A PCK Rubric to Measure Teachers' Knowledge of Inquiry-Based Instruction Using Three Data Sources

April L. Gardner, BSCS
AGardner@bscs.org

Julie Gess-Newsome, Northern Arizona University
Julie.Gess-Newsome@nau.edu

A paper presented within the Strand 10 paper set
Measuring Teacher Inquiry Knowledge
at the
Annual Meeting of the National Association for Research in Science Teaching
Orlando, FL April 6, 2011

Abstract:

For science, the qualities of PCK mirror those of inquiry-based teaching. We developed a rubric that used eight criteria to examine teacher knowledge and practice. The rubric was organized into three elements: content knowledge, pedagogical knowledge, and contextual knowledge. A factor analysis revealed only two factors, and contextual knowledge was combined into pedagogical knowledge. We used the rubric with three data sources: written reflections, interviews, and video of classroom practice. Inter-rater reliability for the scoring was high. Analysis of the written reflections was complicated by extremely short responses. Interviews resulted in somewhat more information, as did use of the rubric on classroom videotapes for some participants. Teachers admitted that they rarely reflected on their classroom instruction in the manner that we were requesting and had limited motivation to provide answers that were more extensive. Content knowledge scores were generally higher than pedagogical scores. These results suggest that teachers have limited practice in reflection, or that knowledge for classroom practice is tacit and not easily expressed outside of action. Changes in inquiry-oriented practice over a two-year intervention occurred only between baseline and an initial professional development session, suggesting that extensive time is needed to effect change in classroom practice.

Work made possible with a grant from NSF: ESI-0455846

We acknowledge the contributions of the following members of the PRIME research team: Kristin Anderson, Nena Bloom, Janet Carlson, Kenric Kesler, Jane Larson, Joe Taylor, & Chris Wilson.

To access this paper and power point presentation: www.bscs.org/sessions

Also access: Gess-Newsome, J., Cardenas, S., Austin, B.A., Carlson, J., Gardner, A.L., Stuhlsatz, M., Taylor, J.A, & Wilson, C.D. (2011, April). *Impact of educative materials and transformative professional development on teachers' PCK, practice, and student achievement.* A paper set presented at the annual meeting of the National Association for Research in Science Teaching, Orlando, FL.

A PCK Rubric to Measure Teachers' Knowledge of Inquiry-Based Instruction Using Three Data Sources

Purpose

Project PRIME (Promoting Reform through Instructional Materials that Educate) was a multi-year study to examine the nature and growth of pedagogical content knowledge (PCK) among high school biology teachers who implemented educative curriculum materials (Schneider & Krajcik, 2002) following transformative professional development (Thompson & Zeuli, 1999) based on these materials. Thus, the study required a measure for assessing PCK. Unfortunately, measures to capture pedagogical content knowledge are in their infancy. The lack of appropriate measures of PCK for Project PRIME required us to create and test such an instrument within the context of the larger project. This paper describes the development, use, and findings of the PRIME PCK Rubric and Reflection.

Theoretical Framework

Researchers have long sought explanations for classroom variables that improve student learning outcomes. While the process-product research of the 1980's identified teaching behaviors that increased student achievement, none of these relationships were found to be strongly correlated to student learning (Brophy & Good, 1986). Additionally, implementing any number of combinations of behaviors produced similar results making it difficult to determine the most important teaching behaviors. The examination of teacher background characteristics such as content knowledge and teaching experience proved equally futile with only a limited relationship to student outcomes.

To move the field past this impasse, Shulman (1986) proposed a "missing paradigm" in educational research – pedagogical content knowledge, or PCK. The construct challenged past practices of examining knowledge of subject matter and pedagogy separately. Instead, pedagogical content knowledge, or PCK, recognizes the melding of subject matter expertise with pedagogical strategies and knowledge of the learner to produce high-quality classroom practice. For Shulman, PCK is a unique knowledge base held by teachers that allows them to consider the structure and importance of an instructional topic, recognize the features that will make it more or less accessible to students, and justify the selection of teaching practices based on student learning needs. With PCK, neither content knowledge nor generic teaching skills alone are sufficient for effective teaching.

