National Center for Research on Rural Education (R²Ed)

The Effectiveness of Technology-Delivered Science Instructional Coaching in Middle and High School

Gwen Nugent, Gina Kunz, James Houston, Irina Kalutskaya, ChaoRong Wu, Jon Pedersen
University of Nebraska–Lincoln

SoonChun Lee
Wichita State University

Sue Ellen DeChenne
University of Northern Colorado

LinLin Luo
University of Nebraska–Lincoln

Brandi Berry
Wisconsin Department of Corrections

Development of this working paper was completed at the National Center for Research on Rural Education (R^2Ed), funded by a grant from the U.S. Department of Education's Institute of Educational Sciences (R305C090022).

University of Nebraska Lincoln

The Effectiveness of Technology-Delivered Science Instructional Coaching in Middle and High School

Although results showing coaching effectiveness are accumulating, coaching is often included with other forms of PD support including teacher in-service (Powell, Diamond, Burchinal, & Koehler, 2010; Kretlow et al., 2011), access to an annotated video library (Allen et al., 2011), and access to ongoing learning communities (Gallucci, Van Lare, Yoon, & Boatright, 2010; Matsumura, Garnier, & Spybrook, 2012). The presence of multiple intervention components obscures the unique effect of coaching and makes drawing conclusions about coaching effectiveness impossible.

A needed area for science PD is helping teachers acquire expertise in instructional approaches for students to develop appropriate science practices. Science practices and their integration into core disciplinary concepts are of central importance in *Next Generation Science Standards* (NGSS; National Research Council [NRC], 2013), and student acquisition of science practice skills is expected in national and state educational mandates (Common Core State Standards Initiative, 2010; National Research Council, 2011). Research shows that instruction that infuses science practice skills into content can improve science achievement and process skills (Bransford, Brown, & Cocking, 1999; Donovan & Bransford, 2005; Llewellyn, 2002; Minner, Levy, & Century, 2010; Schroeder, Scott, Tolson, Huang, & Lee, 2007), but many science teachers report not knowing how to successfully teach these skills (Anderson & Michener, 1994; Bybee & Fuchs, 2006; Capps & Crawford, 2013). U.S. students also do much better identifying the correct answers to simple scientific tasks than using evidence from their experiments to explain those answers (Parnass, 2012). Thus, there is a critical need for identifying effective, sustainable approaches for teacher PD in delivering instruction to foster science practice skills. The purpose of this research study was to a) determine the effects of a professional development intervention comprised of a summer institute and follow-up technology-delivered instructional coaching on teacher and student science practice knowledge, skills, self-efficacy, and engagement and b) isolate specific effects of coaching when combined with more traditional teacher workshops.

Theoretical and Empirical Background

Professional development (PD). Recent professional development research literature has identified several characteristics that influence positive teacher outcomes including an emphasis on deepening teachers' knowledge of content and pedagogy, active teacher engagement in learning opportunities, and experiences encouraging collaboration among teachers (Darling-Hammond et al., 2009; Desimone, 2009). The PD should also be of sufficient duration; a comprehensive study examining 1,300 studies addressing the effect of PD on student achievement concluded that more than 14 PD hours showed a significant effect (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). The PD should also promote continuity to other in- and out-of-school experiences (Garet, Porter, Andrew, & Desimone, 2001; Loucks-Horsely et al., 2003). Graduated experiences including didactic instruction, modeling, practice, feedback, and opportunities to adapt new skills into natural classroom contexts (e.g., via coaching) are also necessary to achieve desired experiential and learning outcomes (Ingersoll & Kralik, 2004; Luft, 2001; Pianta, 2005). Such characteristics reinforce teachers' development of evidence-based instructional strategies and application of these skills in relevant instructional contexts (Akerson & Hanuscin, 2007; Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005).

Instructional coaching. The enactment of No Child Left Behind (NCLB) legislation provided an impetus for the introduction of coaching into the schools through creation of the Reading First Initiative where coaching was suggested "as a viable way to provide sustained and effective PD support to teachers" (Denton & Hasbrouck, 2009, p. 153). Further NCLB provisions created thousands of reading coaching positions by mandating that each Reading First school be served by a reading coach. This influx of coaches into the schools was the start of a new generation of teacher PD, and coaching was rapidly extended into mathematics. Unfortunately, this introduction of coaching was not accompanied by empirical research on coaching effectiveness. The research base on coaching is limited and often focuses on descriptive and case study approaches and reported best practices (Borman & Feger, 2006; Cornett & Knight, 2009). For example, a review of mathematics specialists and coaching research for the National Council of Teachers of Mathematics showed that only a small portion of the studies focused on improving instructional practices and student achievement (McGatha, 2009). But despite its limitations, research with literacy and mathematics coaching suggests promise (Campbell & Malkus, 2011; Foster & Noyce, 2004; Sailors & Shanklin, 2010). Coaching also has led to impacts beyond teacher improvement to student achievement (Lockwood, McCombs, & Marsh, 2010; Powell et al., 2010). While this research contributes to our understanding of coaching impacts, there are limited science coaches in K–12 education, and research is extremely limited, with few evidence-based instructional science coaching programs. The studies that do exist have shown that coaching helps teachers understand inquiry-based teaching practices (Lotter, Yow, & Peters, 2014), produces higher student science achievement compared to a control (Vogt & Rogalla, 2009), and improves student achievement through a focus on teacher–student interactions (Allen et al., 2011).

Recently, technology has been used to deliver coaching, eliminating the need for the coach to be present in the teacher's school for observation and conducting the coaching session. Research results have confirmed technology-delivered coaching to be an effective and efficient delivery method. It has been shown to be equally effective in comparison to on-site coaching

(Powell et al., 2010) and better than receiving video exemplars of "best practices" (Pianta, Mashburn, Downer, Hamre, & Justice, 2008) and regular in-service training (Allen et al., 2011).

Theory of change. Figure 1 shows the theory of change guiding the research (adapted from Desmoines, 2009) which is based on both theory and empirical research. We hypothesize that teachers' ability to make changes in instructional practice reflects their level of foundational content and pedagogical *knowledge* for effective instruction in science practice skills, the degree

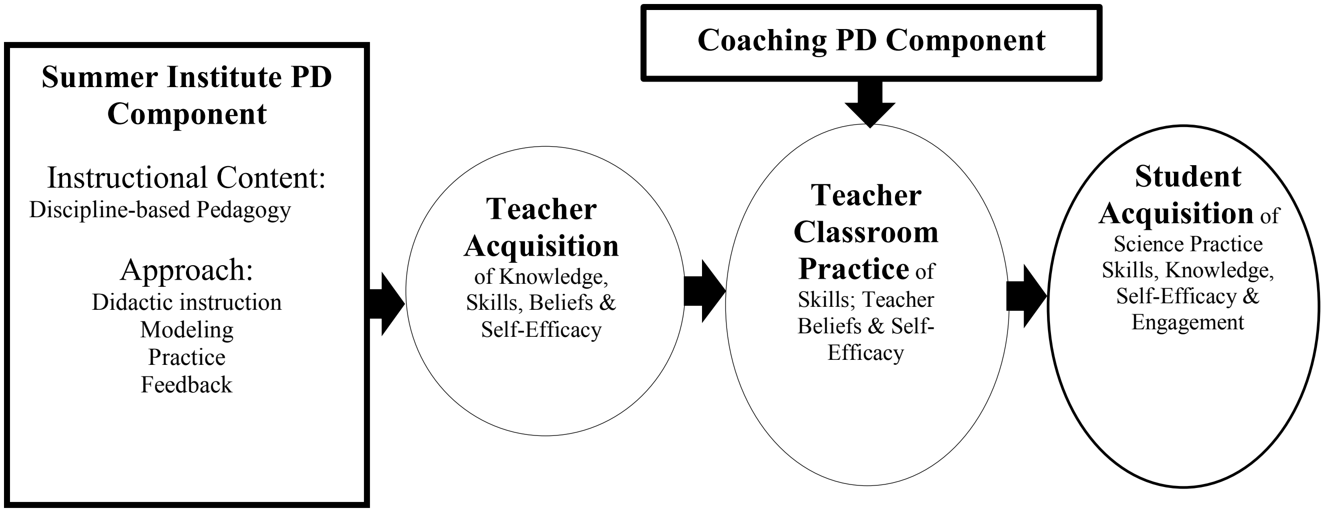

Figure 1. Theory of Change

to which they *believe* such instruction effects meaningful change in student learning, their *self-efficacy* for teaching science practices, and the opportunity for repeated *practice* with feedback. The theoretical basis relies on social cognitive theory which suggests that "people act on their judgments of what they can do [self-efficacy], as well as on their beliefs about the likely effects of various actions" (Bandura, l986, p. 231.) Theory of planned behavior (Ajzen, 1985) identifies beliefs as a predictor of one's intention to engage in a behavior, and holding appropriate beliefs about inquiry-based teaching has been shown to be important for teachers to fully take advantage of professional development and subsequent transfer to classroom practice (Lotter, Rushton, & Singer, 2013). Practice with feedback is well established in the literature. The law of frequencies (Malone, 1990) suggests that accurate acquisition and successful performance of a newly acquired skill requires practice with feedback (Haring, Lovitt, Eaton, & Hansen, 1978; Renaissance Learning, 2015). Indeed, one comprehensive analysis of staff development literature suggested that teachers need to implement a complex teaching practice 25 times with feedback to insure transfer (Showers et al., 1987).

