International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: https://www.tandfonline.com/loi/tsed20

Teacher pedagogical content knowledge, practice, and student achievement

Julie Gess-Newsome, Joseph A. Taylor, Janet Carlson, April L. Gardner, Christopher D. Wilson & Molly A. M. Stuhlsatz

To cite this article: Julie Gess-Newsome, Joseph A. Taylor, Janet Carlson, April L. Gardner, Christopher D. Wilson & Molly A. M. Stuhlsatz (2019) Teacher pedagogical content knowledge, practice, and student achievement , International Journal of Science Education, 41:7, 944-963, DOI: 10.1080/09500693.2016.1265158

To link to this article: https://doi.org/10.1080/09500693.2016.1265158

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
2019, VOL. 41, NO. 7, 944–963
https://doi.org/10.1080/09500693.2016.1265158

Teacher pedagogical content knowledge, practice, and student achievement†

Julie Gess-Newsome[a], Joseph A. Taylor[b], Janet Carlson[c], April L. Gardner[b], Christopher D. Wilson[b] and Molly A. M. Stuhlsatz[b]

[a]Division of Human Health and Wellness, Oregon State University-Cascades, Bend, OR, USA; [b]BSCS, Colorado Springs, CO, USA; [c]CSET, CERAS, Stanford University, Stanford, CA, USA

ABSTRACT
In this exploratory study, we attempted to measure potential changes in teacher knowledge and practice as a result of an intervention, as well as trace such changes through a theoretical path of influence that could inform a model of teacher professional knowledge. We created an instrument to measure pedagogical content knowledge (PCK), studied the impact of a two-year professional development intervention, explored the relationships among teacher variables to attempt to validate a model of teacher professional knowledge, and examined the relationship of teacher professional knowledge and classroom practice on student achievement. Teacher professional knowledge and skill was measured in terms of academic content knowledge (ACK), general pedagogical knowledge (GenPK), PCK and teacher practice. Our PCK instrument identified two factors within PCK: PCK-content knowledge and PCK-pedagogical knowledge. Teacher gains existed for all variables. Only GenPK had a significant relationship to teacher practice. ACK was the only variable that explained a substantial portion of student achievement. Our findings provide empirical evidence that we interpret through the lens of the model of teacher professional knowledge and skill, including PCK [Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). London: Routledge Press], highlighting the complexity of measuring teacher professional knowledge and skill.

ARTICLE HISTORY
Received 3 May 2016
Accepted 22 November 2016

KEYWORDS
Pedagogical content knowledge; teacher knowledge

Conceptual framework and theoretical path of influence

Higher-quality teaching resulting in increased student learning should be the outcome of professional development. Unfortunately, there are few studies that examine the impact of

CONTACT Julie Gess-Newsome julie.gess-newsome@osucascades.edu Oregon State University-Cascades, 650 SW Columbia St., Bend, OR 97702, USA

†This manuscript is based on material previously reported in the following presentation: Gess-Newsome, J., Carlson, J., Gardner, A. L., Taylor, J. A., Wilson, C., Stuhlsatz, M. (2012, March). *Promoting reform through instructional materials that educate*. A paper presented at the annual meeting of the National Association for Research in Science Teaching, Indianapolis, IN.

Supplemental data for this article can be accessed at https://doi.org/10.1080/09500693.2016.1265158.

professional development on teacher learning, classroom practice, or student outcomes. To complicate this lack of research, the typical science teacher participates in less than 10 hours of content specific professional development each year (Smith, Banilower, McMahon, & Weiss, 2002). We define high-quality teaching as the purposeful and deliberate planning of instruction that results in a coherent learning experience for students. To qualify as a high-quality instructional episode, a teacher must justify the importance and structure of the *content* they are teaching, their selection of *pedagogical* techniques, and how this combination effectively attends to the teaching *context* as it relates to meeting student-learning needs (background knowledge, preconceptions, and common areas of conceptual difficulty).

Teacher's professional knowledge has been a topic of interest to researchers for more than 50 years. Attempts to relate teacher background characteristics, such as courses taken and grade point average, found only very low correlations of these variables to student achievement (Wayne & Youngs, 2003). In examining teacher classroom actions, researchers were able to identify behaviours that predicted student achievement, but individual correlations were low and no combination of practices achieved superior results (Brophy & Good, 1986). In 1986, Lee Shulman proposed that educational researchers look more broadly at multiple aspects of teachers' professional knowledge. In this reformulation, one knowledge base, pedagogical content knowledge (PCK), captured great interest because it situated teacher knowledge and practice within the teaching of a specific discipline.

Researchers gravitated towards the idea of PCK even though it lacked a clear and consistent operational definition for research and measurement. When the research reported here started, there were no models of PCK that were built on empirical evidence and only a handful of research studies that were attempting to measure or capture aspects of PCK. Given the close match of our definition of high-quality teaching to PCK, we used the framework of teacher professional knowledge to guide our research. In this study we attempted to:

(1) Create an instrument to measure the professional knowledge, including PCK, of secondary biology teachers.
(2) Study the impact of a carefully designed two-year professional development intervention that included educative curriculum materials on teacher knowledge and classroom practices.
(3) Explore the relationship among teacher variables to attempt to validate a model of teacher professional knowledge.
(4) Examine the relationships among teacher professional knowledge, classroom practice, and student achievement.

Teacher professional knowledge

For this study, we specifically examined three aspects of the professional knowledge base for teaching: academic content knowledge (ACK, our reframing of Shulman's category of subject matter knowledge), general pedagogical knowledge (GenPK) and PCK. ACK was defined as the general factual knowledge that a teacher possesses about a specific topic. This knowledge was measured through a multiple-choice test. GenPK was defined as the ability to implement general teaching skills in the classroom and was measured using a classroom observation protocol. We proposed that PCK exists as a knowledge base as well as an aspect of practice. Based on this assumption, we measured PCK through written reflections, interviews, and observations of classroom practice using a PCK rubric. Similar to Magnusson, Krajcik, and Borko (1999), within PCK, we hypothesised that there were three internal constructs: content knowledge (PCK-CK), pedagogical knowledge (PCK-PK), and contextual knowledge (PCK-CxK). We define these constructs and describe their measurement as well as the intervention in the 'Methodology' section of this paper.

Since we started this study, there has been a significant expansion of the research related to PCK. Many researchers continue in the tradition of the early researchers by examining what teachers do and do not know about the teaching of a topic (Henze, van Driel, & Verloop, 2008), or the characteristics of teaching practice (Park & Oliver, 2008). Others (Baumert, *et al*., 2010; Hill, Rowan & Ball, 2005) have constructed careful, large-scale measures of PCK and used this information to examine the construct as well as teacher knowledge, practice, and student achievement. Perhaps more importantly, in 2012, 11 international research teams participated in a 5-day PCK summit (see Carlson, Stokes, Helms, Gess-Newsome, & Gardner, 2015 for details). At the PCK summit, participants struggled to make sense of disparate definitions of PCK, research programmes, instruments, and analysis techniques to fashion a vision for a structure of teacher professional knowledge that had its origins in empirical data. A result from the PCK summit was what is called the consensus model of teacher professional knowledge, including PCK. We draw on the results of the PCK summit as we examine the findings from this study and extend them to the consensus model.

