classroom. On the other hand, the 5Es measure is tailored to a theory of inquiry instruction and therefore is likely to be more specific to inquiry-focused classrooms. This interpretation was supported by the relatively strong correlation between this scale and teachers' attitudes toward inquiry as measured by the POSITT-F. In addition, the student survey indicated a negative correlation between the IS scale and their perceptions of the frequencies of inquiry. Future research should explore whether students are sensitive to the use of 5E-related strategies and can accurately report their use in the classroom. This would allow for a student–teacher report comparison for this scale as well.

Contrasts of IS strategies across teachers with high and low 5Es scores highlighted at least two strategies on the IS scale that are especially sensitive to inquiry use in the classroom: exploring student questions and writing up results of a scientific investigation. Use of these strategies may be the clearest signal that a teacher is using inquiry in their classroom versus simply preferring student-centered strategies. More research on teacher attitudes and practices for these strategies is warranted.

Further research is also warranted regarding the results we found comparing inquiry knowledge and PCK with preferences for and frequency of the use of inquiry practices. The knowledge scales showed no correlation with inquiry frequency use and a significant negative correlation with inquiry use preference. One suggestion is to use qualitative research in the form of interviews or focus group interviews, along with classroom observations, to study self-reported high- and low-frequency users and high- and low-preference users to probe the reasons behind these self-reports. Classroom observation data could also triangulate these forms of self-report data.

Conclusions

Self-report of instructional strategy use is known to be problematic at times and may not always align with observation (Supovitz & Turner, 2000), but the efficiency of self-report means that it will continue to play a role in teacher evaluation research. In particular, as long as the self-report scales allow for growth in ratings (i.e., there is sufficient ceiling when teachers are assessed prior to an intervention), a positive

bias in self-report scales would not preclude their use for assessing increases over time. Both Inquiry Strategies and 5Es scales studied here allowed some room for growth, but could be improved by modifying the rating scale to increase the room for improvement. Some possibilities to explore include *positively packing* the scale (Klockars & Yamagishi, 1988) or considering retrospective pre-post assessments (Hill & Betz, 2005).

We conclude that the 5Es scale is more valid for measuring frequency of use than the IS scale due to its specificity and moderate correlation with the measure of inquiry preferences. However, there is a need for more research on unpacking the complexities this study raised, including: (a) the discrepancy between knowledge of and preferences for inquiry; (b) the science teachers' understandings of just what inquiry-based teaching is; and (c) how to modify professional development programs in light of the research information.

The current study indicated that some caution is warranted when relying on self-report of inquiry in professional development programs. The teachers self-reported higher levels of the use of inquiry-based strategies than was reflected either in their students' ratings or in conjunction with their knowledge of inquiry practices, preferences, or inquiry-based PCK. There are at least two possible reasons for this discrepancy. First, teachers may have been influenced by social desirability or impression management biases. Second, teachers' relatively weak understandings of inquiry practices may be contributing to an overestimation of their own use of inquiry. If a teacher does not have a strong understanding of what constitutes inquiry, he/she is not likely to recognize it in practice or know when it is being used ineffectively. Both of these hypotheses are important possibilities and supported by other studies.

Lee et al. (2004) also found that teachers self-reported a high use of inquiry strategies, even at the beginning of their professional development program. However, classroom observations by researchers in their study indicated a more modest use of inquiry-based instruction in the classroom. A study by Seung et al. (2014) similarly indicated that teachers and pre-service tended to overestimate inquiry use, characterizing some traditional strategies as inquiry based. The findings from these two studies, along with the current study (see also Lakin & Shannon, 2013), appear to confirm a trend for teachers in professional development programs to overrate their own use of inquiry strategies, especially prior to a professional development intervention.

Further research exploring this phenomenon implies that program designers could include instruction in the characteristics of inquiry practices earlier within the professional development program. For example, participants could analyze videos of science teachers in action and discuss examples and non-examples of inquiry. Self-report surveys could then be administered a second time, a few days into the intervention to obtain more accurate baseline data of where teachers are starting in the inquiry development process. Further, a discussion of how teachers changed their views of their own perceptions of inquiry could be instructive for the group as a whole.