Shulman proposed that PCK was one of seven knowledge bases in the professional knowledge of teachers. Other knowledge bases included subject matter knowledge, pedagogical knowledge, curricular knowledge, knowledge of students, knowledge of context, and knowledge of educational goals. Based on this early conceptualization, many researchers attempted to operationally define and distinguish these knowledge bases. In science, Magnusson, Krajcik, and Borko (1999) offered one of the initial characterizations of the PCK construct (Figure 1). In this conceptualization, PCK is influenced by (and in turn influences) subject matter knowledge and beliefs, pedagogical knowledge and beliefs, and knowledge and beliefs about context. Within PCK there are a number of additional features of PCK that were proposed for measurement.

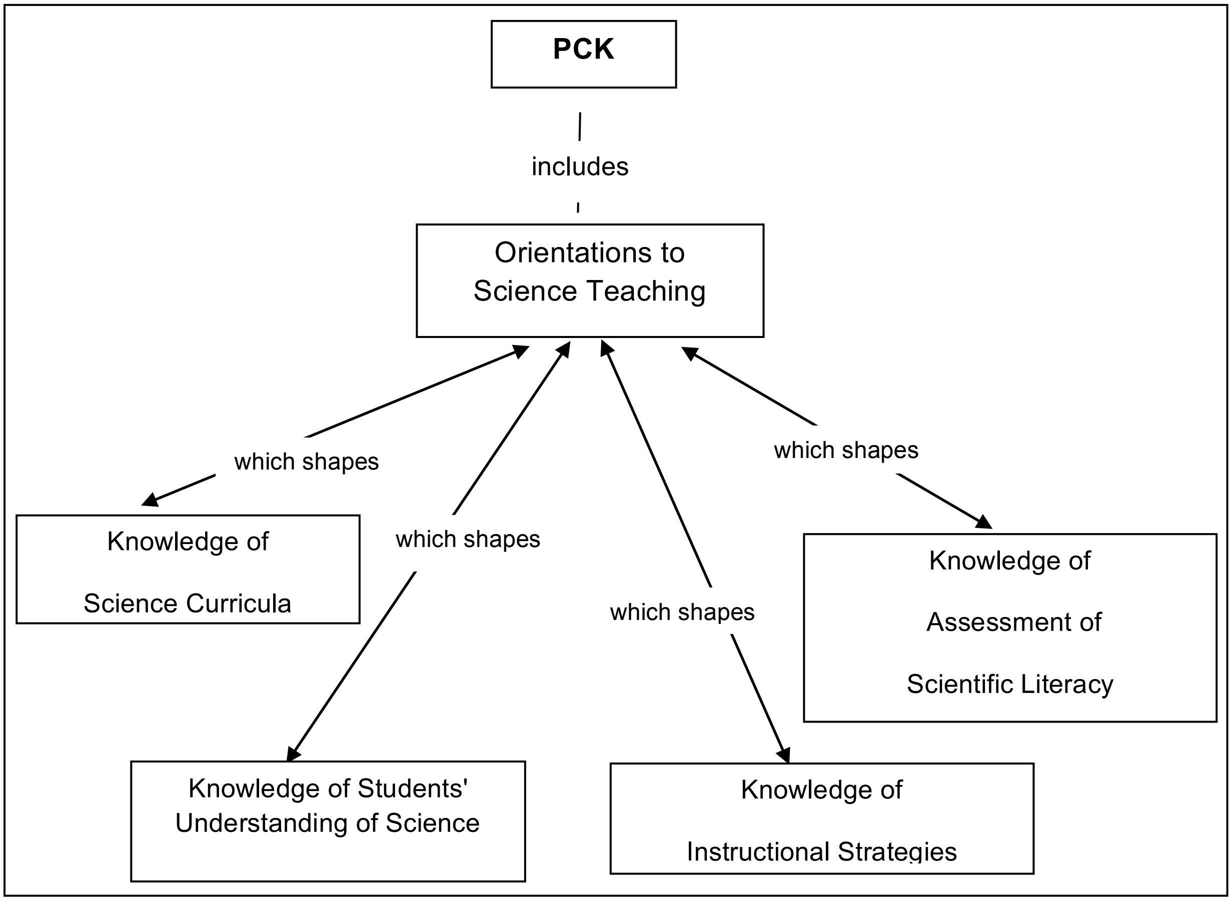
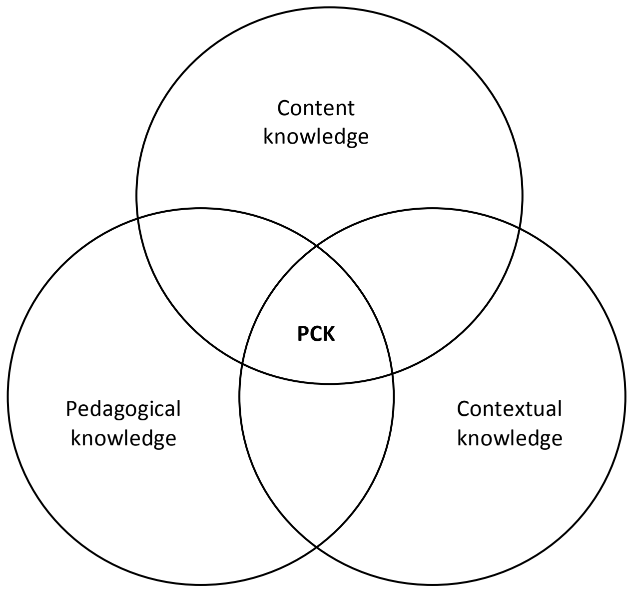