Effective classroom practice, in turn, leads to improved student outcomes. We hypothesized that the summer institute would be responsible for a direct impact on teacher knowledge, beliefs, skills, and self-efficacy for teaching science practice skills. Follow-up instructional coaching was hypothesized to promote teacher transfer of skills to classroom practice and maintenance over time. Teacher implementation of these skills was in turn believed to foster positive student outcomes as observed in improved student science practice knowledge, performance, confidence in "doing" science, and engagement in the instructional process.

Overview of the Intervention

Our PD model was designed to equip middle and high school science teachers with knowledge and skills to use a guided scientific inquiry approach to teach science practice skills integrated into content (Capps, Crawford, & Constas, 2012; Nugent et al., 2012) as specified by NGSS. The new standards have conceptualized science inquiry as a "practice" and emphasize that engaging in scientific investigations requires "not only skill but also knowledge that is specific to that practice" (NRC, 2011, p. 30). The NGSS practices identified as essential for students and that formed the basis of our intervention include the following: (a) asking scientific questions; (b) planning and carrying out investigations; (c) analyzing and interpreting data; (d) developing explanations based on the evidence gathered through data collection; and (e) communicating findings.

The guided scientific inquiry approach used by the project is grounded in student data collection and analysis that leads to student formulation of an underlying science concept or principle. It is also teacher-facilitated, requiring extensive use of teacher questioning and scaffolding to guide students to greater understanding of science concepts, science content, and science practice skills.

The intervention embodied many critical evidence-based PD elements identified in the literature as constituting high quality PD (e.g., modeling and practice with guided feedback). It consisted of a 5-day training for coaches prior to the summer institute and an intensive 8-day summer institute for teachers (over two weeks) followed by 8–12 technology-delivered, asynchronous (delivered outside the classroom instruction time) coaching sessions across 6–8 consecutive weeks during the school year.

Summer Institute

The summer institute, led by project coaches with support from university project faculty, aimed to promote knowledge and skills through use of didactic presentation, modeling skills by science educators and coaches, teacher practice of new skills, and feedback provided to teachers by content experts, including coaches. During the summer institute, project coaches and science educators introduced teachers to the guided scientific inquiry approach by modeling lessons in which participants served as "students." Group discussion followed to clarify concerns or questions. As additional support, teachers were given 36 sample middle and high school 6–8 week unit lessons that integrate science content (life science, physical science, earth science, and chemistry) and practices as instructional models that could be implemented in their classrooms during the following school year.

At the end of week one of the summer institute, teachers identified a sample lesson and prepared to present it the following week. This presentation in week two gave teachers an opportunity to enact strategies learned in the previous week through delivering a practice lesson and receiving feedback from coaches and peers. It also allowed teachers to view each other's lesson implementation and experience additional lessons. Throughout the institute we also interspersed discussions and exercises about posing various types and levels of questions to students and scaffolding student knowledge and skill acquisition. Modeling and video examples of classroom teaching using the guided scientific inquiry instructional method were also used.

Instructional Coaching

The instructional coaching, which occurred during the subsequent school year, aimed to support teacher transfer of knowledge, skills and self-efficacy gained in the summer institute regarding the guided inquiry instructional approach to classroom practice. Table 1 represents the framework for our coaching approach (Hanft, Rush, & Shelden, 2004). Primary features

Table 1. Coaching Framework

Features	Participants	Definition
Joint planning	Teacher & coach together	Discuss/agree on actions before/during implementation
		Occurs as part of all coaching conversations
Action/practice	Teacher & coach	Spontaneous/planned opportunities for teacher to practice, refine, analyze new/existing skills, determined by joint planning
Observation	Teacher & coach together	Examination of another's practices (coach or teacher) to develop new skills, strategies, or ideas
		Can involve use of video or live modeling by coach
Reflection	Teacher & coach	Analyze actions, practices, strategies, ideas, in light of new or intended outcomes
		Video/digital audio can be used as tool to guide reflection
Feedback	Teacher & coach together	Information provided by coach after teacher implements new skill or reflected on own observations/actions
		Can take form of affirmations, examples, descriptions, data, suggested areas for improvement and resources
		Purpose is to promote teacher's new skills, strategies, and ideas as they relate to intended outcomes

involved joint teacher–coach planning followed by opportunities for teachers to practice, refine, and analyze new and existing skills; opportunities for coaches to observe teacher instruction; coach and teacher reflection; and joint feedback. Coaching sessions followed a coaching protocol that included (a) positive coach feedback; (b) review of desired student outcomes and teaching strategies that promote student inquiry skills; (c) detailed discussion of the lesson including sharing time-stamped video clips to demonstrate what worked well and why, and what student outcomes need to be addressed or improved; and (d) exchange of ideas about strategies to address areas for improvement.

Coaches utilized questioning techniques and scaffolding to guide teachers to greater understanding and proficiency in implementing guided scientific inquiry instruction in the classroom. Coaches helped teachers identify instructional strategies to support student skill acquisition. Teachers and coaches jointly developed a data collection procedure for the coach to collect evidence of strategies and student outcomes during observation of the teacher's instruction. During the next coaching session, they shared and discussed the data collected.

Figure 2 depicts the technology-delivered coaching process which involved bi-directional discussion and feedback based on video-recorded classroom lesson implementation. The teacher

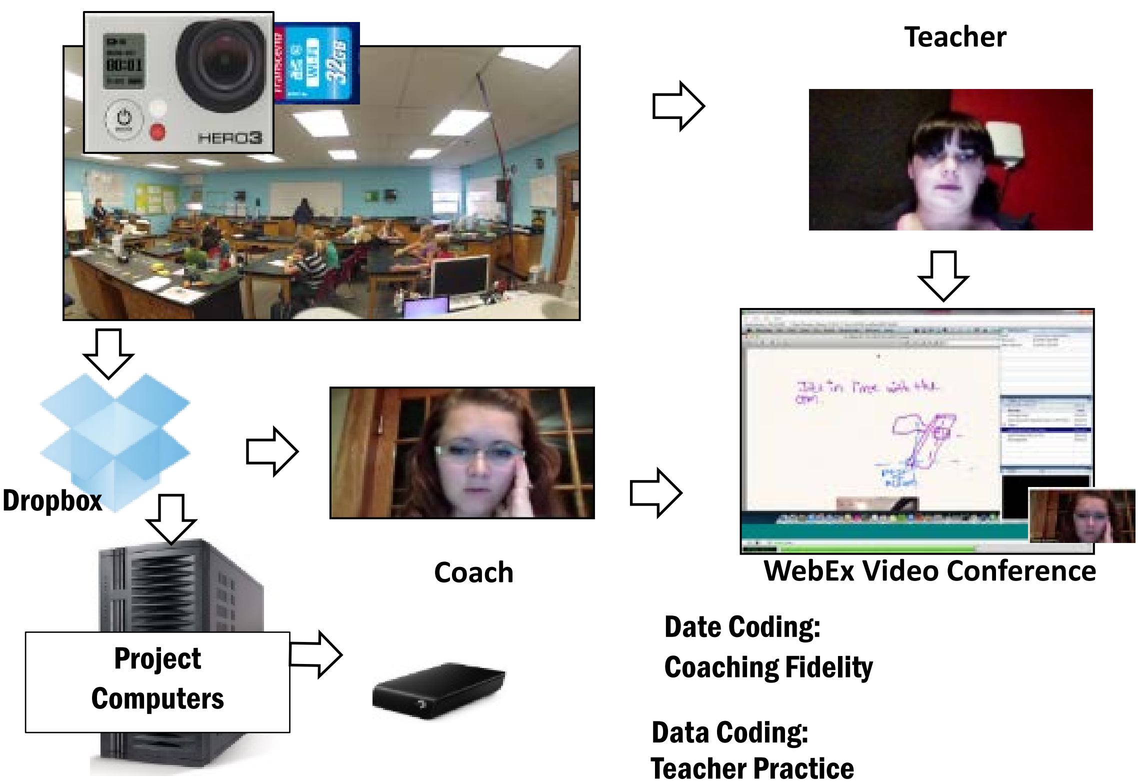

Figure 2. Coaching Process

video-recorded his/her lesson implementation during classroom instruction and uploaded the digital video recording to Dropbox. Teacher and coach prepared for the coaching session by independently reviewing the video prior to the scheduled coaching session. Project coders also reviewed the videos to code observation instruments measuring classroom practice.