Theoretical path of influence

Figure 1 presents our theoretical path of influence that the authors are attempting to validate in the study in order to inform a model of teacher professional learning. We propose that our *intervention*, the use of educative instructional materials featuring an inquiry-based learning approach combined with professional development, influences teacher knowledge bases (component 1 – impact of the intervention). These knowledge bases include ACK, GenPK, and PCK. These knowledge bases were the most closely related to the design of the intervention and thus the most likely to change. Next, these enhanced knowledge bases influence *teacher practice* to become more inquiry-oriented (component 2 – relationships among teacher variables). Finally, the changes in teacher knowledge and/or practice lead to greater *student achievement* (component 3 – relationship of teacher variables to student achievement). We used the three components in the theoretical path of influence for the purposes of data analysis and to organise our findings.

Design and procedure

Participants

A total of 50 self-selected high-school teachers from 15 different districts in the same state in the USA were accepted into Project PRIME in 2 cohorts staggered by 1 year. Participants committed to two years of teaching at least one section of biology using one of two highly educative curriculum programmes and to attend a professional development

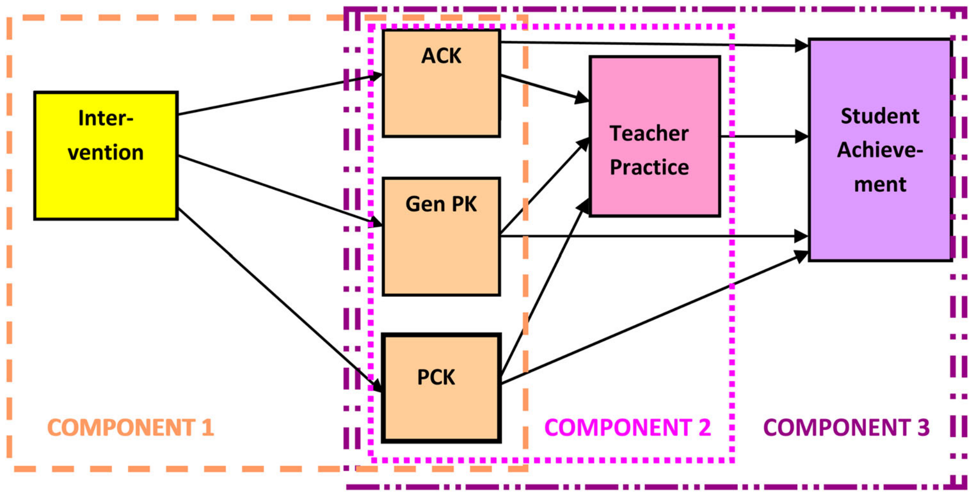

Figure 1. Project PRIME theoretical path of influence

programme in each summer with additional days during the two school years. Thirty-five of the participants completed the full two years of the programme.

There were more male teachers (67%) than female teachers (32%). Seventy-one percent of the teachers held a master's degree, while 29% reported only a bachelor's degree. The mean number of years teaching biology for the group was 7.4 years, with 1–20 years. Forty-four percent of the Project PRIME teachers reported having had at least 15 semester credits of prior biology course work (~5 courses) before participation with Project PRIME. The remainder reported less than 15 semester credits of prior biology course work.[1] Two teachers worked in self-contained special education settings; all others were in traditional classrooms.

Ninety-one percent of the teachers taught at urban schools, with the remainder teaching at rural schools. The mean percentage of limited English proficient students at the schools was 7% (range 0–23%). Eight teachers reported that over 50% of the students within their school were considered economically disadvantaged (average 24%). Nearly 30% of the student population at the Project PRIME schools was Hispanic, with six teachers reporting over 60% of the student population as Hispanic. Less than 5% of the total populations of the schools were African American, Asian, or Native American, but one school reported Native American populations of 24%. About 60% of the student population was reported to be White.

Intervention

Schneider and Krajcik (2002) define educative curriculum materials as those materials 'designed to address teacher learning as well as student learning' (p. 221). In addition, they say that:

> curriculum materials can be educative for teachers by offering support for teachers in thinking about: (a) content beyond the level suggested for students, (b) underlying pedagogy, (c) developing content and community across time, (d) students, and (e) the broader community. (p. 223)

Project PRIME commissioned a research team directed by Joseph Krajcik and Elizabeth Davis at the University of Michigan to review eight sets of high school biology curriculum materials with the goal of determining their potential for promoting teacher learning (see Beyer, Delgado, Davis, & Krajcik, 2009 for details of the analysis). Two programmes scored higher than the rest: *Insights in Biology*, developed by Education Development Center (EDC) and *BSCS Biology: A Human Approach* [*AHA*], developed by BSCS.[2] Both programmes were designed to support teaching biology content with conceptual coherence through the use of inquiry-oriented teaching practices. For this project, school teams selected one of the two curricula to implement over the two-year span of the study. Each curriculum contained units that corresponded to the state science standards (cell biology; heredity; interdependence; evolution; and matter, energy and organisation). For the purposes of this study, only data related to the teachers who selected the AHA curriculum are included because the majority of teachers selected this curriculum.

The professional development programme began with a guided curriculum selection process to select one of the two biology programmes noted above. After selecting a new programme, participants spent two or three weeks in the summer engaged in professional development experiences that focused on studying the curriculum materials, deepening CK, and expanding understanding of effective pedagogy related to two of the five biology topics from the state standards. During the next academic year, we requested that the teachers implement the full curriculum (all five topics). There were also collaborative lesson study sessions during the school year in which the teachers shared videos of their teaching and the associated student work based on lessons they co-planned during the summer as well as an additional meeting held in conjunction with the state professional science teachers' association conference. Attendance at this conference focused on content updates. During summer 2, participants attended a three-week professional development experience that focused on the last three units of the curriculum and again implemented the curriculum during the academic year. Overall, the teachers participated in over 250 hours of professional development across the two-years in addition to the time spent in curriculum implementation.

Measures

We used a mixed-methods data collection and analysis in order to capture the richness of the data we were seeking. Quantitative measures were balanced by qualitative data and analysis in order to triangulate the findings and enhance the meaning of the data. This article focuses primarily on the quantitative findings though we supplement this with trends observed in the qualitative data and representative quotes from the participants.

