CrossMark

The Effect of Cognitive Apprenticeship-Based Professional Development on Teacher Self-Efficacy of Science Teaching, Motivation, Knowledge Calibration, and Perceptions of Inquiry-Based Teaching

Erin E. Peters-Burton[1] · Sydney A. Merz[1] ·
Erin M. Ramirez[1] · Maryam Saroughi[1]

Abstract This study investigated the effects of a 1-year professional development (PD) based on a cognitive apprenticeship model of research experiences on inservice teacher self-efficacy of science teaching, motivation, knowledge calibration, and perceptions of inquiry of 19 secondary earth science and biology teachers. The PD facilitator, who serves a dual role as a scientist and teacher educator, utilized a cognitive apprenticeship model to shape both scientific thinking and inquiry instruction with 19 inservice teachers. Results indicated that inservice teachers changed their perceptions of inquiry and maintained high self-efficacy throughout all phases of the study. However, teachers refrained from making long-term changes in their cognitive strategy instruction. Implications provide a fuller picture of teacher learning during a RET program, supported with inquiry instruction and the implications of cognitive apprenticeships in offering authentic science research experiences with minimal laboratory resources.

Keywords Professional development · Cognitive apprenticeship · Inservice teachers · Teacher self-efficacy · Motivation · Inquiry-based learning

Introduction

A major goal of teacher professional development (PD) programs in science is to offer experiences that build upon teacher knowledge and pedagogical content knowledge that result in increased implementation of inquiry and scientific practices in science classrooms (National Research Council, 1996; NGSS Lead States, 2013; SRI International, 2007). Not only have PD programs been shown to impact

✉ Erin E. Peters-Burton
  epeters1@gmu.edu

[1] College of Education and Human Development, George Mason University, 4400 University
     Drive, West Hall 2001, MSN 6D2, Fairfax, VA 22030, USA

improvement in the quality of teaching (Desimone, Porter, Garet, Yoon, & Birman, 2002), but also found to improve student achievement (Kardash, 2000; Seymour, Hunter, Laursen, & Deantoni, 2003). A popular and well-regarded type of PD for science teachers is the Research Experiences for Teachers (RET) program, where inservice teachers spend time in a research setting learning as an apprentice scientist (National Science Foundation, n.d.).

RET programs support mentoring activities common in the science community and model the mentoring of new scientists with experienced scientists to learn the nature of the discipline (Campanile Faurot, Doe, Jacobs, Lederman, & Brey, 2013; Campanile Faurot, Lederman, Jacobs, & Brey, 2014). Teachers who participated in scientific research experiences are reported to have used more authentic science concepts in their classes, were more confident in teaching science, and were more enthusiastic about science after the PD (Pop, Dixon, & Grove, 2010). Although RETs provide experience in the science discipline, they do not have intentionally structured components that support teaching in their inquiry instruction (Lederman & Lederman, 2014). The PD program featured in this study was unique because it provided inservice teachers with authentic science research experience in an economical fashion by using less resources and adding scientific inquiry to the RET experience.

The PD adopted a cognitive apprenticeship model, which has been found to be helpful in scaffolding the thinking and knowledge of the science discipline (Kardash, 2000) and could potentially be helpful in situating knowledge to transfer into curriculum. Cognitive apprenticeships offer the same mentorship experiences as RETs but can be conducted with fewer resources. For example with a cognitive apprenticeship, only one scientist is needed and that scientist can come to the school site to work with teachers rather than finding multiple professional science placements for teachers. Cognitive apprenticeships involved six characteristics as articulated by Collins, Brown, and Newman (1987): modeling, coaching, scaffolding (which develop cognitive and metacognitive awareness), articulation, reflection (which build problem-solving strategies), and exploration (which encourages independent practice). PDs involving authentic research experiences where teachers learn through cognitive apprenticeships with scientists have yielded positive results regarding teacher cognitive knowledge and noncognitive features of learning (Feldman, Divoll, & Rogan, 2007; Fernandez-Esquinas, 2003; Kardash, 2000). Cognitive apprenticeships that mentor scientific thinking and skills foster critical thinking, active construction of new knowledge, a collaborative climate, and intrinsically motivate learners (Pop et al., 2010). Learning is a complex phenomenon and is not only relegated to the accumulation of academic knowledge. Noncognitive skills such as learner's attitudes, motivation, and performance also influence learning achievement. A learner in a cognitive apprenticeship relationship needs to have confidence in what they are learning, be self-motivated to learn actively in the setting, be aware of their own capacity for learning, and be willing to teach using inquiry-based instruction for the PD to be successful. Given the cognitive and noncognitive outcomes of learners in a cognitive apprenticeship, this investigation explored how the specially designed RET experience affected the teachers' self-efficacy of science teaching, motivation, calibration of content

knowledge, and their perceptions of inquiry teaching before, during, and after a cognitive apprenticeship PD experience. Because key components of successful research experiences include intrinsically motivated learners and active construction of new knowledge through inquiry, it is fundamental to measure teacher motivation, calibration of content knowledge, and perception of inquiry over the PD experience. Due to the need for teachers to translate their experiences as students into learning environments, it is also important to measure self-efficacy of science teaching.

Self-Efficacy of Science Teaching

Teachers are considered to be one of the most important factors that impact student learning (Bandura, 1993; Tschannen-Moran & Woolfolk Hoy, 2001), and teacher efficacy is tightly connected with student achievement (Tucker et al., 2005). Teacher efficacy is defined by two important constructs, perceived self-efficacy and outcome expectancies. Perceived self-efficacy is a teacher's judgment of their capability for success in a particular domain (Bandura, 1986). A positive perceived self-efficacy of teaching can mediate the ability of teachers to maintain student engagement (Marzano, Pickering, & Pollock, 2001). High perceived self-efficacy of teaching is considered an important factor in successful implementation of differentiation of instruction, difficulty of task selection, student motivation, and positively affects teachers' beliefs about students, teaching and instructional behaviors (Klassen, Tze, Betts, & Gordon, 2011; Tschannen-Moran & Johnson, 2011; Tschannen-Moran & Woolfolk Hoy, 2001). The second factor of teacher efficacy is outcome expectancies, which are the levels of confidence a teacher has that her students will learn the content being taught (Bandura, 1997; Gibson & Dembo, 1984). Although teachers can be confident in their science teaching abilities, they may not always have the same level of confidence that their students will be able to perform or demonstrate their knowledge of the content (Bandura, 1993; Enochs & Riggs, 1990; Gibson & Dembo, 1984). A clear understanding of PD outcomes involves obtaining knowledge about both teacher personal self-efficacy and outcome expectations for their students.

Teacher Motivation

Teaching is a profession that requires lifelong learning in practice and learning from the context of the work (Darling-Hammond & Sykes, 1999). It demands being motivated to seek out possible opportunities to acquire knowledge and grow professionally (Randi, 2004). For cognitive apprenticeships like RET programs, motivation has been viewed through expectancy-value theory (Wigfield & Eccles, 2000), which suggests that a person's motivation to behave in a particular way is the product of expectations about his ability to perform a task (such as teaching science through inquiry) and the value the person has for that task. A person having high expectation for success and high value for the task will tend to persist and perform well in accomplishing the task successfully (Atkinson, 1964; Eccles, Adler, Futterman, Goff, Kaczala, Meece, & Midgley, 1983). Contextual factors also play a role in one's motivation to engage and be successful in a task; thus, learners who

have a part in choosing to perform a task tend to be more highly motivated (Ryan & Deci, 2000). Particularly in contexts such as cognitive apprenticeships, inservice teachers have many choices in terms of what they would like to learn from the experience and what they will bring to the classroom, and attention toward their motivational levels can help inform structures in the PD that will support high levels of learning and transfer of inquiry to the classroom (Zubrowski, 2007).