Coaching sessions used WebEx, a video-conferencing program that allowed sharing of files and classroom videos, as well as session recording. The coaching sessions were delivered for about an hour and were scheduled around the teacher's schedule. They involved discussion and feedback of the prior video-recorded lesson and planning for the next classroom instruction and ended with scheduling the next coaching session. The teacher followed up with any preparation needed for the next class implementation, and the coach completed a self-reflection protocol designed to help him/her improve his/her coaching content and process skills. Coach skills building also was enhanced by weekly team meetings including the coaches and research team. The teacher then implemented the lesson plan that was developed during the coaching session. This basic process was conducted for approximately 1–2 sessions per week across 6–8 weeks until formal coaching sessions were mutually terminated by teacher and coach. Generally, the sessions were discontinued when a) teachers believed they were able to successfully implement a guided inquiry approach in the classroom and b) the coach had documented that teachers had demonstrated the skills represented on the teacher observation instruments.

One to two coaching sessions in a 6–8 week period was selected for several reasons. First, the project-provided lessons represented approximately a 6–8 week unit. Second, having 1–2 sessions per week for 6–8 weeks allowed for a high dosage of coaching sessions over a sufficient period of time (approximately one quarter of the school year). Third, use of 6–8 week coaching periods allowed coaches to stagger their coaching workload across the academic year.

Each coach's caseload was approximately 13 teachers per year; the per-coach load at any one time was approximately five teachers.

Method

The project involved a randomized controlled trial aimed at addressing the following research question: *What is the impact of a summer institute focused on guided scientific inquiry with follow-up coaching (treatment) versus no professional development (control) on (a) teacher science practices knowledge, skills, self-efficacy, and beliefs and (b) student science practices knowledge, skills, engagement and self-efficacy?* A secondary question involved the independent effects of the summer institute and coaching: *What were the separate effects of the summer institute and coaching on teacher and student outcomes?*

Participants

The study was conducted with 124 science teachers (63 treatment and 61 control) from 110 rural schools (61 treatment and 49 control) in Nebraska and Iowa. The average teacher's age was 41 years (SD = 10.84), and the gender split was 70% female and 30% male. Average teaching experience was 13 years (SD = 9.33), and 48% of the teachers had a master's degree. Twenty-eight percent of teachers taught in middle schools, 42% in high schools, and 30% in schools that served elementary, middle and high school students.

Students of these teachers comprised the student study sample. The numbers of students completing each of the instruments varied, but there were approximately 1,000 participating students, split nearly equally between middle (48%) and high school (52%). Forty-nine percent were male and 51% were female. In terms of ethnicity, 83% were White (non-Latino), followed by Hispanic/Latino (8%), and multi-racial (3%). The remainder was divided among Asian/Pacific Islander, Native American and Black/African-American.

Research Design

The research study used a two-cohort, group-randomized experimental group design examining differences between a treatment and control group. The treatment group received the summer institute plus coaching; the control group teachers (who taught in the same manner as usual) did not participate in either the summer institute or coaching. The study independent variable (treatment versus control) was manipulated at the school level; thus, schools were the unit of randomization. This method also prevented any contamination of control and treatment teachers within the same school. Summer institutes were conducted in two consecutive years for the two separate teacher cohorts. Teachers were assigned to coaches based on grade level and subject area. Focus groups were also conducted with 16 teachers who participated in the first year of the Coaching Science Inquiry (CSI) study (Authors, 2015), and quotes from these focus groups were used to provide insight into the quantitative results.

Data Analysis

Mixed linear modeling with restricted maximum likelihood (REML) estimation and the Kenward-Roger adjustment for standard errors and denominator degrees of freedom (Kenward & Roger, 1997, as recommended by Stroup, 2012) was used to examine the effectiveness of summer institute with follow-up coaching. Separate models were conducted for each of the teacher and student outcomes. The teacher model structure included time nested within teacher and teacher nested within school. The four-level student model accounted for time nested within student, student nested within teacher, and teacher nested within school. Fixed effects were the effects of group (treatment vs. control), time point (teachers: pre-summer institute/baseline, post-summer institute, post-implementation of unit and end of the year; students: beginning of school year, post-unit and end of year) and the group X time interaction. Teacher cohort was included as covariate for all outcomes, and level (middle/high) was a student covariate for the self-efficacy measure. Random teacher and school effect were included for teacher outcomes.

For student outcomes, random student effect was also included if there were multiple time points. Analysis was conducted using SAS Proc Glimmix.

In order to address research question 2, which dealt with the separate effects of the summer institute and coaching, we conducted a priori planned comparisons between treatment and control groups for specific time segments: baseline to post-summer institute (effect of summer institute), post-summer institute to post-implementation of unit (effect of coaching), and post-implementation of unit to end of the year (sustainability of effects). To control for false positive results in the planned comparisons the Benjamin Hockberg method (SAS PROC MULTTEST) was used to produce FDR-adjusted (false discovery rate) p values. These analyses, along with results from baseline to the final data collection time point for each outcome, provide insight into the total effect of the intervention, as well as the individual components of summer institute and coaching.

Data Collection and Instrumentation

Selection and development of instruments was driven by the need to the measure science practices specified in the NGGS: questioning, investigating, analyzing and interpreting data, explaining, and communicating. Teachers completed measures at four time points: baseline, post-summer institute, after they implemented their science practice unit (6–8 week period during the school year), and at the end of the school year. Not all instruments were completed at each time point. Teachers in the control condition completed the same measures at the same time intervals with one exception. Collecting control group data at the post-summer institute time point, which occurred only two weeks after the previous data collection period, would have placed an undue burden on control group teachers. Because this 2-week time period occurred early summer, teachers would not have gained additional classroom teaching experience or would have had significant opportunities for professional development. Thus, control teachers' baseline data was also used for the post-summer institute time point. Students completed instruments three times: beginning of school year (baseline), after teachers had delivered their science practice unit, and at the end of the year.

Teaching of student science practice skills was assessed by observation of teacher classroom instructional practice by trained, independent coders who reviewed videotapes of the teaching at baseline, post-summer institute (treatment only), and after the teacher had completed their science inquiry unit (post-unit). Baseline videos were obtained in spring semester prior to the summer institute. Both treatment and control teachers video-recorded one to four classroom periods that they believed represented instruction regarding science practices. The post-institute time point represented videos of treatment teachers practice lesson delivered during the summer institute. For the post-unit time point, teachers in the control condition identified a 6–8 week period for the intervention which they taught a science unit in one of the areas of life, physical or earth science and recorded four classroom periods for coding that, taken together, represented the inquiry cycle. Teachers in the treatment condition recorded approximately two class periods per week while they implemented their unit and received the coaching. All videos were reviewed by the coaches as the basis for the coaching sessions; however, to align with the control condition, only four videos were reviewed and scored by the independent project coders. The four videos consisted of the first recorded video, the final recorded video and two others identified by the teachers as representative of components of an inquiry cycle.

Independent project coders conducted observations from the recordings to determine measures of teacher science practice instruction and student engagement. Coders did not know whether teachers were in the treatment or control group. Coders went through an extensive training period and could not perform project coding until they had coded four practice videos showing interrater agreement of 80–85% with videos previously coded by project staff. All of an individual teacher's videos were assigned to a single coder who coded videos in chronological order. Coders were thus able to follow a teacher throughout the instruction and provide an overall, cumulative assessment that took into account all of that teacher's videos. Twenty-five percent of videos were coded by a primary coder (a project graduate assistant) and a secondary data coder in order to establish interrater reliability. The interrater reliability procedures required that independent coders meet to discuss any discrepancies in coding.

Teacher-Completed Instruments

Beliefs. This instrument, adapted from Teaching Beliefs in Inquiry-based Teaching (Duran, Ballone-Duran, Haney, & Beltyukova, 2009), was administered at baseline and the end of the year. The decision for the extended time period between data collection points was due to the fact that teaching beliefs are often deeply held and difficult to alter; they must be tested and found effective in order to change (Jones & Carter, 2007; Pajares, 1992). The instrument consisted of 26 Likert-type items measuring teacher beliefs about inquiry-based teaching, including issues of student engagement and learning, as well as barriers to implementation. Cronbach's alpha coefficient reported by the original authors was .76. Cronbach alpha for this research was .82.

Knowledge. This project-developed instrument (Authors, 2011) contains 25 multiple-choice items developed to measure teacher knowledge of (a) nature of science, (b) student science practice skills as reflected in NGSS, and (c) guided inquiry pedagogical content knowledge. The assessment was administered at baseline, post-summer institute, and after teachers had delivered their science practice unit. The pedagogical content knowledge questions used brief teaching scenarios as a context for teachers to identify appropriate guided inquiry-based approaches (Schuster, Cobern, & Applegate, 2011). Validation was established through direct alignment with the standards and through expert review. Cronbach's alpha coefficient was

.60. The development of this instrument included a thorough field test using item response theory, which involved an item analysis and item deletion process resulting in increased breadth of applicability for diverse teacher abilities.

Self-efficacy. This project-developed instrument was designed to assess teacher self-efficacy in promoting the student science practice skills. Items were rated on a 0–100% confidence scale. Validation was established through an iterative review and revision of items by project coaches and science educators. The survey was administered at all four time points. Cronbach's alpha coefficient was .97.