ACK was measured using the Major Field Test in Biology (MFTB). This test was developed by the Educational Testing Service as an exit examination for undergraduate biology majors.[3] Because the teachers in the programme were teaching biology, we believed that this test would be a reasonable measure of their CK. We administered the 150 question multiple-choice test upon entry to the project and at the exit from the programme. Categories of test questions (cell biology; molecular biology and genetics; organismal biology; populations biology, evolution, and ecology) mapped reasonably well to the five biology content standards addressed in the biology curriculum and the student achievement tests (cell biology; heredity; interdependence; evolution; and matter, energy and organisation) administered by the state. Scores were reported on a scale from 120 to 200 for the total test. Total scores from different forms of the test are comparable through equating an anchor block of common questions (www.ets.org/mft/faq). ETS scored the assessments. We used matched pair *t*-tests to examine knowledge growth from baseline to the end of the project.

To evaluate *GenPK* and *teacher practice*, we used the Reformed Teaching Observation Protocol (RTOP) to assess 4–6 randomly selected videorecorded classroom sessions from each year of teaching (Piburn & Sawada, 2000). The RTOP assesses the level of inquiry-based teaching that occurs in secondary science and mathematics classrooms. The RTOP consists of 25 questions. Upon observing a video of classroom teaching, each question was rated on a scale of 0–4 based on whether the behaviour never occurred (0) or was very descriptive of the lesson (4), with a total score range from 0 to 100. When used in secondary and university classrooms, the developers report final scores ranging typically from 18 to 98, with a mean score of 51.3 (SD = 20.1). Overall reliability for this instrument is reported as .954 (Piburn & Sawada, 2000). The developers established predictive validity of RTOP to student achievement through a .88 correlation between mean RTOP scores and normalised student gain scores in six university physics classes.

For the purposes of this study, we used a subset of questions to measure GenPK and teacher practice, as described in the following paragraphs (see Appendix S1 for supplementary material accompanying the online article). Item #8 did not load on any factor identified in the factor analysis and was not used in our study.

General pedagogical knowledge

We used the scores on eight items in the community of learners area as a measure of teachers' GenPK. According to the protocol developers, 'All of these items reflect the central notion of a classroom as a place where students work together to learn. This is distinct from the content of the lesson, and goes beyond a simplistic notion of inquiry' (Piburn & Sawada, 2000, p. 20). In selecting a classroom observation protocol to measure GenPK, we recognise that we are in fact defaulting to the measurement of general pedagogical skill. We believe that general teaching knowledge has the potential to supersede specific content, but are not convinced that there is a simple translation of this knowledge to practice. Thus, a measure of general pedagogical skills seemed to be a more direct measure of the pedagogy that students experienced.

Teacher practice

The scores on the remaining 18 items in the RTOP were used as a measure of teachers' inquiry-oriented practice, and thus became our measure of teacher practice. Since the educative curriculum materials were designed to support inquiry-oriented instruction, we felt that this would be an appropriate measure of the influence of the intervention. More specifically, the test developers defined these questions as representing inquiry-based teaching practices, including the pedagogy of inquiry, an understanding of the content needed to facilitate inquiry instruction, and the ability of a teacher to refocus the nature of the lesson to attend to student comments. We believe that this skill set represents the intersection of the knowledge and skill to effectively teach a science lesson in the spirit of inquiry-based instruction.

Total RTOP scores from our population ranged from 9 to 80 (maximum score is 96 points), with a mean of 39.9 (SD = 14.6). These numbers compare reasonably well with those of the norm group used by the test developers: (range = 18–98, $X = 51.3$; SD = 20.1). Gains in GenPK and teaching practice were determined by matched pair *t*-tests comparing baseline scores to the highest score obtained in Year 2.

Pedagogical content knowledge

Based on a review of the literature, we envisioned PCK consisting of three internal constructs: PCK-CK, PCK-PK, and PCK-CxK. From this review, we operationally defined the internal constructs as follows:

- *PCK-CK* including accuracy of CK; connections within and between topics and the nature of science; and use of multiple modes of representation or examples of a topic;
- *PCK-PK* including a rationale linking teaching strategies to student learning; strategies for eliciting student prior understandings; and strategies to promote student examination of their own thinking; and
- *PCK-CxK* including understanding how student variations, such as student prior conceptions, impact instructional decisions.

Based on these definitions, we designed (1) the Project PRIME PCK Reflection Instrument and (2) the Project PRIME PCK Rubric. (See Appendix S2 for supplementary material accompanying the online article. Also see Gardner & Gess-Newsome, 2011, for a detailed discussion of instrument development, use, and theoretical framing.) The Project PRIME PCK Reflection Instrument was based on the assumption that PCK is embedded in the act of planning for and envisioning instruction (knowledge). Based on a review of the literature at the time, we focused on capturing teachers' thinking about what and how they plan, and then probed for specific information as one mechanism for eliciting PCK. Proceeding with the assumption that PCK is topic-specific, we based data collection on the teaching of a specific topic (i.e. cell transport mechanisms, photosynthesis). The Project PRIME PCK Reflection has two parts (see Appendix S3 for supplementary material accompanying the online article). In the first section, teachers were asked to think about teaching a specific topic in their classroom. Through a series of steps, they were asked to describe a lesson that exemplified the topic by explaining, in a step-by-step format, what he or she was doing, what the students were doing, and the rationale for the instructional decisions included in that step. Teachers could use instructional materials to assist them in completion of the reflection. Most teachers described their instruction for a given lesson in 4–10 steps. After describing the lesson sequence, teachers responded to set of questions designed to elicit their instructional decision-making. The Project PRIME PCK Reflection Instrument was the basis of both written reflections and interviews.

The Project PRIME PCK Rubric was designed and organised by each component of PCK. Within each component are criteria that earn a score of 0, 1, 2, and 3, corresponding to limited, basic, proficient, or advanced teacher knowledge (see Gardner & Gess-Newsome, 2011). Possible scores by components are: PCK-CK: 0–12, PCK-CxK: 0–3, PCK-PK: 0–9. Total scores could range from 0 to 24. For our population, total PCK scores (averaged across the topics) ranged from 2.5 to 16 with a mean of 9.03, SD =

3.31. In this project, teachers completed the PCK Written Reflections on five different lessons (one from each area of the state standards: cell transport mechanisms, protein synthesis, carrying capacity, natural selection, and photosynthesis) four times across the two-year study. All participants also completed a PCK Interview Reflection on one lesson at the end of the project. In addition, the Project PRIME PCK Rubric was used to assess teachers' PCK as exhibited in the classroom based on video-recordings of their instruction for two of the topics (natural selection and cell transport mechanisms).