Calibration

Calibration has been defined as the relationship between one's perceived competence and one's actual performance (Stolp & Zabrucky, 2009). The cognitive processes required by calibration have been linked to those required of metacognition where one's confidence regarding the accuracy of his or her attained knowledge is evaluated. Therefore, one's calibration accuracy is determined by comparing the consonance between one's self-efficacy beliefs and actual performance (Klassen, 2002). As stated in the prior section, high self-efficacy contributes to effective instruction. However, a teacher must demonstrate competence in addition to having high self-efficacy in the learning domain in order to promote student success.

Much of the research in the field of calibration has been completed with students, but can be informative for PD contexts, where teachers are in the role of learners. Such studies involve students in clinical settings who read passages of text and then predict how well they will perform on ensuing comprehension tests (Butterfield & Metcalfe, 2006; Zabrucky, Agler, & Moore, 2009). Other research within the classroom setting has produced results that indicate students have moderate capability to calibrate their comprehension and perhaps even improve their calibration under certain conditions (Theide & Anderson, 2003). Findings consistently have demonstrated that high-achieving students calibrate more accurately as compared to lower-achieving students as they possess a higher metacognitive skill-set (Bol & Hacker, 2001; Pajares & Kranzler, 1995). It has also been suggested that the accuracy of calibration among lower-achieving students is due to the lack of knowledge about cognition, lack of ability to regulate cognition, or both (Schraw, Crippen, & Hartley, 2006). However, it must also be considered that the failure to improve calibration also has been attributed to lack of motivation among students (Hacker, Bol, & Bahbahani, 2008). Because cognitive apprenticeships have choices about what participants learn and their complex learning goals, an understanding of teacher ability to calibrate their knowledge of science practices can be another measure in determining the effectiveness of PD experiences.

Inquiry Teaching and Learning

Although science education reform documents have highlighted the need for more inquiry-based teaching for the past 20 years (Peters-Burton & Frazier, 2012), little has changed in the majority of classrooms in the USA (Capps, Crawford, & Constas, 2012). The variance in inquiry-based teaching definitions (Abrams, Southerland, & Evans, 2008) may contribute to the reluctance of teachers to adopt

this pedagogy, and it has been a factor in the shift from the use of the word "inquiry" to "science and engineering practices" in the NGSS (NGSS Lead States, 2013). For the sake of clarity in referencing studies prior to the adoption of NGSS, the term inquiry will be used in this section.

PD experiences play a critical role in the teachers' development of understanding inquiry and enacting inquiry (Blanchard, Southerland, & Granger, 2009; Breslyn & McGinnis, 2012; Dresner & Worley, 2006; Gyllenpalm, Wickman, & Holmgren, 2010; Ruebush et al., 2009). Opportunities for collaborative reflection support inquiry-based teaching (Lebak & Tinsley, 2010), and the amount of PD including collegial support in conjunction with the value that teachers placed on inquiry learning was significantly associated with student performance in inquiry-based instruction (Liu, Lee, & Linn, 2010). More specifically, PDs involving authentic research experiences using real-world data (Crawford & Cullin, 2004) and other experiences involving true immersion in an inquiry-based exercise allowed teachers to better understand authentic inquiry (Ruebush et al., 2009). Quigley, Marshall, Deaton, Cooke, and Padilla (2011) found that facilitating a discourse of hierarchical complex questioning of scientific content and inquiry to authentic materials using critical reasoning skills was more effective than teaching the scientific method in isolation. Breslyn and McGinnis (2012) found that the PD experience involved in obtaining a National Board Certificate encouraged biology and earth science teachers to increasingly enact inquiry in the classroom. Conversely, barriers to teaching inquiry include inadequate content preparation in science (Krajcik, Blumenfeld, Marx, & Soloway, 2000) and unfamiliarity with how science is practiced as a discipline (Deboer, 2004). The mounting evidence regarding the design PDs resulting in effective teacher implementation of inquiry in the classroom points toward having connected experiences with authentic research and immersion into inquiry instruction.

Purpose of Study

The purpose of this study was to examine inservice teacher self-efficacy, motivation, calibration of content knowledge, and perception of inquiry at different developmental times during a year-long PD experience. These variables were of interest to determine whether the PD featuring a cognitive apprenticeship with one scientist changed both cognitive and noncognitive features of teacher professional learning. The following questions were derived for this project:

(a) How did secondary biology and earth science teachers' *self-efficacy of and motivation for teaching inquiry-based teaching* change as they participated in a year-long cognitive apprenticeship-based PD program?

(b) How did secondary biology and earth science teachers' *calibration of their content knowledge* change as they participated in a year-long cognitive apprenticeship-based PD program?

(c) How did secondary biology and earth science teachers' *perception of inquiry teaching* change as they participated in a year-long cognitive apprenticeship-based PD program?

Methods

PD Framework: Cognitive Apprenticeship Model

In addition to adopting general characteristics of effective PD programs (Darling-Hammond & McLaughlin, 1995; Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2003), this PD utilized a cognitive apprenticeship model (Kardash, 2000) as a mechanism to meet the two intended goals: providing teachers authentic research experiences and developing inquiry-based units of study. An important characteristic of a cognitive apprenticeship model is the simultaneous exposure to the nature of learning and the practice of science. Teachers play the role of both learner and teacher while they acquire the knowledge and skills in their scientific domain. When teachers are competent learners, they can transfer content knowledge into pedagogical content knowledge, resulting in effective inquiry-based learning environments for their secondary students (Hashweh, 2003).

The science supervisors from the school district involved in this study were cognizant of the need for developing teachers' skills in scientific research, but did not have the resources to conduct an RET. Instead, they hired a scientist with some secondary education experience as a full-time professional development expert. The school district tasked this scientist initially to work with a small group of teachers who wanted more experience with scientific research and then help them develop appropriate units of study. The district anticipated that more teachers would get involved as the teachers became aware of the results of this program. The role of the research team was to evaluate teacher beliefs and perceptions during the initial year of this program. The scientist and the district science supervisors who designed the PD conducted a literature review on cognitive apprenticeships and noted the features of this type of educational experience based on the empirical articles. In the PD, the scientist deliberately exposed his thinking to the teachers about designing statistically sound scientific research, evaluating secondary data for validity of answering a research question, conducting descriptive and inferential statistical analysis, and communicating results as appropriate graphical representations. It was important that the designer of the PD had actually worked in the field of science and could authentically model scientific practices designs and/or implements a cognitive apprenticeship-based PD, otherwise the principles that anchor the mentorship may not be authentic to the scientific discipline.

PD Goals and Schedule

The goals of the year-long PD were to provide teachers a guided experience in scientific inquiry with the support of a scientist who was also a trained science educator, and work individually to produce inquiry units of study related to their research experience for secondary students (see Table 1). In Phase 1, the scientist who facilitated the cognitive apprenticeship modeled the processes of quantitative scientific research and mentored 19 teachers in scientific research, focusing on large secondary data set and its analysis during a 1-week, intensive summer institute.