Teacher Classroom Observation Measures

Three separate observation instruments were used in order to provide a comprehensive assessment of teacher performance:

The Teacher Inquiry Rubric (TIR; Authors, 2013) is a four-level rubric assessing teacher proficiency in guiding students to develop necessary practices in science questioning, investigating, collecting data, explaining, communicating, and applying science knowledge to a new situation. The instrument was constructed so that within each of the six constructs, there are four levels: "pre" (teachers show no evidence of promoting student acquisition of the practices); "developing" (teachers directly present science practice topics through lecture or demonstration); "proficient" (teachers use guiding questions, experiences, and feedback to help students differentiate between examples and non-examples of the practices); and "exemplary" (the teacher uses guiding questions, scaffolding, and feedback to directly elicit student skill performance). The instrument was developed to provide specific behavioral indicators (a total of 31 indicators) which could be used by observers to provide a quality rating of teacher guided inquiry skills. A total construct score was assigned for each of the six constructs, as well as an overall lesson score. This cumulative, overall score was used as the measure of teacher practice. The Kappa statistic was .93.

The Partial Interval Classroom Inquiry Observation System–Teacher Version (PICI-T; Authors, 2013) uses a 15-second partial interval recording procedure (e.g., Cooper, 1987; Fisher, Piazza, & Roane, 2011; Rapp, Colby-Dirksen, Michalski, Carroll, & Lindenberg, 2008; Shapiro & Kratochwill, 2000) to identify specific teacher behaviors that provide opportunities for students to engage in science practices or that do not promote opportunities for student engagement. Definitions for "science practices instruction" (e.g., "instruction delivered within the context of a student-conducted science practice activity in which it precedes the development of the concept") and "non-science practices instruction" were developed to align with guided inquiry instruction and reviewed by a science educator as a face validity form of construct validity. Interrater reliability showed substantial Kappa agreement ($\kappa = .89$).

Electronic Quality of Inquiry Protocol (EQUIP; Marshall, Smart, & Horton, 2009) is a four-level rubric with 19 indicators aligned with four overall constructs: instruction, curriculum, assessment, and discourse. As with the TIR described above, the framework allows a micro (e.g., individual indicators) to macro (e.g., larger constructs such as assessment) appraisal of teacher effectiveness. The instrument provides a method for analyzing the quantity and quality of instruction implemented, which is beneficial in evaluating professional development projects. Cronbach's alpha values as reported by the original author ranged from .88–.89, demonstrating strong internal consistency. Kappa interrater reliability statistics averaged .6 or higher for each observation, showing moderate to substantial agreement. EQUIP, a published instrument which has been widely used across the U.S., provided a level of convergent validity for the two project-developed instruments (TIR and PICI-T). The correlation between TIR and EQUIP was .69; between PICI-T and EQUIP was .63. These correlations provided evidence that there were similar constructs underlying all three measures, but that each embodied unique information regarding science practice instruction.

Student-Completed Measures

Science knowledge. There were separate multiple choice knowledge measures for middle and high school which drew upon publically available inquiry questions from *State Collaboratives on Assessment and Student Standards* (*SCASS),* National Assessment of Educational Progress (NAEP), and Trends in International Math and Science Study (TIMSS; NCES, 2007). Knowledge assessments were completed at the beginning and end of the school year. The Cronbach alpha statistics were .80 for the high school and .65 for middle school.

Science practices self-efficacy. This project-developed instrument contained a total of 10 items measuring student science practice self-efficacy. The same instrument was used for middle and high school. Students completed the survey three times: baseline, at the end of the teacher's science practice unit, and end of year. Responses were provided to Likert-type items on a 5-point scale, ranging from strongly disagree (rating of 1) to strongly agree (rating of 5). They were adapted from an instrument developed for an NSF-funded middle school science/engineering project (Nugent, Barker, Toland, Grandgenett, Hampton, & Adamchuk, 2009). Alpha statistic was .90.

Science practice skills. Teachers completed the Student Inquiry Rubric (SIR; Anthony & Person-Pandil, 2001), a four-level rubric rating each student's skills in the five areas of science practices as specified in the standards: question, investigate, analyze and collect data, explain, and communicate. This instrument was completed after teacher implementation of their 6–8 week unit. While the basic constructs representing science practices were the same for middle and high school, specific indicators for the constructs differed.

Student classroom observation measure. The Partial Interval Classroom Inquiry Observation System–Student Version (PICI-S; Authors, 2013) was a companion to the PICI-T (see description above), and coded student on- and off-task behavior during classroom instruction, as well as student science practice engagement while the teacher was delivering science practices instruction. Definitions for student response behaviors of "on-task" and "off-task" were informed by previous research (e.g., Haley, Heick, & Luiselli, 2010; Kern & Dunlap, 1994; Northup et al., 1999). The PICI-S generates an estimate of the behavior of the whole class based on rotational observation of all students. For four consecutive 15-second intervals totaling 1 minute, a single student was selected at random, and that student's individual behavior was coded for those four intervals. For the next four intervals, another student was selected at random. This same process continued until the observation period ended. In general, each student was observed approximately two to three times. Interrater agreement for PICI-S data showed substantial Kappa agreement ($\kappa = .85$).

Results

Teacher Results

Intervention effects and planned comparisons can be found in Tables 2 and 3; descriptives are reported in Table 4. There were significantly higher results for the treatment group compared to the control for all teacher outcomes. Cohort effect was not significant, meaning that teachers from different years of participation did not differ significantly. Random teacher effects were generally significant (except for PICI), meaning that there was variation for these outcomes across teachers at baseline. Random school effects for teacher outcomes were not significant meaning that there was no variation across schools. In addition to documenting the effect of the summer institute plus coaching intervention, the a priori planned comparisons also showed the separate effects of the two intervention components to address the second research question.

Table 2. Teacher Science Practices Knowledge, Self-efficacy, and Beliefs Results

Term	Knowledge					Self-Efficacy					Beliefs				
	Est	SE	df	*p*	*d*	Est	SE	df	*p*	*d*	Est	SE	df	*p*	*d*
Fixed effect															
Intercept	.59	.02				80.39	2.19				3.79	.07			
Cohort (ref = 2013)	.01	.02	228	.55		-.8	2.08	306	.7		-.07	.07	102	.26	
Group (ref = control)	-.01	.02	228	.61		-2.16	2.05	306	.29		-.05	.06	102	.41	
Time (ref = Time baseline)[1]															
Time 2 (PtSumInst)	0	.01	228	1		0	1.27	306	1						
Time 3 (PtUnit)	-.04	.01	228	<.01		-1.58	1.47	306	.28						
Time 4 (EndofYear)						-.78	1.36	306	.57		-.04	.05	102	.36	
Group by Time															
Time 2 (PtSumInst)	.1	.02	228	<.01		13.26	1.78	306	<.01						
Time 3 (PtUnit)	.11	.02	228	<.01		17.87	2.02	306	<.01						
Time 4 (EndofYear)						16.54	1.89	306	<.01		.34	.06	102	<.01	
Random Effect															
Teacher Level	.01	.00		<.01		58.69	19.19		<.01		.06	.01		<.01	
School Level	0	.00		.53		19.43	17.94		.28		.001[3]				
residual	0	.00		<.01		48.9	3.96		<.01		.06	.01		<.01	
Planned Comparisons[2]															
SumInst	.10	.02	228	.01	.89	13.26	1.78	306	.01	1.21					
Coaching	.01	.02	228	.42	.16	4.62	2.02	306	.02	.53					
SumInst + Coaching	.11	.02	228	.01	1.06	16.54	1.89	306	.01	1.4	.34	.06	102	.01	1.03

[1]Time Point 2 is from baseline to after summer institute (effect of summer institute); time point 3 from baseline to post unit (effect of summer institute + coaching); and time point 3 for baseline to end of school year.
[2]SumInst is treatment–control comparison from baseline to PtSumInst; coaching is from PtSumInst to PtUnit.
[3]School effect is set to be .001.

Table 3. Teacher Science Practice Results

Term	TIR					EQUIP					PICI				
	Est	SE	df	*p*	*d*	Est	SE	df	*p*	*d*	Est	SE	df	*p*	*d*
Fixed effect															
Intercept	1.63	.18				1.95	.13				.07	.07			
Cohort	-.09	.14	168	.53		0	.1	168	.99		.14	.06	168	.02	
Group (ref = control)	.09	.18	168	.62		-.04	.13	168	.75		.04	.07	168	.56	
Time (ref = baseline)[1]															
Time 2 (PtSumInst)	0	.18	168	1		0	.12	168	1		0	.06	168	1	
Time 3 (PtUnit)	.24	.17	168	.16		.14	.11	168	.21		.07	.06	168	.28	
Group by Time															
Time 2 (PtSumInst)	.85	.24	168	<.01		.64	.16	168	<.01		.3	.09	168	<.01	
Time 3 (PtUnit)	1.08	.23	168	<.01		.79	.15	168	<.01		.4	.08	168	<.01	
Random Effect															
Teacher Level	.09	.05	168	.06		.08	.03		<.01		.01	.02		.65	
School Level	.001[3]					.001[3]					.02	.02		.35	
residual	.6	.06	168	<.01		.26	.03		<.01		.08	.01		<.01	
Planned Comparisons[2]															
SumInst	.85	.24	168	.01	.73	.64	.16	168	.01	.80	.30	.09	168	.01	.51
Coaching	.23	.22	168	.31	.05	.16	.15	168	.31	.15	.10	.08	168	.26	.23
SumInst + Coaching	1.08	.23	168	.01	.67	.79	.15	168	.01	1.00	.40	.08	168	.01	.99

[1]Time Point 2 is from baseline to after summer institute (effect of summer institute); time point 3 from baseline to post unit (effect of summer institute + coaching); and time point 3 for baseline to end of school year.
[2]SumInst is treatment–control comparison from baseline to PtSumInst; coaching is from PtSumInst to PtUnit.
[3]School effect is set to be .001.