Inter-rater reliability for the Project PRIME PCK Rubric was calculated to test for agreement in scoring between raters. A sample of approximately 10% of the Written Reflections was scored by two raters, and inter-rater reliability was calculated using the intraclass correlation coefficient. The results showed no significant differences between raters. Throughout all scoring, researchers were blind to both the teacher's identity and the timing of the Reflection.

The two raters also used the Project PRIME PCK Rubric to score video recordings of class sessions (skill) based on the same topics as the Reflections. Several videos were scored by both raters to facilitate an estimate of inter-rater reliability, which was calculated using the intraclass correlation coefficient. These results also showed no significant differences between raters [$F(1, 6) = 2.455$, $p = .178$]. As with the Written Reflections, scorers were blind to the identity of the participant and to the timing of the lesson.

We determined gains in PCK using matched pair *t*-tests comparing baseline subscale scores on the PCK reflections to the highest subsequent score. We based the decision to use the highest subsequent score on the assumption that participants would not *lose* knowledge, although they may put less effort into completing the reflections at points later in the study. Participants confirmed this assumption in post-programme interviews.

Student achievement was measured using a researcher-developed test drawn from a large test bank of items from BSCS. We selected items to create five subscales (cell; heredity; interdependence; evolution; and matter, energy, and organisation) with 20 items each, and to match the state biology standards and the content from *AHA*. Items were selected according to a table of specifications to address state standards. In each subscale, 16 questions measured conceptual knowledge and 4 measured application of knowledge. Subsequent analyses eliminated items that were not functioning well; so the final instrument included 5 subscales with 16–20 items. The final test consisted of 90 items. The participating teachers' students completed the test at the beginning and end of each academic year. Cronbach's α for the five subscales on the instrument ranged from .68 to .80.

Teacher interviews were conducted at two points – at the conclusion of the programme and a year or two later. We interviewed all 35 final participants at programme end asking for their reflections on the project and its impacts. Nineteen teachers participated in the second interview.[4] Specific questions included their views of exemplary teaching, their definitions of the components of PCK, and if and how the project had changed their thinking or teaching practice. We asked the teachers to examine their personal teacher and student-level data collected across the project, including MFTB scores, PCK Reflection scores, RTOP scores, and student achievement gains and explain the results presented. Each interview set was analysed separately using open coding for themes. In addition, case synopses from the first interview were provided for member checking in the second interview. The interview results were compared to each other and to the quantitative data as a means to assist in creating meaning from the data through a cross-case analysis.

Analysis and findings

Pedagogical content knowledge

While our initial model of PCK included three constructs (PCK-CK, PCK-PK, PCK-CxK), confirmatory factor analysis revealed that PCK-PK and PCK-CxK were not fully distinct. The PCK-CK construct remained mostly distinct, with factor loadings for the four items ranging from .613 to .740. Factor loadings for the remaining four items (three from our PK subscale and one from our CxK subscale) indicated that these two subscales were not distinct. The factor loadings on the amalgam of PCK-PK and PCK-CxK for these items ranged from .470 to .789, while all loadings on PCK-CK were less than .40. Thus, based on our data, we revised our three-construct model of PCK to include only two distinct constructs that we call PCK-CK and PCK-PK. This structure was supported by teacher interviews and researcher perceptions. Both groups found difficulty in distinguishing the consideration of student variations from the selection of instructional strategies and attention to student prior misconceptions. According to our participants, 'context knowledge is huge and it recognises your students' prior conceptions and give you a starting point [for planning instruction]'. When asked if PCK-CxK was the same or different from PCK-PK, several teachers indicated that they saw a close relationship between the two:

> It is part of the teaching aspect; understanding the variations in students and what they are coming in with.

> I didn't differentiate between PCK-PK and PCK-CxK. I think the context is somewhat in the pedagogical knowledge.

> You are really looking at instruction and how it impacts student learning and I think that is the big goal here. We want these kids to learn.

We believe that this research constitutes one of the few studies that has empirically and quantitatively examined the construct of PCK, especially in science. While our definition of PCK varies from that of others (see Kirschner, Borowski, Fischer, Gess-Newsome, and von Aufschnaiter 2016 for an analysis of definitions of PCK), there seems to be sufficient evidence that there are at least two internal constructs of PCK related to application of topic-specific content understandings to the classroom, and the considerations of student understanding variables to the selection of instructional practices. This structure has a strong correspondence to the definition of PCK proposed by Shulman in 1986: knowledge of instructional strategies and subject matter representations for teaching, and knowledge of student understandings related to a specific topic.

Impact of the intervention

Component 1 of our theoretical path of influence concerns the impact of the intervention on teacher knowledge bases. We used two-tailed, matched pair *t*-tests to examine these relationships, comparing baseline scores to the average of Year 1 and 2 scores for ACK, GenPCK, and teacher practice, and comparing baseline scores on the PCK Written Reflections to the highest subsequent score for PCK-CK and PCK-PK. From baseline to programme end, all of the teacher knowledge bases increased significantly following the intervention (Table 1).

Table 1. Impact of intervention on teacher knowledge bases and practice.

Knowledge base or classroom practice	df	*t*	*p*	Pre–post effect size
ACK	33	5.114	<.001	0.88
GenPK	17	1.441	<.001	0.88
PCK-CK	26	6.094	<.0001	1.66
PCK-PK component	26	5.645	<.0001	1.59
Teaching practice	17	1.210	<.0001	0.71

Teachers'*ACK* scores increased significantly on the total MFTB test from baseline (X = 154.97, SD = 14.173, judged to be at the 45th percentile of national test takers) to post programme (X = 161.56, SD = 13.632, 65th percentile). This finding is of interest because during the interviews only a small number of teachers mentioned increases in CK as something they thought improved and many admitted to carelessly completing later administrations of the test and yet we see positive and significant gains in CK.

Teachers'*general PK* as measured through classroom observations increased over the two years (baseline: X = 7.78, SD = 4.28; Years 1 and 2 average: X = 12.89, SD = 2.72). This result indicates an improvement in overall skills related to creating a community of learners in the classroom, including being a skilled listener and acting as a classroom resource as opposed to an academic authority.

Scores for both *PCK-CK* and *PCK-PK* increased (PCK-CK baseline from X = 4.52, SD = 1.57 to highest subsequent X = 7.11, SD = 1.55; PCK-PK baseline from X = 2.24, SD = 1.22 to highest subsequent X = 4.19, SD = 1.30). Although these changes suggest that teacher's topic-specific consideration of content connections and representations, inquiry-based instructional strategies, and attention to student backgrounds and misconceptions increased as a result of the use of the curriculum materials and the professional development, participant views about where learning occurred varied.

According to our participants:

> The professional development was extremely helpful because I could understand the rationale for doing certain activities in the order they were meant to be completed.

> The curricular materials and the professional development were essential. They gave me the content and the pedagogy to walk into the classroom knowing what I was going to teach and why I was teaching it.