Table 1 Cognitive apprenticeship-based professional development phases

Phase 1	Phase 2	Phase 3
Duration: 5 days (8 h per day)	*Duration*: 4 months (six meetings)	*Duration*: 6 months (six meetings)
Format: Face-to-face meeting	*Format*: Face-to-face meeting	*Format*: Remotely
Purpose: To conduct scientific research using secondary data sets because the teachers would not have access to the tools in a professional laboratory at their schools	*Purpose*: To have scientist provide follow-up personal assistance with teachers as they create their research projects	*Purpose*: To translate science research project created in Phase 2 into an inquiry-based curriculum for their secondary education students
Scientist's role: Revealed scientific thinking while answering research questions through a cognitive apprenticeship using large secondary data set (i.e., Weather Underground)	*Scientist's role*: Facilitated cognitive apprenticeship by providing assistance to teachers as they answer their research questions using secondary data sets	*Scientist's role*: Facilitated cognitive apprenticeship by providing assistance to teachers translate their science research projects using large secondary data sets into an inquiry-based curriculum
Teachers' task: Work in collaborative peer groups mentored by the scientist to engage in process of scientific research by locating reliable data, performing statistical analyses, making conclusions, and communicating conclusions	*Teachers' task*: Work independently with scientist to obtain background information, develop, and answer research questions for their choice of secondary data sets (i.e., NASA, NOAA)	*Teachers' task*: Work collaboratively with scientist in translating their science research project into curriculum for secondary science classrooms
Outcomes: Teacher experience in answering scientific research questions using secondary data sets from reliable, scientific resources	*Outcomes*: Teachers emulate scientific thinking and seek assistance when needed to answer a research question using secondary data set that is aligned with a topic from science curriculum	*Outcomes*: Teachers develop inquiry-based curriculum that reflects the content and processes learned in carrying out the original scientific research project

Since the teachers had little knowledge of statistics, learning how to use statistics to make claims in science was the focus of the PD. The scientist posed such questions as "Does the distance from the Earth to the sun influence temperature on the Earth?" and "What are the factors that influence peak colors of leaves in the fall?" to the teachers and scaffolded the teachers' processes of finding solutions to the questions. The teachers examined data from public Web sites such as Weather Underground, The Weather Channel, Leaf Watch, and Google Earth to test their null and alternative hypotheses. The PD included instruction for the teachers to be able to conduct descriptive and inferential statistics, both parametric and nonparametric, so that they were able to ask scientific questions and find solutions within secondary data sets. During this phase, teachers participated in a group activity with a common data set to reinforce analysis and statistical techniques. The scientist formatively assessed teacher understanding of scientific processes and content by engaging with one-on-one conversations with teachers daily to evaluate progress in

understanding how to find reliable data online, how to perform sound statistical analyses, and how to communicate data with graphical representations. The school district intentionally designed the PD to focus on secondary data sets because teachers would not have access to the tools in a professional laboratory. However, teachers and students had equal access to secondary data sets from reliable, scientific resources such as NASA, NOAA, and USGS public databases.

Armed with new knowledge about the role of statistics in science during Phase 2, 12 teachers developed their own projects for research with the support of the same scientist/science educator. This took place during the final days of the week-long summer institute and continued into the following 4 months. The teachers dug deeper into their inquiry to answer one of the two questions posed during the PD, "Does the distance from the Earth to the sun influence temperature on the Earth?" or "What are the factors that influence peak colors of leaves in the fall?" The teachers worked to obtain additional background information and develop research questions related to their choice of data sets that were available through online science databases in their content area. The teachers shared information via email about their progress when they needed assistance, and the scientist visited all 12 teachers four times during the 4-month period to discuss any issues with their analysis. These visits were required in the PD design. Finally in Phase 3, 12 teachers worked collaboratively over the next 6 months in two content area groups (biology and earth science), facilitated by the scientist, and translated their learning from the research project into inquiry units of study for their secondary science students. The scientist visited the teachers as needed and reviewed the final unit for scientific accuracy. The units of study were expected to model the process of inquiry that the teacher experienced during Phases 1 and 2 of the PD. Biology teachers modified their research project on leaf color to units of study for the secondary classroom, by modifying their research design or adding more samples for students. Similarly, earth science teachers modified their research project about the relationship of the distance between the Earth and the sun and temperature trends on Earth.

Participants

Nineteen secondary teachers participated in the study ($n = 9$ earth science teachers, $n = 10$ biology teachers), although seven of those teachers left the PD after the intensive week in the summer. All were high school teachers from the same school district in a suburban area of the mid-Atlantic region of the USA. The average teaching experience of the 19 teachers was 11.7 years. There were 7 males and 12 females; all teachers were white, non-Hispanic. Teachers who were participants in the PD volunteered for the experience.

Data Sources

Quantitative Measures

The study used a longitudinal, parallel mixed methods approach (Creswell, 2007) to determine the effectiveness of the PD with regard to self-efficacy of science

teaching, motivation, content knowledge calibration, and perceptions of inquiry over a year-long PD experience. Self-efficacy was measured with the STEBI-B (Enochs & Riggs, 1990). The STEBI-B was revalidated with high reliability ($\alpha = .90$) by Bleicher (2004) which reconfirmed the loadings of the two subscales of personal self-efficacy of teaching and outcome expectancies.

Motivation was considered an adaptation of a general expectancy-value model (see Pintrich, 1988; Pintrich & DeGroot, 1990) and was linked to three components (a) an expectancy component that included learner's beliefs about their ability to perform a task, (b) a value component that included learner's goals and beliefs about the importance and interest in a task, and (c) an affective component, which included learner's emotional reactions to a task and was measured by the widely used MSLQ. The test-anxiety subscale was removed from the MSLQ for this study because it was irrelevant. The subscales on the MSLQ showed high reliability, self-efficacy ($\alpha = .89$), intrinsic value ($\alpha = .87$), and cognitive strategy ($\alpha = .83$) (Pintrich & DeGroot, 1990).

Knowledge calibration has been typically studied with reading comprehension, but was adapted for use in the domain of scientific reasoning and content knowledge for this study. Knowledge calibration was measured by the SASKS, an National Science Foundation-funded assessment for scientific reasoning (Lawson, 2000). The SASKS was adapted for calibration measurement by placing an additional space for
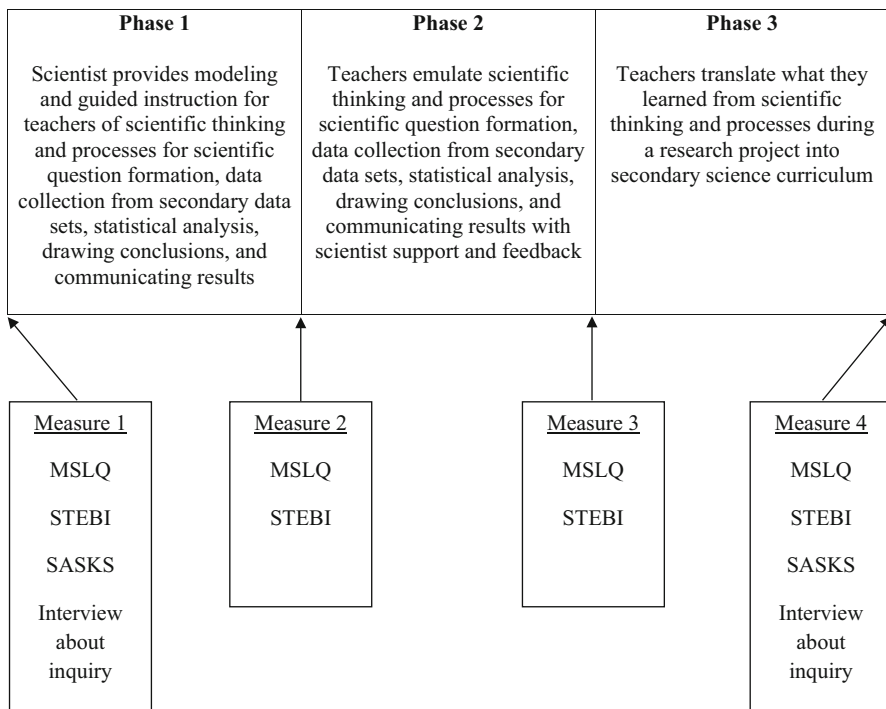

Fig. 1 Cognitive apprenticeship-based professional development measures

teachers to estimate their confidence in their ability to answer each question correctly on a scale of 1 (not confident) to 4 (very confident). For the sake of coherence, the correctness of the items on the SASKS was based on a scale of 1–4 to match the same interval for confidence ratings.