Table 4. Teacher Descriptives																								
	Group																							
	Treatment												Control											
	Baseline			Post-SumInst			Post Unit			End Yr			Baseline			Post-SumInst			Post Unit			End Yr		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n
Knowledge (%)	60	9	63	69	11	62	67	10	57				60	12	61	60	12	61	57	13	63			
Beliefs	3.68	.34	63							3.99	.33	54	3.73	.35	61							3.70	.37	51
Self-efficacy (%)	78	13	63	90	8	62	94	5	46	93	7	54	80	13	61	79	13	61	79	13	40	80	13	50
Practice																								
TIR	1.63	.84	45	2.49	.82	53	2.96	.93	57				1.56	.73	39	1.56	.73	39	1.79	.86	53			
EQUIP	1.92	.54	45	2.55	.72	53	2.84	.41	57				1.94	.65	39	1.94	.65	39	2.08	.47	53			
PICI	.21	.36	45	.52	.41	53	.68	.22	57				.19	.34	39	.19	.34	39	.22	.27	53			

Results from all teacher-completed measures showed that teachers in the treatment condition had significant gains compared to the control condition when looking at total effects of the summer institute and coaching. The pattern of teacher outcomes differed across time points, however (see Figures 3 and 4). With only two data collection points, *beliefs* results showed a high effect size (d = 1.03) for baseline through both the summer institute and coaching. *Knowledge* results showed significant effects for the summer institute (d = .89) and sustainability throughout the coaching and classroom implementation (d = 1.06). There were four measurement points for *self-efficacy*. Results showed that gains for the treatment group were significantly higher compared to the control group for the effect of the summer institute (d = 1.21) and the coaching (d = .53), leading to a very large effect size for the total intervention (d = 1.40).

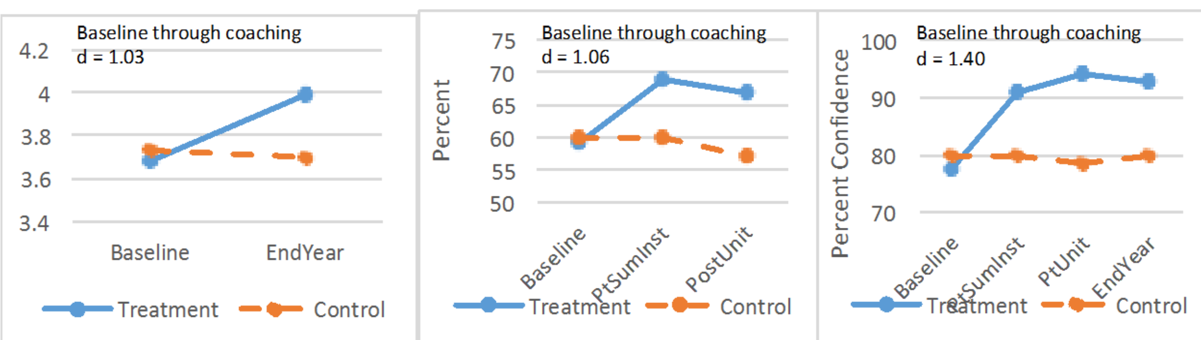

Figure 3. Teacher inquiry beliefs (left), knowledge (center) and self-efficacy (right)

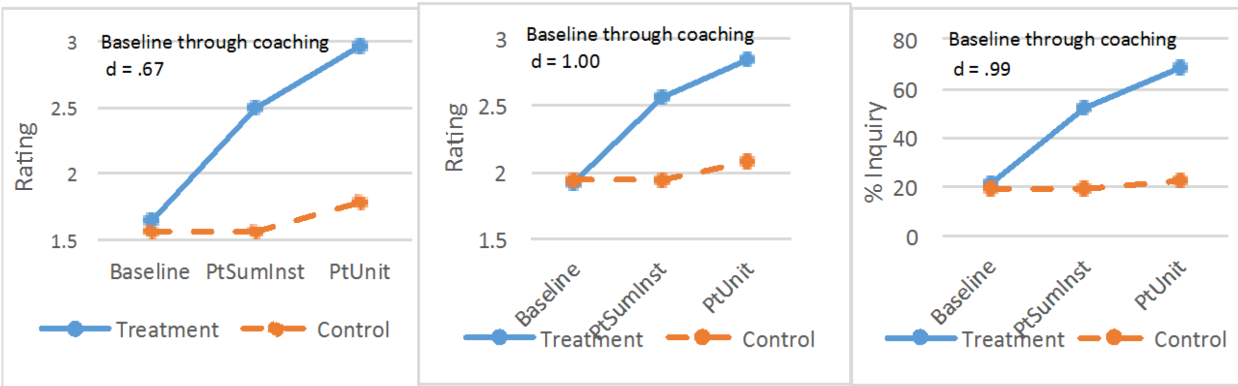

Figure 4. Inquiry instructional practice gains: TIR (left), EQUIP (middle), PICI-T (right)

Observations of teacher classroom implementation or *practice* of inquiry show a similar pattern of results for the three instruments, with all having significant effects for the combined summer institute plus coaching (Figure 4). When looking at a priori comparison results, the major contribution for this total intervention difference appears to come from the summer institute. Treatment–control comparisons for the summer institute were significant for all three measures while increases due to coaching were not. However, the graphs (Figure 4) confirm that there were increases as a result of the coaching, which contributed to the large effects for the total intervention.

In looking at the descriptive data, overall treatment teacher mean classroom performance ratings (on a 4-point scale) for the practice measures of the TIR and EQUIP were 2.96 and 2.84, showing that teachers basically performed between the "developing" and "proficient" levels. In contrast, control teacher means were 1.79 and 2.07, suggesting they basically performed between the "pre-inquiry" and "developing inquiry" levels. The PICI-T showed a significant difference between treatment and control conditions in percentages of time teachers delivered guided inquiry instruction in the classroom (70% treatment and 22% control).

Student Results

Table 5 presents student descriptives and Table 6 shows statistical results, which generally show significantly higher results for the treatment group. The one exception was for the high school knowledge outcome, which showed no significant treatment–control difference. However, the middle school results were significant, although with a small effect size (d = .18). Student *practice* of science, as measured by teachers rating each student's inquiry practices as

Table 5. Student Descriptives

	Group																	
	Treatment									Control								
	Baseline			Post Unit			End Yr			Baseline			Post Unit			End Yr		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n
MS Knowledge (%)	.66	.17	469	.73	.17	419				.66	.18	403	.70	.18	403			
HS Knowledge	.61	.20	425	.64	.20	426				.59	.21	480	.62	.22	389			
Self-efficacy	3.68	.67	890	3.79	.69	689	3.81	.72	690	3.79	.63	918	3.81	.70	574	3.83	.74	622
MS Practice				3.01	.76	409							2.75	.76	361			
HS Practice				3.06	.69	423							2.76	.85	403			

Table 6. Student Results

Term	Knowledge (Middle School)				Knowledge (High School)				Self-efficacy (Combined MS/HS)				MS Practice				HS Practice			
	Est	SE	df	*p*	Est	SE	df	*p*	Est	SE	df	*p*	Est	SE	df	*p*	Est	SE	df	*p*
Fixed effect																				
Intercept	.61	.03			.6	.03			3.75	.06			2.72	.16			2.62	.1		
Cohort	.05	.02	657	.04	0	.03	675	.95	.1	.05	2240	.06	.02	.16	721	.91	.25	.11	768	.02
Group (ref = control)	.01	.02	657	.59	.01	.03	675	.68	-.13	.05	2240	<.01	.24	.13	721	.06	.3	.11	768	<.01
MS/HS									-.05	.05	2240	.3								
Time (ref = baseline)[1]																				
Time 2 (PtUnit)	.05	.01	657	<.01	.03	.01	675	<.01	-.01	.03	2240	.67								
Time 3 (EndYr)									.05	.03	2240	.08								
Group by Time																				
Time 2	.02	.01	657	<.05	0	.01	675	.97	.14	.04	2240	<.01								
Time 3									.1	.04	2240	<.01								
Random Effect																				
Student Level	.015	0		<.01	.02	0		<.01	.22	.01		<.01								
Teacher Level	.003	0		<.01	.01	0		<.01	.03	.01		<.01	.08	.11		.47	.1	.06		.07
School Level	.001[2]				0	0		.86	.01	.01		.23	.06	.12		.59	.02	.05		.7
residual	.013	0		<.01	.01	0		<.01	.21	.01		<.01	.45	.02		<.01	.47	.02		<.01

[1]Time Point 2 is from baseline to post unit (effect of summer institute + coaching); time 3 is from baseline to end of year.
[2]School effect is set to be .001.

specified in the NGSS, was found to be significantly higher for treatment versus control at the high school level (d = .40), but not the middle school. However, the middle school probability level approached significance (.056) and showed a small–medium effect size (d = .34).