> [In Project PRIME] I got content knowledge, pedagogical knowledge, learning how to teach, and understanding the reasons why you do things a certain way. [It caused me] to understand and think about what I am doing and why. I didn't really do that in my other science classes. I think about it now throughout my day, not just in biology.

> I would say that content is probably not the most important. Pedagogy [is important] because it has to do with the strategies involved in helping kids learn. I am still learning content but it doesn't mean that I can't get the content out to them in an accurate manner. But if I can't figure out a way to get it to them, it doesn't matter how much content I know.

> Specifically, using the evaluate activities and the analysis questions [in the curriculum] has helped me with pedagogical knowledge; ways of looking at whether my students learned. Before, I did not have enough higher level questions or activities that would lead me to know if the students really understood what they were taught.

> Project PRIME has shown me how to teach students how to be learners rather than just giving them information.

Prior to Project PRIME, teachers stated that purposeful consideration of their instructional choices was uncommon and that they rarely engaged in reflection. About 25% of the teachers credited Project PRIME with helping them better understand or articulate the rationale for their teaching practices. Teachers credited the instructional materials and professional development with providing concrete support for increased inquiry-based instruction. For many, inquiry teaching represented a change in practice and a new teaching commitment. Many teachers also noted a new focus on student learning with increased attention to prior knowledge, questioning techniques, and instructional practices that facilitated metacognition. These changes support our quantitative findings of an increase in all areas of teacher professional knowledge and the impact of the intervention on teacher knowledge bases of ACK, GenPK, PCK-CK, and PCK-PK.

The *Teaching Practice* of the participants became significantly more inquiry-oriented from baseline to the average of Years 1 and 2 (baseline: $X = 15.72$, SD = 6.86; Years 1 and 2 average: $X = 23.87$, SD = 3.98). This change represents an increase in inquiry-oriented teaching practices and skills such as involving students in active exploration of the content, hypothesis generation, examining multiple representations of data, connecting CK to other understandings, and refocusing instruction based on student feedback. In addition, s*tudent achievement* scores increased significantly from pre-test to post-test [t (4717) = 58.39, $p < .001$]. With regard to the *practical significance* of this gain in student achievement, we calculated a mean normalised gain score of 24%. That is, students gained 24% of the gain that was possible, pre-test to post-test. The pre–post effect size (d) = .98 is considered 'large' by some researchers though we interpret this finding with caution because there are no literature-based benchmarks for longitudinal teacher effects to which we could defensibly compare.

Relationships among teacher variables

Component 2 of our theoretical path of influence focuses on the relationships between teacher practice (skill) and the teacher knowledge bases: ACK, GenPK, PCK-CK, and PCK-PK. Because the most complete teacher data was from instruction in the areas of the cell and evolution (i.e. areas taught by almost all of the teachers), we used the data from those areas to check for correlations among the teacher variables. Many teachers encountered difficulties in video recording their classroom instruction; so we only had complete data for 18 teachers. Tests run within SPSS suggested that the teacher data were not missing at random. Thus, we employed a missing data analysis procedure to impute missing teacher data.

A correlation analysis (Table 2) indicates that moderate correlations exist between the teacher variables. However, regression diagnostics suggest that the teacher variables are not correlated strongly enough to constitute a multi-collinearity problem in the multiple regression or hierarchical linear modelling (HLM) analyses.

As above, the teacher variables are generally related to one another with 4 of the 10 correlations statistically significant. In particular, ACK is related to PCK-CK. This relationship suggests that the two knowledge bases may build upon one another, though direction of influence cannot be determined from this analysis. Further, ACK appears to be more distinct than any of the other teacher variables because it is significantly correlated *only* with PCK-CK.

Our revised theoretical model proposed that PCK consisted of two factors, PCK-CK and PCK-PK. Further analysis indicated that the correlation between these two factors is moderate, suggesting some degree of distinctness, and providing cautious evidence in support of our assertion that PCK is a unique construct separate from the other knowledge bases. Additionally, PCK-CK was correlated with GenPK, suggesting that knowledge related to creating a classroom community characterised by active participation, respect, and student engagement has application to topic-specific teaching. Of all the variables, only GenPK was significantly correlated to classroom practice. Further, no statistically significant relationship was found between years of teaching experience (as measured by teacher self-report) and any of the other teacher variables.

Regression of teacher practice on ACK, GenPK, PCK-CK, and PCK-PK using the missing data-imputed data files gave the results shown in Table 3. Teachers with higher GenPK scores exhibited more inquiry-based teaching practices potentially indicating that knowledge of why certain teaching practices were used increased the potential that these practices would be realised in the classroom. This relationship was statistically significant. In contrast, no statistically significant relationships were found between ACK, PCK-CK, or PCK-PK and teacher practice.

While we are encouraged by the significant relationship between GenPK and teacher practice, we are uncertain about how to explain the limited relationship of the other teacher variables to practice. We expected, based on theory, to see a stronger relationship between classroom practice and each of the content connections (PCK-CK), knowledge of students, and the selection of appropriate instructional strategies (PCK-PK); however, this is an area for which teachers admitted to having limited knowledge of and experience. One explanation may be that Teaching Practice and GenPK were both measured as skills. The lack of correlation of the other measures, which were all measuring knowledge, may have to do with the difficulty of translating what one knows into practice and/or the need for time and experience to support that translation. In addition, careful consideration of student learning as the metric through which to make instructional decisions was one example of where the project introduced a dramatically new way of thinking about teaching. Inquiry teaching was another. One teacher noted, 'One of things that I have pulled from this programme is to base everything on student learning.' Another concurred, 'That is something I will continue to look at: how to know if the kids really conquered this information, that they really understand the concepts.' A third teacher attributed this new focus directly on the analysis of student work that occurred as part of the project:

Table 2. Correlation matrix of teacher variables ($n = 40$).

	ACK	GenPK	PCK-CK	PCK-PK	Practice
ACK	–	.354	.672**	.237	.283
GenPK		–	.522**	.190	.727**
PCK-CK			–	.503*	.299
PCK-PK				–	.092
Practice					–

*Correlation is significant at the .05 level (two-tailed).
**Correlation is significant at the .01 level (two-tailed).

Table 3. Multiple regression of teacher variables to teacher practice ($N = 40$).

	B	SE	*t*-Ratio	*p*
ACK	.091	.114	0.797	.426
GenPK	.537	.128	4.178	<.001
PCK-CK	−.223	.290	−0.769	.443
PCK-PK	.011	.244	0.045	.964

> It was huge to look at student work and really think about what you are doing and why, and the close analysis [of student thinking]. I had never done that on my own and I probably wouldn't have. I didn't have the colleagues to bounce ideas off and just forcing you to go through those steps, to analyze what you are doing and why, and then figuring out if it works or not [has been a change for me].