The self-efficacy (STEBI) and self-regulation (MSLQ) measures were administered during four points in time: (1) Phase 1; (2) after Phase 1; (3) after Phase 2; and (4) after Phase 3. The teachers completed the SASKS calibration test Phase 1 and after Phase 3 (Fig. 1).

Qualitative Measures

Qualitative measures were employed to gain an in-depth understanding of the participants' understandings of inquiry-based learning. Each participant participated in one face-to-face interview that lasted approximately 30 min before Phase 1 of the study ($n = 19$) and after Phase 3 ($n = 19$). The semi-structure interview protocol was adapted from Breslyn and McGinnis (2012) work regarding teachers' conceptions and enactment of inquiry. All interviews were audio-recorded with consent from the participants, transcribed verbatim, coded, and analyzed for salient themes. Qualitative and quantitative data were triangulated to determine the strength of the trends in the findings.

The scientist providing the PD was interviewed before the PD regarding the design, the intentions of the cognitive apprenticeship, and the learning outcomes. The scientist was also interviewed during the PD regarding the interactions of the day and changes to the PD design. Finally the scientist was interviewed during the time he was supporting the teachers in their independent research to determine the fidelity of teacher performance to the ways scientific research is conducted.

Findings

Quantitative Findings

Descriptive statistics were run on the responses to the MSLQ (motivation and self-regulation), STEBI (self-efficacy), SASKS (knowledge calibration) content questions, and the SASKS calibration evaluations. Inferential statistics for all measures were conducted using an ANOVA to determine significant differences in mean scores over time. Post hoc tests were performed to determine the statistically significant differences across each interval of time. In addition to the overall scores of each measure, the same tests were run for the subscales of the MSLQ and the STEBI. Three subscales remained intact for the modified MSLQ (self-efficacy, cognitive strategy use, and intrinsic value), and two subscales remained intact for the STEBI (personal science teaching efficacy and outcome expectancy). Correlations were run between the items on the SASKS and teacher confidence ratings for the items.

Modified MSLQ

There were no significant differences reported when the repeated measures were compared immediately before ($n = 18$, $M = 5.66$, SD $= .16$), after Phase 1 ($n = 17$, $M = 5.88$, SD $= .65$) and after Phase 2 ($n = 9$, $M = 6.15$, SD $= .37$) of the PD on the overall MSLQ scale, $F(1,19) = .98$, $p = .342$. Although the mean increased across the group, it was not significant, which was not surprising given the short-term and guided experience. Significant differences across time were approaching significance for one of the three subscales of the MSLQ, the cognitive strategies subscale $F(19,1) = 2.27$, $p = .06$. When post hoc tests were performed, it was found that there were significant differences between the cognitive strategy scales immediately before and after Phase 1 of the PD, $t(19,1) = 2.01$, $p = .06$, and after Phase 1 and after Phase 2 of the PD, $t(19,1) = 2.98$, $p = .05$, but no significant differences between the administration at Phase 1 and after Phase 2, $t(19,1) = .52$, $p = .61$. These results indicated that the participants changed their cognitive strategies while engaged in the intensive face-to-face week-long PD, but reverted back to their prior cognitive strategies 4 months after the face-to-face PD (after Phase 2). The other two subscales, self-efficacy and intrinsic value, did not demonstrate significance.

Scores for the self-efficacy were self-reported as high at the beginning of the PD and continued to be high after Phase 1 of the PD, resulting in no significant differences across time. Cognitive strategies may have approached significance because the PD experience was unique from teachers' past PD opportunities due to the concentration on content area rather than pedagogy. Participant scores increased after Phase 3 ($n = 4$, $M = 6.57$, SD $= .54$); however, these data were not included in the repeated measures analysis due to the low number of participants that had scores at all four points of measure ($n = 3$).

STEBI

No significant differences were found between Phase 1 ($n = 15$) of the PD ($M = 3.76$, SD $= .21$) and after Phase 1 ($n = 16$) of the PD ($M = 3.71$, SD $= .26$) and after Phase 2 ($n = 9$) of the PD ($M = 3.67$, SD $= .25$) mainly due to the high reporting of self-efficacy of teaching, similar to the MSLQ subscale. Similar means were reported for the PSTE and STOE subscales, which did not indicate significant differences over time. In congruence with previous findings, no significant differences were found after Phase 3 ($n = 9$, $M = 2.97$, SD $= .24$).

Calibration of Science Reasoning

Calibration was measured at the beginning of Phase 1 ($n = 19$) and was measured again after completion of Phase 3 ($n = 19$), the inquiry curriculum development phase. At Phase 1, teachers had low ability to calibrate the correctness of their answers. Teachers were often confident that their answers were correct when they were actually incorrect before the PD began ($r = -.45$). Although rare, the highest positive correlation between confidence of correctness and actual accuracy

was with questions involving volume comparisons ($r = .67$). The highest negative correlation occurred with questions involving predicted outcomes of physical interactions ($r = -.87$).

After Phase 3 of the PD experience, the teachers improved calibration of their scientific reasoning ($r = .67$). All teachers improved the calibration between correctness of the responses and the confidence about being correct. The calibration of knowledge and confidence levels on questions involving predicted outcomes of physical interactions, previously negatively correlated, demonstrated that teachers gained more insight into their awareness of how well they performed ($r = .32$).

Summary

The two notable instances of teacher change due to the PD, as measured by the three quantitative instruments, centered on cognition. The MSLQ subscale on cognitive strategies changed significantly over the course of the intensive week of PD and then reverted to prior cognitive strategies. This could be due to the nature of the PD, centered on statistics, which was an admitted weakness of the teachers. The teachers were learning intensely about statistics as applied to science and changed their cognitive structures during this time of learning new material and then assimilated the knowledge into their prior cognitive structures. Teachers calibration of knowledge on the scientific reasoning test (SASKS) also changed, which could be due to their experience with statistics in the PD and then grappling with making claims and reasoning in Phase 2 of the PD. The self-efficacy measures on both the STEBI and the MSLQ subscale did not show a significant difference, mainly due to the initial high scores. Since this was the first year of the PD, district science supervisors recruited early adopters for the project, since they were most likely to have buy-in. The teachers who are eager to learn new techniques and information tend to have higher self-efficacy (Bandura, 1986), and therefore, didn't produce differences in their scores before and after the PD.