The results of this study also indicate promise for achieving a better understanding of scientific inquiry through the use and dissemination of the term *scientific practices*, rather than *inquiry*. The construct of inquiry has been fraught by

misunderstanding and miscommunication for some four decades. The use of scientific practices as defined in the NGSS presents a more crystallized vision of what one does while engaged in science, such as asking questions or analyzing data that may be more easily recognized in the classroom. There are many specific examples of scientific practices in the NGSS, making it possible to teachers to develop a tangible sense of these practices. In contrast, the term inquiry has over the years referred not only to scientific practices, but also to entities such as understandings of the nature of science, a philosophy, a guiding principle of instructional design, a type of curriculum, and a form of pedagogy (Settlage, 2013). A further challenge for future researchers is to develop both observation and self-report scales more tailored for collecting data on teachers' use of scientific practices in their classroom teaching. As professional development continues to play an important role in improving science classroom practices (Elmore, 2002), evaluation research on both the training and the use of strategies will be critical.

Acknowledgments This material is based upon work supported by the National Science Foundation under Grant No. MSP-1102997.

References

- Alonzo, A. C., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher–student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, *49*, 1211–1239.
- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press
- Avraamidou, L. (2013). Prospective elementary teachers' science teaching orientations and experiences that impacted their development. *International Journal of Science Education*, *35*, 1698–1724.
- Bodzin, A. M., & Beerer, K. M. (2003). Promoting inquiry-based science instruction: The validation of the science teacher inquiry rubric (STIR). *Journal of Elementary Science Education*, *15*, 39–49.
- Bracken, B. A. (2007). Creating the optimal preschool testing situation. In B. A. Bracken & R. J. Nagle (Eds.), *Psychoeducational assessment of preschool children* (4th ed., pp. 137–154). Mahwah, NJ: Lawrence Erlbaum Associates.
- Butler, D. L., Novak Lauscher, H., Jarvis-Selinger, S., & Beckingham, B. (2004). Collaboration and self-regulation in teachers' professional development. *Teaching and Teacher Education*, *20*, 435–455.
- Bybee, R. W., Buchwald, C. E., Crissman, S., Heil, D. R., Kuerbis, P., Matsumoto, C., & McInerney, J. D. (1989). *Science and technology education for the elementary years: Frameworks for curriculum and instruction*. Washington, DC: National Center for Improving Science Education, Washington and Colorado Springs CO.
- Capps, D. K., Crawford, B. A., & Constanas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. *Journal of Science Teacher Education*, *23*, 291–318.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7). Retrieved from <http://pareonline.net/getvn.asp?v=10&n=7>
- Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Washington, DC: Albert Shanker Institute. Retrieved from http://www.shankerinstitute.org/Downloads/Bridging_Gap.pdf
- Goldston, M. J., Dantzler, J., Day, J., & Webb, B. (2013). A psychometric approach to the development of a 5E lesson plan scoring instrument for inquiry-based teaching. *Journal of Science Teacher Education*, *24*, 527–551.
- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, *26*, 501–517.

- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*, 800–802.
- Huffman, D., Thomas, K., & Lawrenz, F. (2003). Relationships between professional development, teachers' instructional practices, and the achievement of students in science and mathematics. *School Science and Mathematics*, *103*, 378–387.
- Jeanpierre, B., Oberhauser, K., & Freeman, C. (2005). Characteristics of professional development that effect change in secondary science teachers' classroom practices. *Journal of Research in Science Teaching*, *42*, 668–690.
- Jeffers, A. T., Safferman, A. G., & Safferman, S. I. (2004). Understanding K-12 engineering outreach programs. *Journal of Professional Issues in Engineering Education and Practice*, *130*, 95–108.
- Johnson, C. C. (2007). Whole-school collaborative sustained professional development and science teacher change: Signs of progress. *Journal of Science Teacher Education*, *18*, 629–661.
- Johnson, C. C. (2009). An examination of effective practice: Moving toward elimination of achievement gaps in science. *Journal of Science Teacher Education*, *20*, 287–306.
- Johnson, C. C. (2010). Making the case for school-based systemic reform in science education. *Journal of Science Teacher Education*, *21*, 279–282.
- Karabenik, S. A., & Maehr, M. L. (2007). *Final report to the National Science Foundation: MSP—Motivation Assessment Program*. Retrieved from <http://hub.mspnet.org/index.cfm/15222Karplus>
- Karplus, R. (1979). Teaching for the development of reasoning. In A. Lawson (Ed.), *1980 AETS yearbook: The psychology of teaching for thinking and creativity*. Columbus, OH: ERIC/SMEAC.
- Kimble, L. L., Yager, R. E., & Yager, S. O. (2006). Success of a professional-development model in assisting teachers to change their teaching to match the more emphasis conditions urged in the National Science Education Standards. *Journal of Science Teacher Education*, *17*, 309–322.
- Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. *Journal of Educational Measurement*, *25*, 85–96.
- Lakin, J. M., & Shannon, D. M. (2013). The role of treatment acceptability and understanding in formative evaluation: Predicting implementation variation in a middle school science program. Paper presented at the American Evaluation Association, Washington, DC.
- Lee, O., Deaktor, R. A., Hart, J. E., Cuevas, P., & Enders, C. (2005). An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, *42*, 857–887.
- Lee, O., Hart, J. E., Cuevas, P., & Enders, C. (2004). Professional development in inquiry-based science for elementary teachers of diverse student groups. *Journal of Research in Science Teaching*, *41*, 1021–1043.
- Lee, O., Penfield, R., & Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching*, *46*, 836–859.
- Lotter, C., Harwood, W. S., & Bonner, J. J. (2007). The influence of core teaching conceptions on teachers' use of inquiry teaching practices. *Journal of Research in Science Teaching*, *44*, 1318–1347.
- Luft, J. A., Firestone, J. B., Wong, S. S., Ortega, I., Adams, K., & Bang, E. J. (2011). Beginning secondary science teacher instruction: A two-year, mixed-methods study. *Journal of Research in Science Teaching*, *48*, 1199–1224.
- Lumpe, A. T., Haney, J. J., & Czereniak, C. M. (2000). Assessing teachers' beliefs about their science teaching context. *Journal of Research in Science Teaching*, *37*, 275–292.
- Minner, D. D., Jurist Levy, A., & Century, J. (2010). Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, *47*, 474–496.
- Mullens, J. E., Gayler, K., Goldstein, D., Hildreth, J., Rubenstein, M., Spiggle, T., ... Welsh, M. (1999). *Measuring classroom instructional processes: Using survey and case study field test results to improve item construction* (working paper 1999–2008). Washington, DC: Policy Studies Associates. Retrieved from <http://files.eric.ed.gov/fulltext/ED434144.pdf>
- National Research Council. (1998). *Inquiry and the National Science Education Standards*. Washington, DC: The National Academies Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- National Research Council. (2013). *The next generation science standards*. Washington, DC: The National Academies Press.

- Olson, S., & Loucks-Horsley, S. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning*. Washington, DC: National Academy Press. Retrieved from <http://www.nap.edu/catalog/9596.html>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.
- Pruitt, S., & Wallace, C. S. (2012). The effect of a state department teacher mentor initiative on science achievement. *Journal of Science Teacher Education, 23*, 367–385.
- Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10*, 805–817.
- Schuster, D., Cobern, W. W., Applegate, B., Schwartz, R., Vellom, P., & Undreiu, A. (2007). *Assessing pedagogical content knowledge of inquiry science teaching—Developing an assessment instrument to support the undergraduate preparation of elementary teachers to teach science as inquiry*. Proceedings of the National STEM Conference on Assessment of Student Achievement, hosted by the National Science Foundation and Drury University, Washington, DC, October 19–21, 2007.
- Schwab, J. J., & Brandwein, P. F. (1966). The teaching of science as enquiry. In J. J. Schwab & P. F. Brandwein (Eds.), *The teaching of science* (pp. 3–103). Cambridge, MA: Harvard University Press.
- Settlage, J. (2013). On acknowledging PCK's shortcomings. *Journal of Science Teacher Education, 24*, 1–12.
- Seung, E., Park, S., & Jung, J. J. (2014). Exploring preservice elementary teachers' understanding of the essential features of inquiry-based science teaching using evidence-based reflection. *Research in Science Education, 44*, 507–529.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1–22.
- Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy, 14*, 331–356.
- Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*, 963–980.
- Trumbell, D. J., Scarano, G., & Bonney, R. (2006). Relations among two teachers' practices and beliefs, conceptualizations of the nature of science, and their implementation of student independent inquiry projects. *International Journal of Science Education, 28*, 1717–1750.
- Welch, W. W., Klopfer, L. E., Aikenhead, G. S., & Robinson, J. T. (1981). The role of inquiry in science education: Analysis and recommendations. *Science Education, 65*, 33–50.