From other participants:

> [From Project PRIME] I got better classroom management and a better way of teaching. I had heard about teaching inquiry and I knew it was right but trying to implement it myself was pretty pitiful. I would give up easily. So one of the big plusses of PRIME was learning how to really implement inquiry in the classroom in a manageable way and actually see results.

> One of the things that I got out of this program that I was not expecting was to make a shift in my teaching that I'm using every day. It is inquiry based. I find myself asking probing questions in all of my classes, whether it is science or something else. I'm really pushing kids to think for themselves.

Relationships of teacher variables to student achievement

Component 3 of our theoretical path of influence considers the relationship between teacher knowledge bases and practice, and student achievement. We used HLM with the missing data-imputed files to determine the amount of variance in student achievement scores that was accounted for by the teacher factors in Years 1 and/or 2 (Table 4).

Only teachers' ACK appears to be influential, though not significant, in student achievement ($t = 1.91$, $p = .064$). The .30 coefficient can be interpreted as the increase in average student post-test score for every *one* percentage point increase in teachers MFTB score, controlling for the other independent variables in the model. The effect of CK on average student achievement in the larger population could be as small as .00 or as large as .59. The true effect likely falls between these values.

As ACK is the only variable that approaches statistical significance as a predictor of student achievement, we conclude that the majority of this effect on student achievement is attributable to differences in teachers' ACK. Although there is little research like the present study to compare to, we suspect that this is a small but noteworthy effect.

Table 4. Teacher factors that impact student achievement (fixed effects with non-robust standard errors).

	B	SE	*t*-Ratio	df	*p*
ACK	.30	.16	1.91	34	.064
GenPK	.15	.22	0.67	34	.530
PCK-CK	−.51	.39	−1.31	34	.201
PCK-PK	.11	.27	0.410	34	.687
Practice	.30	.32	0.94	34	.350

Table 5. Estimation of variance components.

Random effect	Standard deviation	Variance component	df	χ^2	p
Level 2 (μ)	3.23	10.45	34	1621.37	<.001
Level 1	5.65	31.88			

The relationship between teacher practice and student achievement was weak within the fully specified model, suggesting that teacher practice is not likely a strong mediator of any of the teacher knowledge bases.

The statistically significant χ^2 value in Table 5 for the random teacher effect (μ) confirms that the model, as is, leaves much variance unexplained. Our interpretation is that the model is somewhat underspecified and that the addition of other teacher-level variables as predictors could increase the model's explanatory power. There is more to learn in further study about the effect of teacher variables on student achievement.

The lack of correlation between the teacher knowledge bases and skills to student achievement is inconsistent with our qualitative data. In reflecting on the successes and challenges of implementing the curriculum and practices learned in Project PRIME, our participants had the following to say:

> I think every teacher in the world would get better if somebody could go through their [text]-book with them and show them how to present it and do the activities. You just get so much out of it. Otherwise, even after you have taught it for a few years, you are still shooting in the dark.

> I could understand the rationale for doing certain activities in the order that they were meant to be completed, … like what the students were going to get out of it, which is helpful.

> By far this year, I have had more students say, number one, this was their best class; number two, they never learned stuff in science until this class …. They finally saw connections throughout the entire year of class. I've never had kids say that to me before and I had easily fifty or sixty kids saying things like that. Not necessarily my A kids either. It was my lower level kids that were saying it just as much.

In contrast to this high praise for the programme, other teacher expressed different views and noted barriers to implementing the work of the project in their schools:

> Before [in my teaching], there was a lot more direct instruction. That was a huge change for me. It was a hard change … I imagine I will go back to using more direct instruction than in AHA [in the future], but probably less than before.

> I just wanted to throw [the curriculum] away sometimes and do something else. Sometimes it taught things out of the natural progression order that I had in my head from teaching it before.

> It is kind of tough because the district is not really facilitating this [implementation]. After we had been in [the project] for a year, they were not very happy with it and they wanted to count our test scores.

> A challenge has been for me to get students to think about how they think, to think about what they know. They are not used to doing that, they don't want to do that, and they don't see the point of doing that. Learning new ways to do that has been a challenge for me.

> It is really a struggle because they have been taught to memorize things and spit it back out for a test. That is what our district wants. It is a struggle because they don't want to think, they want you to tell them because that is what they are used to.

> Hopefully I can still bring in some of the labs and the brainstorming questions [in the future]. I will not be using the program in this entire format just because it does not follow our district's curriculum. Therefore, we can't get through enough material in time for the district tests that we give, so our scores suffer and that reflects poorly on us teachers.

When asked about the limited relationship of knowledge, skills, and practice in the teacher interviews, the teachers admitted teaching in a mechanical manner in Year 1, resulting in slow pacing and uneven coverage across topics. As confidence with the instructional materials grew in Year 2, their pacing improved and a willingness to substitute tried-and-true teaching practices from the past into their curriculum. These substitutions resulted in a regression to former teacher practice, many of which lacked the inquiry-oriented qualities that we measured and the programme encouraged. In addition, our participants described a number of school-based variables that prevented them from fully implementing the curriculum materials, including resistant students and colleagues, and pressures to adhere to district-based curriculum and testing in order to 'best' prepare for the state end-of-level test. Since our measure of teaching was based on reform-based teaching practices, the mixture of 'old' and 'new' curriculum materials may account for the non-significant relationships between knowledge, skill, and student outcomes.

Our results may also be explained by the time needed to manifest new learning. Our definitions for PCK represented new understandings introduced as part of the project. While we saw significant growth in individual aspects of teachers' knowledge, with such nascent understandings, it is not surprising that corresponding classroom practices may not exist after a two-year intervention. There appears to be a diminishing impact of the intervention on the pathway from intervention to teacher knowledge bases to teacher practice and then to student achievement. This could be due to the challenges of translating knowledge to practice and achievement, the limited desire to change practice, or a lack of support for navigating other system pressures. It could also be due to the time required for changes in teachers' knowledge to become integrated into teachers' practices, and/or to the need for much greater professional development support in making the translation from knowledge to practice. Finally, our theoretical path of influence is underspecified and does not contain sufficient data to capture the variables that impact student achievement beyond teacher professional knowledge and classroom practice.

Discussion and conclusions

In this exploratory study, we attempted to measure potential changes in teacher knowledge and practice as a result of an intervention, as well as trace such changes through a theoretical path of influence that could inform a model of teacher professional knowledge. We entered the research with a number of assumptions, some of which have been supported by our findings, while others were challenged. In addition, this work contributes to the study of PCK.