Qualitative Findings

All interviews ($n = 19$) conducted at Phase 1 and after Phase 3 ($n = 19$) were transcribed from the audio recordings. Each transcript was coded independently by one of the four researchers on the team and then recoded by a second member of the team. From the interview transcripts, the researchers applied a priori codes adopted from Breslyn and McGinnis (2012) and coded meaningful statements that emerged outside of the a priori codes. The coding scheme required the first coder to highlight the text in the interview transcript that was associated with one (or more) a priori codes and/or an emerging code. The a priori codes included teacher definition of inquiry, teacher attitudes, and beliefs toward inquiry, teacher implementation of inquiry, role of the teacher in inquiry, role of the student in inquiry, implications of academic levels in inquiry, role of mathematics in inquiry, and sources of curriculum for inquiry. Codes that emerged from the data include the difference between theoretical understandings of inquiry and implementation of inquiry in the classroom and teacher perception of barriers that prevent implementation of inquiry-

based instruction. All discrepancies were discussed among the research team and resolved, leaving a very high inter-rater reliability.

Phase 1 Interviews

There was 92 % agreement among coding between raters for all transcripts in the before Phase 1 interviews. As codes were collapsed to make sense of the data, the following two themes formed from the Phase 1 data collection: misalignment between theoretical understanding and implementation and barriers of inquiry-based learning.

Misalignment Between Theoretical Understanding and Implementation

All of the teachers described the definition of inquiry as specific instructional method that tended to focus on the roles of the students, beginning with student-generated questions. Eleven teachers emphasized that inquiry should involve little to no guidance by the teacher, which is contrary to the cognitive apprenticeship model. Illustrating this idea, Teacher 14 stated,

> Inquiry-based to me, means that you really don't give the students a whole lot of guidance, you give them a rough outline of what they're doing…it's very little teacher guidance involved and for the most part they decide kind of where they want the activity to go…

Teachers were conflicted about letting students figure things out for themselves and scaffolding student learning. Although 11 of the teachers explained students should do inquiry with little or no guidance, nine of the 11 teachers also reported that their role in implementation of inquiry was to help students (a) identify reliable resources, (b) have regard for valid data, (c) manipulate the data correctly, (d) make logical conclusions, and (e) determine sources of error.

Complicating the issue of teacher support in inquiry was teacher beliefs about student ability levels. Twelve teachers noted that honors students seemed to possess an intrinsic scientific curiosity as compared to other students, named academic students, who, according to the teachers, relied much more on instructions provided by teachers and were primarily concerned with earning a grade. Teacher 4 articulated the difference she perceived between the students in the honors and the academic levels, "the honors kids are a little more inclined to be brave and show some independent thought whereas the academic [students] are; they're interested in getting it done, getting it in"… Teacher 9 best described the dichotomy of teacher beliefs about how to scaffold students of differing abilities:

> With my lower classes, I am more likely to give them the procedure, than any other classes. With the honors, they are more able to do things on their own…drawing conclusions from all the data that they would have collected because [honors students] like doing experiments and they will like actually like writing the conclusion; to [non-honors students], that is like the worse thing ever.

Barriers

Seventeen of the 19 teachers emphasized that planning inquiry lessons involved barriers including time limitations caused by meeting a large number of state standards and assessments, the availability of planning and implementation resources, incorporating students' prior knowledge, motivating students, and scaffolding the lesson so that all students can feel involved. It was commonly reported that teachers lacked funding for inquiry-based resources, and noted limited access to computers within the classroom. As acknowledged by Teacher 14, the sharing of inquiry-based lesson plans among colleagues could relieve some of the tension that they feel about teaching through inquiry. She noted, "I am pretty sure a lot of teachers are coming up with a lot of inquiry-based learning lessons and it would be a good thing if everyone could share." A number of teachers reported feelings of discouragement as they felt that the majority of students learned science material to achieve a grade as opposed to learning for the love of the discipline. Teachers also reported that students have difficulty answering open-ended questions that cannot easily be derived from the text. As a result, teachers found students using unreliable sources to answer the questions, such as Yahoo! Answers.

Role of Mathematics

Because the cognitive apprenticeship involved analysis of data sets using statistics, the role of mathematics in science instruction was part of the interview protocol. However, it was found that before the PD, mathematics instruction was not important to the teachers. The teachers felt that using measurements during investigations represented doing "a lot of math." Teacher 16 explained, "I personally struggle with statistics. Like with honors, I have to meet with the statistics teacher and he has to remind me about t tests and p values," and Teacher 3 reported "I obviously don't know anything about statistics or don't remember anything, but I know that they need to gather the right material, the right data in order to be able to do the statistics." Thirteen of the teachers were trying to "protect" their students from mathematics engagement during science. Teacher 11 discussed:

> They tend to have a fear of math. So, even when you are doing it and you think you're doing the inquiry-based to get them to the realizations, [the students have great difficulty] and it has to do with math.

Teachers also cited that the topics of biology and earth science were more difficult to incorporate mathematics in a natural way. "We don't have too much math intensity [in biology]. If you look at our state test, they get a four-function calculator. It is very laden with vocab—like learning a new language which terrifies some of them" (Teacher 12). Lack of teacher knowledge about statistics, coupled with teacher belief that students have a fear of mathematics, resulted in few opportunities to analyze quantitatively oriented problems. However, the PD was designed to help teachers learn about how statistics is used in science to make claims through completing the cognitive apprenticeship.

Phase 3 Follow-Up Interviews

All interviews occurred after the teachers attended the PD and implemented a portion of their newly created curriculum of inquiry-based learning in their science classrooms. Because of time constraints, these qualitative data were collected through three groups of 5–6 teachers and two individual interviews. The a priori codes and condensed codes for the after Phase 3 interviews were the same as the Phase 1 interviews, and there was 94 % agreement between raters.

Misalignment Between Theoretical Understanding and Implementation

Similar to the Phase 1 interviews, all teachers continued to describe inquiry as a student-centered approach; however, teachers shifted their beliefs from leaving the students alone to conduct the inquiry, to an environment where the teacher serves as the facilitator of learning. However, Teacher 8 brought up the power of technology in inquiry "as the role of the student and teacher is changing" because technology increasingly provided more access and opportunities to databases of information. The use of the secondary data sets in the PD seemed to shift teacher beliefs that students could engage in quantitative data analysis.

Implementation of inquiry-based learning in the teacher's classrooms produced a wider array of responses than the Phase 1 interviews. Seventeen teachers articulated the importance of building a shared network to increase their resources and ability to have common formative assessments with standards and benchmarks. An awareness of the need for teachers to collaborate and share resources was a shift in thinking from the Phase 1 interviews, suggesting that the PD created a community of scholars and a new learning network to share ideas. Because inquiry-based lessons took much time in preparation, combining resources, ideas, and diverse perspectives allows a new form of communication to occur.

This seemed to be related to how the teachers communicated their views on inquiry-based learning. Again, the majority of teachers spoke of how their ideas of inquiry have changed because of the experience in the PD. Others spoke of the importance of scientific thinking, implying that the PD provided them the opportunity to think like scientists and to apply this specific mindset. Another change from the Phase 1 interview responses was that more teachers ($n = 11$) began to believe that inquiry-based learning was for everyone and not just for honor or high-achieving students. The PD provided specific strategies to aid in this scaffolding process and it seemed to change some of the teachers' ideas about which students benefit from inquiry-based learning. In this, fifteen teachers spoke about giving students the most authentic experience and really pushing them to think on their own and use problem-solving skills. However, the other four teachers felt that pushing students to think on their own was difficult to do because they felt that students are unprepared for this level of thinking.