Figure 1. A transformative model of PCK. Adapted from Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching (p. 99), by Magnusson, S., Krajcik, J., and Borko, H., 1999. In J. Gess-Newsome & N.G. Lederman (Eds.), Examining pedagogical content knowledge: the construct and its implications for science education. Kluwer: Dordrecht.

The Magnusson et al. model (1999) closely follows Shulman's early conception and proposes that other knowledge bases (subject matter knowledge, pedagogical knowledge, and knowledge of contexts) are *transformed* into PCK, which can then be described by the elements above. This conceptualization is in contrast to an *integrative* model proposed by Gess-Newsome (1999) where PCK is the integration of the other professional knowledge bases for teaching (see Figure 2). For instance, teachers can possess academic knowledge of content (such as that measured by multiple-choice tests given to college content majors), general pedagogical practices (such as pacing and classroom management), and the ability to list student misconceptions. That knowledge can be contrasted with that knowledge that exists *within* PCK (the integration of the same knowledge bases for the teaching of specific topics) and is useful for classroom practice. For the Project PRIME study, Powell, Taylor, and Gess-Newsome (2005) proposed that the professional knowledge bases of teachers are integrated to form PCK. Specifically, aspects of PCK include the content knowledge needed to help students understand the topic, pedagogical strategies to engage students and promote conceptual understanding, and the knowledge of student context, such as student variations in background knowledge and experiences that can impact learning.

Figure 2. An integrative model of PCK



Strategies for assessing PCK. Researchers have not agreed on any standardized procedure for evaluating teachers' PCK. When we started our research in 2004, the most common method of capturing PCK was through a case study approach (Borko & Livingston, 1989; Carlsen, 1991, 1993; Hashweh, 1987; Loughran, Mulhall, & Berry, 2004). These approaches typically involve one or more procedures such as teacher interviews, classroom observations, teachers' written reflections, and analysis of lesson plans. More recently, Baumert, et al. (2010) developed a scenario-based written test to examine three aspects of secondary mathematics teachers' PCK: *tasks*, to measure ability to identify multiple solution paths; *students*, to evaluate misconceptions and common difficulties students encounter; and *instruction*, to assess knowledge of different representations and explanations. Consistent with the idea that PCK is a knowledge base, Hill, Rowan, and Ball (2005) created a multiple-choice test to assess elementary teachers' mathematical knowledge of teaching, which includes aspects similar to PCK. In both these last two cases, PCK was measured at the domain level (mathematics) as opposed to the topic level (multiplication of two digit numbers).