Student combined middle/high school self-efficacy results showed significantly higher increases (slope) for the treatment as compared to control as a result of their teacher's participation in coaching, d = .28, and the increase was sustained to the end of the school year, d = .19. There was a significantly higher level (percentage) of student inquiry engagement (student is on-task in response to teacher implementation of inquiry instruction) in the treatment group (67%) as compared to the control group (22%) during teacher implementation of science units.

Implementation Fidelity

Fidelity of instructional coaching was monitored via digital video and audio recordings of the coaching sessions. *Adherence* to coaching implementation was determined by the *Coaching Fidelity Checklist* which provided data on whether each established step of the coaching protocol was followed, including explicit coach discussion of teacher strengths, skills for improvement, and plan development. All recorded coaching sessions with good audio (n = 468) were coded by an independent coder, and 25% of the sessions were randomly selected for interrater agreement by a second independent coder. These randomly selected sessions represented approximately 25% of the coaching meetings conducted by each of the four coaches. Adherence to critical coaching items was coded. Example items were that the coach prompted/was successful in getting the teacher to a) identify positive student outcomes (from Student Inquiry Rubric), b) identify a skill strength that supported student outcomes, c) show video clip(s) to illustrate strength(s), and d) identify and discuss area(s) for improvement to support student outcomes. Coach behaviors coded included the coach showing video clip(s) to illustrate strength(s) in teacher skill that supported student outcomes. Adherence to critical coaching items was 91%. Interrater agreement was very high at 94%.

In addition to high levels of adherence to coaching protocols, the majority of the coaching sessions focused on relevant science skills content. Of the total average session length of 45 min 45 sec, only 39 seconds was casual conversation (i.e., conversation not related to the CSI coaching content, such as discussions about other school-related information such as prom or snow days).

Teacher participant responsiveness was based on the coder's determination of whether the teacher demonstrated evidence of being prepared for the coaching session. On average, teachers were prepared for 89% of the coaching sessions. Information used to help the coder make the decision if teachers appeared prepared included: a) the teacher made reference to something that occurred in the class period being discussed (average of 99%); b) teacher made reference to something about their execution of their teaching practices (average of 98%); c) teacher was prepared with ratings for the student inquiry rubric for the class as a whole (average of 84%); and d) teacher identified a video clip time stamp for an area of strength or improvement (indicated as present for an average of 64% of the sessions).

Discussion

Results showed that the teachers who participated in the summer institute and follow-up coaching had significant gains in beliefs, knowledge, and self-efficacy regarding the teaching of science practices when compared to a "business as usual" control. The intervention also resulted in significant changes in teacher practice. While the results for the total intervention were significant, the pattern of results across the various time points provides additional insight. In keeping with our hypothesized theory of change, teacher knowledge, self-efficacy and practice increased as a result of the summer institute. Results from this study support the idea that a summer institute plays an important role in providing teachers needed foundational knowledge and confidence to effectively implement instruction of science practices. The summer institute format, impacting multiple teachers simultaneously, provided an efficient and cost-effective way to deliver foundational knowledge, provide models of effective science teaching practices, begin building the coach–teacher relationship, and build teaching confidence through delivery of a practice lesson and receipt of feedback. We also found that the summer institute was instrumental in developing a common and shared language about teaching science practices, which carried over to the coaching sessions.

Treatment–control teacher knowledge comparisons showed a significant treatment effect as a result of participating in the summer institute. However, while the knowledge gained from the summer was sustained, it did not increase as a result of the coaching. This result was expected since a major goal of the summer institute was to provide knowledge about teaching science practices, including effective pedagogical approaches. The coaching, in turn, focused on translating this knowledge into practice. This result is in concert with previous research which concluded that conceptual understanding of inquiry-based teaching is a necessary, but not sufficient, prerequisite to teachers' successful implementation of inquiry practices (Lotter et al., 2013).

In contrast to the knowledge results, teacher *self-efficacy* treatment–control a priori comparisons showed significant effects for the coaching as well as the summer institute. These results are impressive since the treatment teachers' confidence scores following the summer institute were high (91% out of a maximum of 100%), making increases difficult. Results substantiate the importance of the coaching component in increasing teacher confidence in implementing a guided inquiry approach to teach student science practices. The support and feedback from a coach appears to be a critical element in increasing teachers' confidence as they move from gaining knowledge and practicing skills in the more controlled environment of a summer institute to the authentic environment of their own classroom.

While this foundational knowledge and self-efficacy were important, the critical outcome in this study was evidence that teachers could actually translate these outcomes into effective classroom instructional practice. Results showed that teachers who participated in the summer institute plus follow-up coaching had significant gains in guided scientific inquiry practice in authentic classroom instruction compared to "business as usual" control. This result was substantiated by three separate observational measures of science practice instructional performance, and the graphs of results across time points (Figure 3) display a strikingly similar pattern. However, a priori results showed that the only significant effect for all three measures was due to the summer institute—a result which supports the importance of providing opportunities for the teacher to practice teaching and receive feedback during a summer workshop. The coaching continued to increase the quality of teacher practice and contributed to the size of the total intervention treatment–control difference. The nonsignificant coaching result can partially be explained by the fact that the instructional practice that teachers experienced during the summer institute was in a controlled environment where teachers presented their lesson to peers, coaches, and project staff who acted as students. It may be that this "friendly" environment, with an informed, encouraging, and supportive audience, resulted in inflated ratings which made it difficult to show additional statistically significant increases as a result of the follow-up coaching. The difference in the two environments of practice teaching in the summer institute versus teaching in the actual classroom must be considered in interpreting these results.

This study showed small–medium effects sizes (d = .34 and .40) for student demonstration of science skills, which was the key student outcome. Student *knowledge* of science practices also showed significant effects at the middle school, but not the high school. We suggest that the results may be due to middle school students and teachers being more open to the guided inquiry approach and more flexible in its classroom implementation. Middle school students may also be more accepting of instructional approaches where teachers do not provide answers but instead encourage students to work problems out for themselves. Teachers in our study reported that high-performing students, particularly at the high school level, simply wanted to be given the material so they could memorize the answers for the test. As one teacher reported, “They’ve been so used to going through their science class and every other class taking notes, getting rote learning, and then turning around and spitting it back out on a test. They really don’t know how to learn.” High school teachers also tended to be very focused on their subject area and were content-driven with the need to cover the standards.

Results also showed a significant treatment–control difference in student science practice self-efficacy, both immediately after teachers had delivered their science practice unit and through the end of the year. The sustainability of this confidence is reflected in comments gleaned from teachers. They reported that student effects were not limited to the 6-week coaching window and were still evident months later as students continued applying the skills they had learned to new science content. As reported by one teacher, “I found three months later, I’d have the kid raise their hand and be like, 'So when we did such and such, you really wanted us to figure out that, you know, abiotic biotic and...' So three months later, it was like their brains were able to wrap around the idea."

This study was significant in a number of ways. Results substantiate the value of combining coaching with another form of professional development, as has been documented in other research (Powell et al., 2010; Kretlow et al., 2011; Allen et al., 2011; Gallucci et al., 2010; Matsumura et al., 2012). The study showed the effectiveness of the summer institute and coaching on teacher instructional practice, with effect sizes ranging from .67 to 1.40. While effects sizes for the isolated coaching component were generally small (.05–. 53), graphs shown in Figures 3 and 4 clearly show the continued improvement as a result of the coaching for the treatment teachers. Other science coaching studies reporting on *teacher practice* outcomes used correlational mediation analyses (Allen et al., 2011) or pre-post analyses (Lotter et al., 2013). In contrast, this study was a randomized controlled trial using rigorous treatment–control comparisons and multi-level modeling to account of the nesting of students within teachers and teachers within schools. In addition, *student* effects from other science coaching studies were derived from a quasi-experimental design (Vogt & Rogalla, 2009) or showed student effects only in the post-intervention year (Allen et al., 2011) after teachers had a year of experience in implementing coached strategies. In contrast, this study generally showed significant effects for student knowledge, practice, and self-efficacy during the year of the intervention (i.e., the first year of teacher implementation).