Barriers

In the follow-up interviews, 15 of the 19 teachers shifted how they viewed barriers in teaching with inquiry-based practices. In the Phase 1 interviews, teachers often expressed being held to standardized testing requirements, the lack of resources, difficulties of tapping into students' prior knowledge, motivating students, classroom meeting time, and scaffolding the lesson so that all students can feel involved. However, in the Phase 3 interviews, eleven teachers reported that they were re-energized to conduct inquiry because they found peers through the PD and at their schools who wanted to collaborate. This enabled teachers to tap into new set of resources as they combined sources and minds to better develop inquiry-based lessons.

Though four teachers reported that they continued to be challenged by the state standardized tests, the other fifteen teachers found ways around this problem and engaged administrative support in using inquiry-based learning (Focus Group 2), asking for longer class periods and better access to technology. Teacher 11 believed that students had better learning retention due to inquiry-based instruction, but didn't practice this type of instruction herself at a large scale. She felt that inquiry-based instruction took longer than traditional methods, and she was unwilling to change because she was concerned that her students wouldn't learn all of subject matter needed to be able to pass the state standardized tests. Additionally, ten teachers found that their students were more motivated and enjoyed inquiry-based learning as it forced them to think differently. These teachers explained that they were successful because they were better able to scaffold learning because of the modeling they received in the PD from the scientist.

Role of Mathematics

In regard to the use of mathematics, twelve teachers articulated a more connected approach to using mathematics in inquiry-based learning compared to the first round of interviews. Teacher 2 stated he "put in as much math as I can" pushing this at all student levels. During the following up interview, no teacher indicated that their students feared mathematics as some did in the Phase 1 interviews. However, Teacher 11 still expressed her desires to implement more mathematics into her science classrooms, but that her flexibility in doing so was constrained by preparing her students explicitly for the science concepts that would appear on the state standardized tests. Additionally, Focus Group 1 stated mathematics was mostly used through the means of graphing in earth science while biology students used statistics often.

Discussion

Capps et al. (2012) conducted a review of empirical literature on inquiry PD and found that no reported study has connected participation in a PD with all outcomes: (a) enhanced teacher knowledge, (b) change in beliefs and practice, and (c) enhanced

student achievement. Further, noncognitive skills are a central feature of successful learning in a cognitive apprenticeship, and few PD experiences have attempted to explain both cognitive and noncognitive changes in teachers in a systematic way. In an effort to address this deficit in the literature, this study measured teacher self-efficacy of teaching science, motivation toward teaching science, calibration of scientific reasoning skills, and perception of inquiry teaching and learning during several points of a year-long PD experience which engaged teachers in a cognitive apprenticeship of quantitative research experiences and immersion in inquiry teaching with a scientist trained in teacher education. The discussion is organized by research questions investigated in the study.

How Did Secondary Biology and Earth Science Teachers' *Self-Efficacy of and Motivation for Teaching Inquiry-Based Teaching* Change as They Participated in a Year-Long Cognitive Apprenticeship-Based PD Program?

Measures of self-efficacy indicated that teachers started the PD with a high self-efficacy in teaching science, maintained this high score across the week, and maintained high self-efficacy of teaching science 4 months after the intensive portion of the PD (Behar-Horenstien, Pajares, & George, 1996). Consistent with their high self-efficacy of teaching science, teachers shifted their concern regarding students with low academic achievement who may not benefit from inquiry before the PD to an expression that all students can learn through inquiry (Bandura, 1993; Gibson & Dembo, 1984) and even pass the state mandated tests. These results correspond to Pop et al.'s (2010) study that show cognitive apprenticeships intrinsically motivate learners. The modeling of science research experiences through the cognitive apprenticeship combined with an inquiry teaching focus may have helped teachers become more aware of how science is done, leading to the belief that all students can implement scientific practices in the classroom.

Although a significant change was not measured through the quantitative tools in the study, teachers explained in interviews before the PD that they had low self-efficacy for teaching mathematics, especially in the context of biology and earth science. This low self-efficacy of teaching mathematics most likely resulted in the belief that adequate mathematics instruction in science consisted of only measuring. However, teachers reported more value on the use of statistics in science instruction after the cognitive apprenticeship. These findings align with prior research demonstrating that self-efficacy is a driving factor in implementation of teaching (Bandura, 1993; Tschannen-Moran & Woolfolk Hoy, 2001). By the end of the Phase 3, teachers suggested that the collaboration during this phase may have helped raise their self-efficacy of teaching mathematics, particularly because they collectively saw the usefulness of knowing statistics to make interpretations in science. The outcome of teacher collaboration raising self-efficacy of teaching mathematics aligned with the outcomes of other cognitive apprenticeship models (Pop et al., 2010). The PD focus on secondary data analysis, rather than data collection, was most likely a factor in raising teacher comfort with teaching

mathematics because of the intense treatment of the importance of statistics in the apprenticeship.

Teachers scored high on the MSLQ immediately before and after Phase 1 of the PD, as well as scoring high 4 months after Phase 1 of the PD, indicating that they were prepared as learners to increase their inquiry teaching activities in the following year (Dembo, 2001; Paris & Paris, 2001), which allowed them to be more aware of their own knowledge and calibrate more accurately. As indicated by the awareness the teachers had for the benefits and barriers of teaching inquiry during the interviews, teachers' high motivation helped them to persist in learning about inquiry and scientific reasoning with the help of a professional scientist (Delfino, Dettori, & Persico, 2010; Randi, 2004). When teachers were assessed again using the MSLQ 9 months after the PD, their scores slightly increased; however, no significant differences were present. This aligns with the notion that the PD would aid in teachers' acquisition of knowledge, but not necessarily impact their cognitive strategy use, as the teachers have assimilated the new knowledge into existing mental models. Additionally, the small sample ($n = 3$) across all four time points made generalizations about all teachers impossible after Phase 3. Based upon this, however motivated the teachers were to learn how to teach inquiry, they did not preserve the ideas of evidence as promoted by the inquiry. It may be the high self-efficacy of teaching science that the teachers held over time that prevented them from making changes in their cognitive strategies in the long term.

How Did Secondary Biology and Earth Science Teachers' *Calibration of Their Content Knowledge* Change as They Participated in a Year-Long Cognitive Apprenticeship-Based PD Program?

One indication that the PD caused change in the teachers' belief structure was the ability to calibrate scientific reasoning (measured by SASKS). Initially, teachers had a negative correlation of their perceived ability to reason scientifically, and after all phases of the PD, the teachers were able to change a great deal to be confident in their correct answers and not confident in their incorrect answers before feedback (Pajares & Kranzler, 1995). Much like Theide and Anderson (2003) found, under the correct conditions, calibration can be improved in a relatively short period of time. The cognitive apprenticeship gave teachers the chance to see a model of scientific reasoning with statistics by the scientist and emulate scientific reasoning using statistics with their own investigation of leaf change or temperature. Their growth was reflected on the calibration measure of their scientific reasoning. Previously it was found that low-achieving students were not able to calibrate well (Schraw et al., 2006), but in this study we examined high-achieving teachers, in which the converse was true. High-achieving teachers were able to improve their calibration. Likewise, Hacker et al. (2008) found that students with low motivation were unable to improve their calibration. Based on the high self-efficacy scores throughout the study and that the teachers persisted throughout the PD, we can assume the teachers were highly motivated and were able to improve their calibration.

How Did Secondary Biology and Earth Science Teachers' *Perception of Inquiry Teaching* Change as They Participated in a Year-Long Cognitive Apprenticeship-Based PD Program?