First, we needed an instrument to measure PCK. Through the creation of the Project PRIME PCK Reflection and Rubric, we developed a measure that could be applied to written reflections, interview reflections, and videorecorded classroom observations to capture robust information about teacher knowledge and skill. This instrument contributes to the field because it can be used reliably by multiple researchers for the assessment of topic-specific PCK (Gardner & Gess-Newsome, 2011). Through a factor analysis, we were able to provide evidence of PCK as consisting of two related constructs (PCK-CK and PCK-PK) as opposed to our original assumption of a three-construct model.

Second, through a carefully designed intervention (including the use of educative curriculum materials and professional development), we provide evidence of increased teacher knowledge and skill in all areas (ACK, GenPK, PCK-CK, PCK-PK, and Teaching Practice) as well as increased student achievement. These quantitative results as well as our qualitative results from participant interviews provide evidence that the intervention was a success and offer a contribution that builds on other studies of professional development. Our results are similar to those of other researchers who have demonstrated that professional development that challenges teachers thinking, provides them with new instructional materials and practices that support a change in knowledge and beliefs, that is situated in the real work of teachers in classrooms, and that is sustained over time can have a significant influence on teacher knowledge and skill (Thompson & Zueli, 1999).

Third, we explored the relationship among teacher variables in an attempt to validate our model of teacher professional knowledge, which was based on a number of assumptions about how the knowledge bases were related. Our findings showed that ACK is related to PCK-CK, as might be expected. PCK-CK was related to GenPK, suggesting a topic-specific nature to teaching. PCK-CK and PCK-PK were moderately correlated with each other, providing empirical evidence that PCK is a unique construct with content and pedagogical components that are distinct from both academic understanding of a topic and GenPK. Only GenPK had a statistically significant relationship to classroom practice, a trend that held in the more sensitive multiple regression analysis. All other measures, including ACK and the two PCK constructs, were non-significant in predicting classroom practice. Interestingly, PCK-CK showed a weak negative relationship to classroom practice while the other variables exerted a weak positive influence. Together, the teacher knowledge and skill variables accounted for only 53% ($R^2 = .525$) of the variance in teacher practice, highlighting that the model used in this study is underspecified. These findings do not fully support our theoretical path of influence. Clearly, other variables in this study influenced teaching as much as teachers' knowledge and skill.

While student-learning gains were found, we recognise that the lack of a comparison group. As a result, we make no claims that the professional development intervention in this study is more efficacious for students than any other programme. Finally, the way teacher knowledge, skill, and practice affect student learning was examined using HLM. These findings do not support the complexity of our theoretical path of influence as only teacher ACK substantially influenced student learning. These findings do not support our assumption that teachers with strong PCK are more likely to positively affect student learning than those teachers with weak PCK. There may be a variety of reasons for this: we may not be measuring PCK in an empirical manner that captures what teachers describe verbally, there may be an influence of variables outside our study, or the significant amount of time that it takes to translate changes in knowledge and beliefs into classroom practice might account for these limited affects.

The relationship of ACK to student learning contrasts early research that used proxies of teacher knowledge (courses taken and GPA) where weak relationships to student learning were found. Our result, however, is aligned with research that examined teacher knowledge through coherent programmes of study (Milken Family Foundation, 2000; Monk, 1994). One explanation is that direct measures of teacher knowledge (tests) are more powerful indicators of teacher knowledge. Again, we are cautious about this finding. Few measures of CK exist that are appropriate for teachers. We selected the MFTB because it was designed for those who had completed a major in biology and it had established psychometrics. While there was a correspondence of the general categories of knowledge in both the test and the concepts of interest in our study, through closer examination of the questions, they had limited application to the types of knowledge that might be developed through the use of the curriculum materials or the professional development. While the positive impact of ACK on student achievement supports our model, a more aligned CK test would have been more sensitive to the types of gains in knowledge that the teachers thought they had made.

The negative relationship of PCK-CK to student learning bears examination. Our multiple-choice test may not have been sensitive to the learning that one might anticipate from instruction characterised by deep and connected CK and multiple representations and connections to relevant events. This reasoning may help explain our negative correlation and underscores the need for both teacher and student content measures that more accurately reflect what we value in learning.

Our findings also call into question the assumption that classroom practice, as measured by the RTOP, is predictive of student learning in high school biology classrooms. The authors (Piburn & Sawada, 2000) reported a .88 predictive validly correlation for college-level students. In our analysis, however, neither RTOP-based measure (GenPK or teacher practice) had a significant relationship to student achievement. Such a correlation was not evident in our study.

Since the conceptualisation of this research and the studies used to inform its theoretical framework and data analysis, the PCK Summit was convened. Our research, as well as the substantial research that has occurred in the last 10 years around teachers' professional knowledge, was influential in crafting the model of Teacher Professional Knowledge and Skill including PCK (Gess-Newsome, 2015; see Figure 2). We now interpret our findings through that lens.

What we called ACK we would now be included within the teacher professional knowledge bases. Our research, as well as others, shows that there is an influence of teacher CK on the selection and implementation of teaching strategies, or in our case, PCK-CK, though that relationship does not imply directionality.

While we had hoped to find a measure of GenPK, we now recognise that what we measured was general teaching skill as measured by a subset of the RTOP instrument that focused on creating a community of learners. In the qualitative analysis of our interview data, one of the greatest impacts of Project PRIME was on the teachers' ability to provide a solid rationale for instructional decisions that they made in the classroom while using student learning as the metric of success. These comments, more so than their actual teaching practice, may be a truer representation of the Teacher Professional Knowledge Base of Pedagogical Knowledge. In our study, it was this teacher variable that had the greatest influence on teacher practice and may account for an important area of emphasis in future professional development interventions. Both these findings lend support to the consensus model.

We propose that our professional development programme with a focus on curriculum materials that were designed to support coherent conceptual learning for both the student and the teacher and the scaffolding that the intervention provided for inquiry-based instructional practices was a direct attempt to influence our participants' Topic Specific