Like previous studies (Allen et al., 2011; Pianta et al., 2008; Powell et al., 2010; Vernon-Feagans et al., 2013), technology was found to be an effective and efficient way to deliver coaching to schools. The use of technology provided maximum flexibility in scheduling the coaching sessions; its use allowed coaching to happen at times convenient for the teacher (often early in the morning or late at night). The video recording and playback were also found to be critical in allowing teacher self-reflection of their teaching. We know from focus groups with teachers and coaches that teacher reflection from watching their videos was perceived as a critical component of the change process. As one teacher reported, “Perhaps the most valuable tool was to go through the videos and look for specific strategies used in class and how it impacted student learning and understanding. I feel much more confident in doing self-reflections of teaching strategies in class and plan to continue to video my classrooms on a periodic basis to critically look at teaching strategies used." This result is supportive of other research showing the importance of this reflection (Borko, Jacobs, Eiteljorg, & Pittman, 2008; Roth et al., 2011; Sherin & Han, 2004).

The technology-delivered coaching was also cost effective. Having coaches drive to participating school districts for face-to-face coaching would have resulted in prohibitive travel and personnel costs. The basic technology costs were $250 for cameras and $150 for microphones used by teachers to record their instruction. Cost analysis figures show a teacher cost of $1,410 (based on 42.75 total hours of preparing for coaching sessions plus direct coaching; see Author, 2015), a coach cost per teacher of $3,776 (based on a caseload of 15 teachers and an annual salary of $55,000), and a per-student cost of $5.76. In sum, the initial, onetime cost to provide teachers with these critical skills through support of science coaches is cost effective when you consider the benefit to teachers and students. These costs are in line with other virtual coaching estimates (Knight, 2012; U.S. Department of Education, 2015) and conclusions that virtual coaching models are a potential solution for providing PD in the midst of diminishing federal, state, and district resources (Ermeling, Tatsui, & Young, 2015).

Limitations and Future Research Directions

Because the sample consisted of rural science teachers and their students, results cannot be generalized to an urban population. However, urban school districts have expressed interest in our coaching model, reporting that the technology aspects of the project offer some clear advantages. First, it can reduce drive-time for coaches serving multiple schools. Second, the use of video recording for teacher self-reflection has distinct instructional advantages regardless of context and student population. However, additional research is needed to determine if the protocols used as the basis for this study produce similar results in other, non-rural teacher populations of the urban and suburban settings. Additionally, the procedures for conducting coaching sessions via distance-based technologies need to be further explored for its application in the more urbanized settings.

Another limitation relates to the teacher beliefs outcome in relation to our secondary study goal to isolate the separate effects of the summer institute and coaching. We were not able to do this for the teacher beliefs outcome, since it was only administered at the beginning and end of the year. Although our results showed that the total intervention of the summer institute plus coaching resulted in a change in teacher beliefs, one could hypothesize that coaching's direct connection to classroom practice would exert a stronger influence on teacher beliefs than the summer institute. To allow a more nuanced understanding of how beliefs may be shaped through the two separate professional development methods and how these two methods influence teacher practice, the beliefs instrument should also be administered immediately following the summer institute and after the completion of coaching. Only by obtaining results at these two time points will we be able to determine the unique effects of the coaching and summer institute components in influencing teacher beliefs about science inquiry.

Results from this study show the promise of coaching and its value added to traditional teacher in-service, but they do not provide definitive evidence regarding the impacts of the two individual components of a summer instructional institute and follow-up coaching. In order to clearly tease out the effects of coaching, the summer institute plus coaching condition must be compared with a summer institute *only* condition. It is entirely possible there would be a *decrease* from post summer institute to the completion of teacher science practice instruction for the summer institute only group due to teachers reverting back to their familiar ways of teaching instead of attempting to implement any new strategies. The coach was there to provide needed encouragement, positive or constructive criticism to improve teaching, and a level of accountability. One teacher succinctly summed up this coach role as "to hold you to it and keep you doing it."

This study focused on changes to teachers and students within a one-year period. We do not know if teachers continued to implement their newly learned skills in subsequent years and if students were able to apply newly learned skills in future science classes. Research has shown that student impacts from teacher coaching are often not realized until the second year—after teachers gain experience and have time to practice and internalize the new skills (Allen et al., 2011; Campbell & Malkus, 2011). This study's student effect sizes are not as strong as those for the teacher. A longitudinal study would provide additional insight into the long-term effects of coaching.

Research is also needed to dig deeper into why or how coaching leads to outcomes. Only with an understanding of the underlying workings of coaching will we be able to design, implement and scale up effective interventions to meet diverse needs of science teachers and their students. While this study shows the promise of coaching in impacting teacher change, we do not know what *specific* aspects of the coaching process (i.e., rapport and trust between teacher and coach, coach qualifications, teacher self-reflection) are most responsible for these effects. We do not know the optimal time that needs to be devoted to each step of the coaching process (i.e., joint planning, practice, reflection, and feedback). A necessary next step is to "unpack" the coaching intervention by operationalizing critical coaching elements and identifying which key components are most important in leading to desired outcomes. Most of the research in this area has utilized a case study approach, and there is a clear need for more quantitative approaches (Anderson, Feldman, & Minstrell, 2014). Such studies can also lead to conceptual models of coaching, showing relationships between critical variables that lead to impacts on teachers and their students.

References

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Action control: From cognition to behavior* (pp. 11–39). New York, NY: Springer-Verlag.

Akerson, V., & Hanuscin, D. (2007). Teaching nature of science through inquiry: Results of a 3-year professional development program. *Journal of Research in Science Teaching, 44*, 653–680. doi:10.1002/tea.20159

Allen, J., Pianta, R., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*, 1034–1037. doi:10.1126/science.1207998

Anderson, R., Feldman, S., & Minstrell, J. (2014). Understanding relationships: Maximizing the effects of science coaching. *Education Policy Analysis Archives, 22*(54). http://dx.doi.org/10.14507/epaa.v22n54.2014

Anderson, R. D., & Michener, C. P. (1994). Research on science teacher education. In D. Gable (Ed.), *Handbook of research on science teaching and learning* (pp. 3–44). New York, NY: McMillan.

Anthony, J., & Person-Pandil, S. (2001). *Student Inquiry Rubric*. Retrieved from http://www.education.ne.gov/science/Documents/InquiryRubric.pdf

Authors. (2015). *Active ingredients of instructional coaching: Developing a conceptual framework.* (Working paper)

Authors. (2015). *Coaching for change: Amount of instructional coaching support to transfer science inquiry skills from professional development to classroom practice* (Unpublished doctoral dissertation).

Authors. (2013, April). [Title masked]. Presentation at the National Association for Research in Science Teacher Conference, Rio Mar, Puerto Rico.

Authors. (2011, April). [Title masked]. Paper presented at 2011 International Conference of the National Association of Research in Science Teaching, Orlando, FL.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. (2008). Video as a tool for fostering production discussions in mathematics professional development. *Teaching & Teacher Education, 24*, 417–436.

Borman, J., & Feger, S. (2006). *Instructional coaching: Key themes from the literature.* Providence, RI: Educational Alliance.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academies Press.

Bybee, R., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science. *Journal of Research in Science Teaching, 43*, 349–352. doi:10.1002/tea.20147

Campbell, P., & Malkus, N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journa1, 111*, 430–454. doi:10.1086/657654

Campbell, P. H., & Milbourne, S. A. (2005). Improving the quality of infant-toddler care through professional development. *Topics in Early Childhood Special Education, 25*, 3–14. doi:10.1177/02711214050250010101

Capps, D., & Crawford, B. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education, 24*, 497–526. doi:10.1007/s10972-012-9314-z

Capps, D. K., Crawford, B. A., & Constas, 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.

Common Core State Standards Initiative. (2010). *Common core state standards for science and technical subjects.* Retrieved from http://www.corestandards.org/ELA-Literacy/RST/introduction

Cooper, J. O. (1987). Measuring and recording behavior. In J. O. Cooper, T. E. Heron, & W. L. Heward (Eds.), *Applied behavior analysis* (pp. 59–80). New York, NY: Prentice Hall.

Cornett, J., & Knight, J. (2009). Research on coaching. In J. Knight (Ed.), *Coaching: Approaches and perspectives* (pp. 192–216). Thousand Oaks, CA: Corwin.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad.* Dallas, TX: National Staff Development Council. Retrieved from http://www.learningforward.org/docs/pdf/nsdcstudy2009.pdf

Denton, C. A., & Hasbrouck, J. A. N. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational & Psychological Consultation, 19*, 150–175. doi:10.1080/10474410802463296

Desimone, L. (2009). Improving impat studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181–199.

Donovan, S. M., & Bransford, J. D. (2005). *How students learn: History, mathematics, and science in the classroom.* Washington, DC: National Academies Press.

Duran, E., Ballone-Duran, L., Haney, J., & Beltyukova, S. (2009). The impact of a professional development program integrating informal science education on early childhood teachers' self-efficacy and beliefs about inquiry-based science teaching. *Journal of Elementary Science Education, 21*(4), 53–70.