In the Phase 1 interviews, teachers' definitions of inquiry were misaligned with their description of their implementation of inquiry. This misalignment corresponds with recent literature which discusses the ways teachers grapple with teaching inquiry (Gyllenpalm et al., 2010). From Phase 1 to Phase 3 interviews, the teachers appeared to be more excited and more connected to using inquiry-based learning and one group contributed such to learning to thinking like scientists, a key characteristic of a cognitive apprenticeship. After the PD, the teachers emphasized the helpfulness of collegial support in learning inquiry, which has been found as a factor in developing teacher inquiry skills (Liu et al., 2010). The format of a cognitive apprenticeship encouraged the communication of ways of thinking in all activities (Collins et al., 1987) and could have also illuminated the benefits of collegial support in learning inquiry. This mental frame of thinking was activated in the PD because the facilitator was a scientist and a science educator, knowledgeable in helping the teachers both to understand inquiry and to scaffold the processes to think like a scientist. In addition, because a cognitive apprenticeship provides a scientist as a model for learning scientific practices, this type of PD has potential to inform teachers about ways scientists participate in the discipline without needing access to a laboratory.

Our second finding revolves around deeper understanding of how inquiry-based learning can be applied for all ability learners. Before the PD, the teachers articulated that inquiry-based learning was more appropriate and often best utilized with their more advanced learners. They struggled to see the fit of using inquiry-based learning with their students in the general classes. However, after the PD, more teachers changed their minds and saw the overall benefit of inquiry-based learning for all learners. As such, teacher perception of what students of different ability levels are motivated by or can accomplish can be a barrier to proving inquiry learning for all students. A clear model of how to support scientific thinking through a cognitive apprenticeship could illustrate ways teachers can help students of all levels of academic performance by supporting their students without giving direct answers.

Our last finding relates to the role of mathematics in inquiry-based learning. While the teachers originally stated the role of mathematics in inquiry-based learning as a barrier, by the end of the year-long PD, many teachers transitioned their thinking of how the role of mathematics can be applied. In addition, no teacher indicated a fear for mathematics at the conclusion of PD. These findings provide evidence that a carefully scaffolded PD, utilizing a scientist as teacher educator, may provide means in helping teachers see mathematics in a different format, especially in biology and earth science. As such, the PD was designed so that it had the potential to address the reported teacher beliefs as it demonstrated how lessons can focus on freely available authentic data sets (i.e., mathematics) that answer a compelling question using science practices such as finding reliable resources.

Limitations

The first limitation is the study design only captured 1 year's data, starting with the PD to the curriculum implementation, and didn't include student outcomes. Furthermore, there are threats to the validity for this study. While rich data were collected using both qualitative and quantitative methods (Maxwell, 2005), the study did not include long-term involvement. Collecting data that relates to how the new curriculum related to student learning outcomes, and follow-up interviews with teachers, a few years of implementing the new curriculum is recommended for future research. Moreover, the original sample of this study included all those who participated in the PD; however, due to attrition, participant numbers dropped to 12. As such, there is "no guarantee that these informants' views are typical" (Maxwell, 2005, p. 91) and these results must be interpreted cautiously.

Another limitation is the use of self-report data. As with all self-report measures, teachers' participation in the PD over time became an issue. In this study, this was especially apparent in teachers' responses on the MSLQ measure in which only 3 out of 19 participants responded at every time point and the size of the sample decreased over time ($n = 18$, 17, 9, and 4, respectively). The STEBI showed similar trends with the sample size changing over time ($n = 15$, 16, 9, and 9, respectively). However, the SASKS was taken by all participants at both time points, perhaps suggesting that repeated measures should be done in less frequency to ensure stronger participant response rates. Further, the STEBI measure used to assess self-efficacy was designed primarily for use with elementary teachers which could have influenced how teachers' responded to the items. However, as the self-efficacy data showed nonsignificant change over time, this limitation does not influence the findings. One suggestion for future research is to gain a stronger response rate from participants over time to ensure that findings represent the total sample, and not simply those that chose to respond to all measures across time.

Implications

Authentic science research experiences as PD have been highly regarded for providing teachers with insight into scientists' work. However, they have been less successful in supporting teachers to use inquiry methods to teach science. This PD based on a cognitive apprenticeship model was unique in that it featured science content engagement similar to a Research Experience for Teachers (RET) program, but went beyond the RET because it was facilitated by a scientist who was a teacher educator expert in teaching through inquiry. The PD facilitator utilized a cognitive apprenticeship model in shaping both scientific thinking and inquiry instruction with the teachers. The format of facilitating thinking about scientific endeavors combined with how inquiry is carried out in the classroom supported the teachers' beliefs about successfully implementing inquiry instruction. Teachers' beliefs that lack of time and mandatory high-stakes testing are barriers for inquiry instruction have limited the shift to more student-centered instruction. However, the combining of cognitive

apprenticeships for science research experiences and inquiry instruction lends itself to a more successful PD structure. Not only that, but science teacher educators have been called upon to incorporate inquiry instruction into RET experiences (Lederman & Lederman, 2014), and this conversation also extends to our teachers and how teachers use and teach for scientific thinking.

The results of this study also support that successful PD for teachers regarding research experiences does not have to be conducted in a laboratory. The focus on secondary data analysis for PD experiences reduces the amount of necessary resources because the PD design is not dependent on laboratory facilities. Focusing on secondary data analysis requires only an Internet connection and computer equipment, resources that most school districts can access. Additionally, the focus in a laboratory RET might lend itself best to data collection rather than focusing on data analysis. The increased comfort with integrating mathematics and increased comfort with statistics as a basis for scientific claims could have been fostered by the focus in the PD on data analysis, which is timely considering the implementation of NGSS.

One consideration of the successes of this PD is that the facilitator was unique because he had experience as a field scientist, had worked professionally with secondary teachers, and taught secondary students, thus having the ability to translate between the science world and the science education world. It is reasonable to think that a pair or team of facilitators, some experts in science research and some experts in science education, could design and implement a PD with the same characteristics. A team of scientists and science educators working collaboratively to prepare a cognitive apprenticeship with the components of scientific research experiences combined with implementation strategies for inquiry teaching could provide a PD that results in increased implementation of inquiry for learning science effectively.

References

Abrams, E., Southerland, S. A., & Evans, C. A. (2008). Inquiry in the classroom: Identifying necessary components of a useful definition. In E. A. Abrams, S. A. Southerland, & P. Silva (Eds.), *Integrating inquiry into the classroom: Realities and opportunities* (pp. 11–12). Hartford, CT: Age of Information Press.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

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

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117–148.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Behar-Horenstien, L., Pajares, F., & George, P. (1996). The effect of teachers' beliefs on students' academic performance during curriculum and adaptations innovation. *High School Journal, 79*, 324–332. doi:10.1007/s10972-006-9015-6

Blanchard, M., Southerland, S., & Granger, E. (2009). No silver bullet for inquiry: Making sense of teacher change following an inquiry-based research experience for teachers. *Science Education, 93*, 322–360. doi:10.1002/sce.20298

Bleicher, R. E. (2004). Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers. *School Science and Mathematics, 104*(8), 383–391.

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *Journal of Experimental Education, 69*, 133–152.