While teacher professional knowledge was originally described as a knowledge base, and thus open to direct measurement, there was developing disagreement in the field as to whether PCK could be measured outside the act of teaching (Kagan, 1990; Lederman & Baxter, 1999). Individuals who believed it could not hypothesized that PCK may be tacit knowledge and difficult for teachers to articulate. For those individuals, such as Park and Oliver (2008), tools like the PCK Evidence Reporting Table (ERT) were developed. This instrument tallies the number of instances of different aspects of PCK identified in classroom observations.

In the above examples, PCK was considered a characteristic of an individual that could be measured directly or through classroom observations. In contrast, Loughran, Mulhall, and Berry (2004) considered PCK the result of community agreement on how to best teach a topic. They formalized a method to obtain such qualitative data from teachers in the Content Representation (CoRe). In this method, teachers complete a matrix with the "big ideas" they identify for a specific content area and respond to questions such as the importance of the ideas and difficulties students frequently encounter with them.

In our consideration of PCK, we maintained that it was topic specific rather than existing at the domain level. We also believed that PCK was an attribute of an individual that could be measured directly as a knowledge base while still having elements that could be captured in the observation of classroom practice. Finally, we believed that there are both qualitative and quantitative aspects to be considered in the measurement of PCK and that the mere counting of instances would not be sufficient.

Drawing from these perspectives, we sought to develop a tool that would combine the following features:

- Capture topic-specific PCK so that it could be submitted to statistical analysis including comparing scores across topics and across points in time.
- Attend to both the quantity and quality of teachers' PCK.
- Be efficient in scoring and support inter-rater reliability calculations.
- Allow us to identify specific aspects of PCK as proposed in our model (PCK-CK, PCK-PK, and PCK-CxK, see below), relate them to other professional knowledge bases (such as academic content knowledge), and empirically test our model of PCK through factor analysis.
- Lend itself to measurement of both teacher knowledge and practice.
- Allow for the correlation of aspects of PCK to student learning outcomes.
- Allow us to use it across multiple data sources.

Perspective on PCK for this study. For the purpose of this study, we viewed PCK as an integrative knowledge base, with three components: content knowledge (PCK-CK), pedagogical knowledge (PCK-PK), and contextual knowledge (PCK-CxK). We also entered our study with a number of assumptions: (1) PCK is topic specific, (2) PCK exists on a continuum from weak to strong, and (3) PCK can be strengthened with professional development or other high quality learning experiences. Through our study we hoped to examine our conceptualization of PCK, trace change in teachers PCK over time following their implementation of educative curriculum materials and participation in transformative professional development, and correlate the impact of PCK on student learning outcomes. The results beyond the scope of this paper are reported in Gess-Newsome et al., 2011.

Within science education, PCK has a strong connection to inquiry-based learning. Because PCK is discipline- and topic-specific, the pedagogical values of the profession act as the metric for its definition. Thus, inquiry-based instruction would be at the heart of a science PCK instrument, and PCK would logically appear to be a prerequisite for inquiry-based learning.

Methodology

As noted previously, our study used an integrative perspective of PCK in which three teacher knowledge bases contribute to their PCK: content knowledge, pedagogical knowledge, and contextual knowledge. In 2007, Taylor and Gess-Newsome hosted a workshop at the annual meeting of the Association of Science Teacher Educators to present our framework, examine and analyze existing PCK data collection tools, and generate topics that participants felt should be included in the conceptualization of PCK. Using this research community input and our analysis of the literature, we developed the PRIME PCK Rubric (Appendix A) and the PRIME PCK Reflection (Appendix B) to which teachers would respond. The rubric allowed us to examine topic-specific PCK, to use multiple data sources for assessing PCK, and to give quantitative scores to teachers' levels of PCK. Thus we were able to follow changes in their levels of

PCK across time and across different topics. We discuss first the development of the rubric, then the data sources we used, and finally the scoring of the PCK Rubric.