Ermeling, B. A., Tatsui, T. T., & Young, K. R. (2015). Virtual coaching for instructional leaders: A multi-method investigation of technology-enabled external assistance. *Teachers College Record, 177*(11), 1–63. Retrieved from www.researchgate.net

Fisher, W. W., Piazza, C. C., & Roane, H. S. (Eds.). (2011). *Handbook of applied behavior analysis*. New York, NY: Guilford.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). Tampa, FL: University of South Florida.

Foster, D., & Noyce, P. (2004). The mathematics assessment collaborative: Performance testing to improve instruction. *Phi Delta Kappan, 85*, 367–374. Retrieved from http://www.svmimac.org/images/Using_Formative_Assessment_to_Drive_Learning_Reduced.pdf

Gallucci, C., Van Lare, M. D., Yoon, I. H., & Boatright, B. (2010). Instructional coaching: Building theory about the role and organizational support for professional learning. *American Educational Research Journal, 47*, 919–963. doi:10.3102/0002831210371497

Garet, M., Porter, S., Andrew, C., & Desimone, L. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*, 915–945. doi:10.3102/00028312038004915

Haley, J. L., Heick, P. F., & Luiselli, J. K. (2010). Use of an antecedent intervention to decrease vocal stereotypy of a student with autism in the general education classroom. *Child & Family Behavior Therapy, 32*, 311–321.

Hanft, B., Rush, D., & Shelden, M. (2004). *Coaching families and colleagues in early childhood.* Baltimore, MD: Paul H. Brookes.

Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Merrill.

Hill, H. C. (2007). Learning in the teaching workforce. *The Future of Children, 17*, 111–127. doi:10.1353/foc.2007.0004

Ingersoll, R. M., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says.* Denver, CO: Education Commission of the States. Retrieved from http://www.ecs.org/clearinghouse/50/36/5036.htm

Jones, M. G., & Carter, G. (2007). Science teacher attitudes and beliefs. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1067–1104). Mahwah, NJ: Erlbaum.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kenward, M., & Roger, J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983–997.

Kern, L., & Dunlap, G. (1994). Use of a classwide self-management program to improve the behavior of students with emotional and behavioral disorders. *Education & Treatment of Children, 17*, 445–458.

Knight, D. (2012). Assessing the cost of instructional coaching. *Journal of Education Finance, 38*, 52–80.

Kretlow, A., Cooke, N., & Wood, C. (2011). Using in-service and coaching to increase teachers' accurate use of research-based strategies. *Remedial and Special Education, 33*, 348–361. doi:10.1177/0741932510395397

Llewellyn, D. (2002). *Inquiry within: Implementing inquiry-based science standards.* Thousand Oaks, CA: Corwin.

Lockwood, J. R., McCombs, J. S., & Marsh, J. (2010). Linking reading coaches and student achievement: Evidence from Florida middle schools. *Educational Evaluation and Policy Analysis, 32*, 372–388. doi:10.3102/0162373710373388

Lotter, C., Rushton, G, & Singer, J. (2013). Teacher enactment patterns: How can we help move all teachers to reform-based inquiry practice through professional development? *Journal of Science Teacher Education, 24*, 1263–1291.

Lotter, C., Yow, J. A., & Peters, T. T. (2014). Building a community of practice around inquiry instruction through a professional development program. *International Journal of Science and Mathematics Education, 12*, 1–23. doi:10.1007/s10763-012-9391-7

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin.

Luft, J. A. (2001). Changing inquiry practices and beliefs: The impact of an inquiry-based professional development programme on beginning and experienced secondary science teachers. *International Journal of Science Education, 23*, 517–534.

Malone, J. C. (1990). *Theories of learning: A historical approach.* Belmont, CA: Wadsworth.

Marshall, J., Smart, J., & Horton, R. (2009). Design and validation of EQUIP: An instrument to assess inquiry-based instruction. *International Journal of Science and Mathematics Education, 8*, 299–321.

Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2012). The effect of content-focused coaching on the quality of classroom text discussions. *Journal of Teacher Education, 63*, 214–228. doi:10.1177/0022487111434985

McGatha, M. (2009). Mathematics coaching: What does the research say? Retrieved from http://www.nctm.org/news/content.aspx?id=22839

Minner, D., Levy, A., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984–2002. *Journal of Research in Science Teaching, 47*, 474–496. doi:10.1002/tea.20347

National Center for Education Statistics (NCES). (2007). *Trends in International Math and Science Study: TIMSS 2007 results.* Retrieved from http://nces.ed.gov/TIMSS/results07.asp

National Research Council. (2011). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=13165

National Research Council. (2013). *Next Generation Science Standards.* Retrieved from http://www.nextgenscience.org/next-generation/science

Northup, J., Fusilier, I., Swanson, V., Huete, J., Bruce, T., Freeland, J.,Gulley, V., & Edwards, S. (1999). Further analysis of the separate and interactive effects of methylphenidate and common classroom contingencies. *Journal of Applied Behavior Analysis, 32*, 35–50. doi:10.1901/jaba.1999.32-35

Nugent, G., Barker, B., Toland, M., Grandgenett, N., Hampton, A., & Adamchuk, V. (2009). Measuring the impact of robotics and geospatial technologies on youth science, technology, engineering and mathematics attitudes. In G. Siemens & C. Fulford (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2009* (pp. 3331−3340). Chesapeake, VA: AACE. Retrieved from http://www.editlib.org/p/31957

Nugent, G., Toland, M., Levy, R., Kunz, G., Harwood, D., Green, D., & Kitts, K. (2012). The impact of an inquiry-based geoscience field course on pre-service teachers. *Journal of Science Teacher Education, 23*, 503–529.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*, 307–322.

Parnass, S. (2012, June 19). *National assessment of educational progress tests show students struggle to explain answers* [News blogs video community]. Retrieved from http://www.huffingtonpost.com/2012/06/19/tests-show-students-strug_0_n_1610213.html

Pianta, R. C. (2005). A new elementary school for American children. *SRCD Social Policy Report, 19*, 4–5.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431–451.

Powell, D., Diamond, K., Burchinal, M., & Koehler, M. (2010). Effects on an early literacy professional development intervention of head start teachers and children. *Journal of Educational Psychology, 102*, 299–312. doi:10.1037/a0017763

Rapp, J. T., Colby-Dirksen, A. M., Michalski, D. N., Carroll, R. A., & Lindenberg, A. M. (2008). Detecting changes in simulated events using partial-interval recording and momentary time sampling. *Behavioral Interventions, 23*, 237–269. doi:10.1002/bin.269

Renaissance Learning. (2015). *Going deeper: The role of effective practice in encouraging profound learning.* Wisconsin Rapids, WI: Author. Retrieved from http://doc.renlearn.com/KMNet/R0054160251DB7C5.pdf

Roth, K., Garnier, H., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. (2011). Videobased lesson analysis: Effective science professional development for teacher and student learning. *Journal of Research in Science Teaching, 48*, 117–148.

Sailors, M., & Shanklin, N. L. (2010). Introduction: Growing evidence to support coaching in literacy and mathematics. *The Elementary School Journal, 111*, 1–6.

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33*, 94–104. doi:10.1207/S15374424JCCP3301_10

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*, 1436–1460. doi:10.1002/tea.20212

Schuster, D., Cobern, B., & Applegate, B. (2011). *Assessing pedagogical content knowledge of inquiry science instruction.* Paper presented at the National Association of Research in Science Teaching annual conference, Orlando, FL.

Shapiro, E. S., & Kratochwill, T. R. (2000). *Behavioral assessment in schools: Theory, research, and clinical foundations*. New York, NY: Guilford.

Sherin, M., & Han, S. (2004). Teacher learning in the context of a video club. *Teaching & Teacher Education, 20*, 163-183.

Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and state-of-the-art analysis. *Educational Leadership, 45,* 77–87.

Sparks, D. (2002). *Designing powerful professional development for teachers and principals.* Oxford, OH: National Staff Development Council.

Strong, M. (2004). Induction, mentoring and teacher retention: A summary of the research. *The New Educator, 1*, 181–198. doi:10.1080/15476880590966295

Stroup, W. W. (2012). *Generalized linear mixed models: Modern concepts, methods and applications*. Boca Raton, FL: CRC Press.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2015). *Teacher training, evaluation and compensation intervention report: MyTeachingPartners—Secondary.* Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_mtp_061615.pdf

Vernon-Feagans, L., Kainz, K., Ginsberg, M., Hedrick, A., & Amendum, S. (2013). Live webcam coaching to help early elementary classroom teachers provide effective literacy instruction for struggling readers: The targeted reading intervention. *Journal of Educational Psychology, 105*, 1175–1187.

Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education, 25*, 1051–1060. doi:10.1016/j.tate.2009.04.002

Wei, R., Darling-Hammond, L., & Adamson, F. (2010). *Professional development in the United States: Trends and challenges.* Dallas, TX: National Staff Development Council.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K–12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.

Wilson, C., Taylor, J., Kowalski, S. M., & Carlson, J. (2010). The relative effects of equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching, 47*, 276–301. doi:10.1002/tea.20329

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee.edlabs