Breslyn, W., & McGinnis, J. (2012). A comparison of exemplary biology, chemistry, earth science, and physics teachers' conceptions and enactment of inquiry. *Science Education, 96*, 48–77. doi:10.1002/sce.20469

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition Learning, 1*, 69–84. doi:10.1007/s11409-006-6894-z

Campanile Faurot, M., Doe, F., Jacobs, E. R., Lederman, N. G., & Brey, E. M. (2013). *From the undergraduate student perspective: The role of graduate students in an undergraduate research program.* Presentation at the national conference of the American Society of Engineering Education, Atlanta, GA.

Campanile Faurot, M., Lederman, N. G., Jacobs, E. R., & Brey, E. M. (2014). *Defining the roles of graduate student mentors and faculty advisors in undergraduate research programs.* Presentation at the international meeting of the National Association for Research in Science Teaching, Boston, MA.

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.

Collins, A., Brown, J. S., & Newman, S. E. (1987). *Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics* (Technical report no. 403). BBN Laboratories, Cambridge, MA. Centre for the Study of Reading, University of Illinois.

Crawford, B., & Cullin, M. (2004). Supporting prospective teachers' conceptions of modeling in science. *International Journal of Science Education, 26*, 1379–1401. doi:10.1080/09500690410001673775

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches.* Thousand Oaks, CA: Sage.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76*, 597–604.

Darling-Hammond, L., & Sykes, G. (Eds.). (1999). *Teaching as the learning profession: Handbook of policy and practice.* San Francisco, CA: Jossey-Bass.

Deboer, G. E. (2004). Historical perspectives on inquiry teaching in schools. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and the nature of science* (pp. 17–35). Boston, MA: Kluwer Academic.

Delfino, M., Dettori, G., & Persico, D. (2010). An online course fostering self-regulation of trainee teachers. *Psicothema, 22*, 299–305.

Dembo, M. H. (2001). Learning to teach is not enough—Future teachers also need to learn how to learn. *Teacher Education Quarterly, 28*, 23–35.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*, 81–112.

Dresner, M., & Worley, E. (2006). Teacher research experiences, partnerships with scientists, and teacher networks sustaining factors from professional development. *Journal of Science Teacher Education, 17*, 1–14. doi:10.1007/s10972-005-9000-5

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: W.H. Freeman.

Enochs, L., & Riggs, I. (1990). Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale. *School Science and Mathematics, 90*, 694–706.

Feldman, A., Divoll, D., & Rogan, A. (2007). *Research education of new scientists: Implications for science teacher education.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Fernandez-Esquinas, M. (2003). *From apprenticeship to training: An empirical enquiry into the preparation of scientist in Spanish Academic Science.* Cordoba, Spain: Institute for Social Studies of Andalusia.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*, 569–582.

Gyllenpalm, J., Wickman, P., & Holmgren, S. (2010). Teachers' language on scientific inquiry: Methods of teaching or methods of inquiry? *International Journal of Science Education, 32*, 1151–1172. doi:10.1080/09500690902977457

Hacker, D., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition Learning, 3*, 101–121. doi:10.1007/s11409-008-9021-5

Hashweh, M. Z. (2003). Teacher accommodative change. *Teaching and Teacher Education, 19*, 421–434.

Kardash, C. M. (2000). Evaluation of an undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. *Journal of Educational Psychology, 92*(1), 191–201.

Klassen, R. (2002). Writing in early adolescence: A review of the role of self-efficacy beliefs. *Educational Psychology Review, 14*, 173–203.

Klassen, R., Tze, V., Betts, S., & Gordon, K. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review, 23*, 21–43. doi:10.1007/s10648-010-9141-8

Krajcik, J. S., Blumenfeld, P., Marx, R. W., & Soloway, E. (2000). Instructional, curricular, and technological supports for inquiry in science classrooms. In J. Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry: Learning and teaching in science* (pp. 283–315). Washington, DC: American Association for the Advancement of Science.

Lawson, A. E. (2000). How do humans acquire knowledge? And what does that imply about the nature of knowledge? *Science and Education, 9*, 577–598.

Lebak, K., & Tinsley, R. (2010). Can inquiry and reflection be contagious? Science teachers, students, and action. *Journal of Science Teacher Education, 21*, 953–970. doi:10.1007/s10972-010-9216-x

Lederman, N. G., & Lederman, J. S. (2014). Talking the talk and walking the walk. *Journal of Science Teacher Education, 25*, 513–517.

Liu, O., Lee, H., & Linn, M. (2010). An investigation of teacher impact on student inquiry on student inquiry science performance using a hierarchical linear model. *Journal of Research in Science Teaching, 47*, 807–819. doi:10.1002/tea.20372

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Marzano, R., Pickering, D., & Pollock, J. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum and Development.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage Publications.

National Research Council. (1996). *National education science standards*. Washington, DC: National Academy Press.

National Science Foundation. (n.d.). *Research experiences for teachers*. Retrieved from http://www.retnetwork.org/about.php

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Pajares, M., & Kranzler, J. (1995). Self-efficacy and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 20*, 426–443.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*, 89–101.

Peters-Burton, E. E., & Frazier, W. (2012). Voices from the front lines: Alignment of reform documents and master teacher instruction. *School Science and Mathematics, 112*, 179–190.

Pintrich, P. R. (1988). A process-oriented view of student motivation and cognition. In J. S. Stark & L. Mets (Eds.), *Improving teaching and learning through research: New directions for institutional research* (pp. 55–70). San Francisco, CA: Jossey-Bass.

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40.

Pop, M. M., Dixon, P., & Grove, C. M. (2010). Research experiences for teachers (RET): Motivation, expectations, and changes to teaching practices due to professional program involvement. *Journal of Science Teacher Education, 21*, 127–147.

Quigley, C., Marshall, J. C., Deaton, C. C. M., Cooke, M. P., & Padilla, M. (2011). Challenges to inquiry teaching and suggestions for how to meet them. *Science Educator, 20*, 55–61.

Randi, J. (2004). Teachers as self-regulated learners. *Teachers College Record, 106*, 1825–1853.

Ruebush, L. E., Grossman, E. L., Miller, S. A., North, S. W., Schielack, J. F., & Simanek, E. E. (2009). Scientists' perspective on introducing authentic inquiry to high school teachers during an intensive three-week summer professional development. *School Science and Mathematics, 109*, 162–174.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111–139.

Seymour, E., Hunter, A. B., Laursen, S. L., & Deantoni, T. (2003). Establishing benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education, 88*, 493–534.

SRI International. (2007). *Evaluation of research experiences for teachers (RET) program: 2001–2006.* Prepared for The National Science Foundation, Division of Engineering Education Centers.

Stolp, S., & Zabrucky, K. (2009). Contributions of metacognitive self-regulated learning theories investigations of calibration of comprehension. *International Electronic Journal of Elementary Education, 2*, 7–31.

Theide, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129–160.

Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. *Teaching and Teacher Education, 27*, 751–761. doi:10.1016/j.tate.2010.12.005

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.

Tucker, C., Porter, T., Reinke, W., Herman, K., Ivery, P., Mack, C., & Jackson, E. (2005). Promoting teacher efficacy for working with culturally diverse students. *Preventing School Failure: Alternative Education for Children and Youth, 50*(1), 29–34. doi:10.3200/PSFL.50.1.29-34

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68–81. doi:10.1006/ceps.1999.1015

Zabrucky, K. M., Agler, L. L., & Moore, D. (2009). Metacognition in Taiwan: Students' calibration of comprehension. *International Journal of Psychology, 44*, 305–312.

Zubrowski, B. (2007). An observational and planning tool for professional development in science education. *Journal of Science Teacher Education, 18*, 861–884.