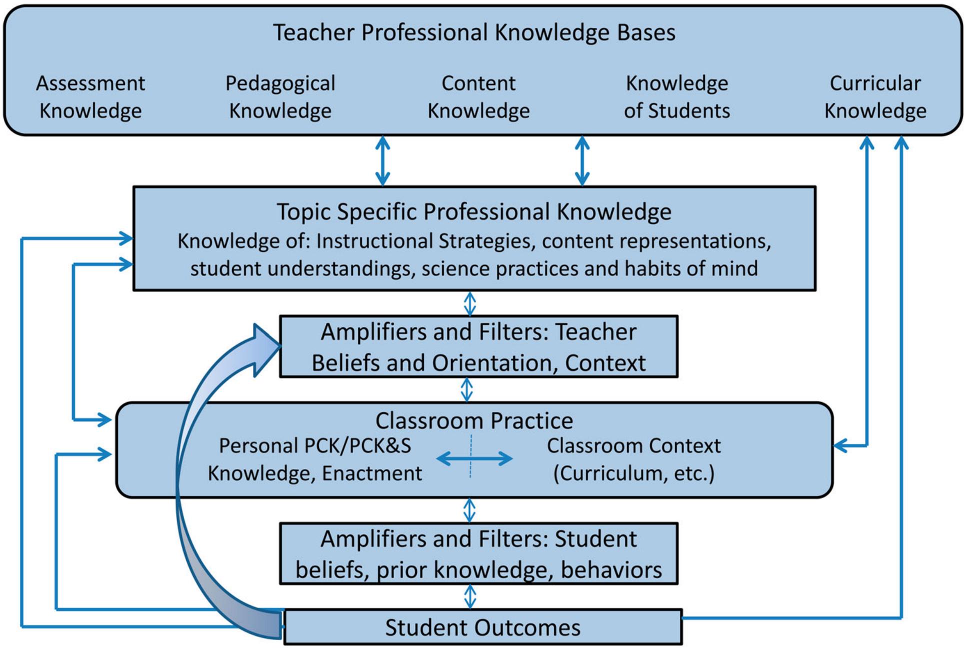

Figure 2. Model of teacher professional knowledge and skill including PCK (Gess- Newsome, 2015)

Professional Knowledge around the topics of focus in Project PRIME. The activities blended content with instructional practices and highlighted student prior conceptions. We believe that the curriculum materials and the professional development were a manifestation of Topic Specific Professional Knowledge. The study provides evidence that a blended focus on content and pedagogy not only improved what teachers learned about the specific topics, but also increased their learning in the Teacher Professional Knowledge Bases of Academic Content Knowledge (as shown by gains on the MFTB), in their PK (as revealed in interviews), and their teaching practice (as demonstrated in increases of Teaching Practice and GenPK scores).

Our study provides additional evidence that there are several components of PCK that have to be looked at simultaneously when capturing or measuring PCK. First, we believe that the PCK reflections and interviews provided us with information about PCK as a knowledge base. PCK-PK was noted as being intertwined with the context of teaching-specific students, while PCK-CK acted as the subject matter context to be considered. This was explicit knowledge that teachers could share with us. Videos of classroom instruction provided evidence of the infusion of pedagogical content and skill. Some of the less impressive correlations in our study may be explained by the fact that while our participants knew what they wanted to do in the classroom, they may not have had the skill to implement it. In addition, the consensus model points out amplifiers and filters that exist between what a teacher knows and does. In this study, teachers noted factors such as personal beliefs, school contexts, and resistant students that decreased their motivation to implement practices promoted in Project PRIME. This is similar to the forces described by Kennedy (2010). Kennedy proposed that the educational research community might be guilty of attribution error where researchers overestimate the power and influence of teacher personal characteristics and practices. Instead, she proposed that situational forces such as teachers' work (time, materials, and work assignments), students, school interruptions into the classroom, and the difficulty of attending to multiple reform efforts simultaneously may have a stronger relationship to student learning than teacher quality. Many of our participants described issues such as deterrents to their attempts to implement the curricular materials used in the study.

As highlighted by the discussion above, our data provide evidence that our theoretical model of influence is underspecified in terms of describing the variables that influence student learning. The consensus model of Teacher Professional Knowledge and Skill including PCK helps highlight the myriad of influences that we were not able to measure in this study and points to the complexity of instruction and learning. It should be clear that developing programmes of research that can trace educational interventions to student learning will always be challenging.

Notes

1. In the US, science teachers in some states are often endorsed to teach *all* science courses regardless of their science background. For example, a teacher with a background in geology could be assigned to teach biology despite limited content preparation. This lack of preparation in biology content was evident in a number of our participants. For the sake of comparison, 18 semester credits is the typical minimum for a minor in a subject area.
2. For more information about EDC, see http://www.edc.org/. For more information about BSCS, see www.bscs.org.
3. For more information, see www.ets.org/MFT/faq
4. We interviewed the 19 teachers several months after the completion of the intervention. This is the total number of participants who responded to multiple invitations to participate in an interview.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation [grant number ESI-0455846] and Division of Elementary, Secondary, and Informal Education.

References

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krass, S., Neubrand, M., & Tsai, Y. (2010). Teachers' mathematical knowledge, cognitive action in the classroom, and student progress. *American Educational Research Journal*, *47*, 133–180.

Beyer, C. J., Delgado, C., Davis, E. A., & Krajcik, J. (2009). Investigating teacher learning supports in high school biology curricular programs to inform the design of educative curriculum materials. *Journal of Research in Science Teaching*, *46* (9), 977–998.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Third handbook of research on teaching* (3rd ed., pp. 328–375). New York, NY: Macmillan.

Carlson, J., Stokes, L., Helms, J., Gess-Newsome, J., & Gardner, A. 2015. The PCK summit: A process and structure for challenging current ideas, provoking future work, and considering new directions. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 14–27). London: Routledge Press.

Gardner, A.L., & Gess-Newsome, J. (2011, April). *A PCK rubric to measure teachers' knowledge of inquiry-based instruction using three data sources*. A paper presented at the annual meeting of the National Association for Research in Science Teaching, Orlando, FL.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). London: Routledge Press.

Henze, I., van Driel, J. H., & Verloop, N. (2008). Development of experienced science teachers' pedagogical content knowledge of models of the solar system and the universe. *International Journal of Science Education*, *30*, 1321–1342.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematics knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406.

Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher*, *39*, 591–598.

Kirschner, S., Borowski, A., Fischer, H. E., Gess-Newsome, J., & von Aufschnaiter, C. (2016). Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge. *International Journal of Science Education*, *38*, 1343–1372.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome, & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Dordrecht: Kluwer.

Milken Family Foundation. (2000). *How teaching matters: Bringing the classroom back into discussion of teacher quality*. Princeton, NJ: ETS.

Monk, D. H. (1994). Subject area preparation of mathematics and science teachers and student achievement. *Economics of Education Review*, *13* (2), 125–145.

Park, S., & Oliver, J. S. (2008). National Board Certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, *45*(7), 812–834.

Piburn, M., & Sawada, D. (2000). *Reformed teaching observation protocol (RTOP) reference manual*. Technical Report. (No. ACEPT-TR-IN00-32).

Schneider, R. M., & Krajcik, J. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education*, *13*, 221–245.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15* (2), 4–14.

Smith, P. S., Banilower, E. R., McMahon, K. C., & Weiss, I. R. (2002). *The national survey of science and mathematics education: Trends from 1977 to 2000*. Horizon Research, Inc. Available at: 2000survey.horizon-research.com.

Thompson, C., & Zueli, J. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond, & G. Sykes (Eds.), *Teaching as the learning profession*. pp. 341–375. San Francisco, CA: Jossey-Bass.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89–122.