Development of questions for the PRIME PCK Written/Interview Reflections. Teachers were asked to recall a topic-specific lesson that they had recently taught (such as transport mechanisms in cells) and to describe the teaching of the lesson in a step-by-step fashion including what they did, what the students did, and a rationale for each step (See Appendix A). Following the lesson description, the teachers were asked to reflect on nine questions related to:

- the placement of the topic within a larger unit of instruction, the overarching unit concept, the connections among sub-concepts, and the importance of the topic to the study of biology (measuring PCK-CK);
- student incoming knowledge of the topics and points of learning difficulty (measuring PCK-CxK); and
- the advantages and disadvantages of the instructional strategies selected, alternative strategies that were considered, and reasons for changing the curriculum (measuring PCK-PK).

Each question set was modified to direct the teacher to answer the questions in relation to a specific topic. We designed the questions that followed the lesson description to expand upon the information in the description and to more clearly elaborate on the information needed to assign an appropriate score on the rubric (see Appendix B). The nature of these questions was strongly influenced by the questions used by Loughran, Mulhall, & Berry (2004) in the design of their Content Representation, and have strong similarities to the questions asked by Gess-Newsome (2002) and Keeley (2005) in prompts to help teachers think about their content , student background knowledge, and teaching strategies before designing instruction. Scores for each rubric row were derived from a combination of information contained in the lesson steps and in the answer to the questions.

Development of the PRIME PCK Rubric. We developed the rubric to capture aspects of content knowledge, pedagogical knowledge, and contextual knowledge within PCK. Within each element are criteria, described below, that earn a score of 0, 1, 2, and 3, corresponding to limited, basic, proficient, or advanced teacher knowledge. A copy of the PCK Rubric is included in Appendix B.

- *PCK-Content Knowledge* (PCK-CK) (12 points – four components) including depth, breadth, and accuracy of content knowledge; connections within and between topics and the nature of science; and fluency with multiple modes of representation or examples of a topic;
- *PCK-Pedagogical Knowledge* (PCK-PK) (9 points – three components) 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-Contextual Knowledge* (PCK-CxK) (3 points – one component) including understanding how student variations, such as student prior conceptions, impact instructional decisions.

We derived the PCK-CK rubric rows from several sources. Fluency with multiple representations for a topic have been pervasive in the literature (Grossman, 1990; Magnusson et al, 1999; Shulman, 1986). While classic forms of content knowledge (courses taken, GPA) have limited correlations with student achievement (Wayne & Youngs, 2003), other findings demonstrate the value of coherent course work, such as a major or minor (Milken Foundation, 2000) or content courses up to a threshold level (Monk, 1994). We believe that these findings may suggest that coherence in content preparation may not only increase the quantity and quality of content knowledge, but may also increase hierarchical organization of ideas, connections among topics including the nature of science and inquiry, and the possession of multiple representations or examples. Such characteristics also come from the synthesis of the cognitive learning research as qualities of experts and the recommendations for teaching based on this research (Bransford, et al., 2000).

The PCK-PK rows are also derived from the cognitive research literature (Bransford et al., 2000). These findings recognize that learning is about change and that for change to occur, a teacher must recognize student incoming knowledge and misconceptions. Through this recognition, the selection of instructional strategies is purposeful and includes helping students to examine their own thinking and define their own learning goals, or metacognition. Finally, student variations including background knowledge, misconceptions, cultural experiences, and developmental level all must frame the consideration of teaching practices if a teacher is to engage, motivate, and facilitate learning. Such knowledge is included in our row for PCK-CxK.

Each row of the rubric was scored using a combination of the quality of teacher responses and the quantity of specific information they included. For example, in the fourth row of the content knowledge component of the PCK rubric, teacher responses are scored based on the quality and quantity of the examples of the relevant concept that are included. A score of 0, or Limited, for this row was given if the teacher did not include any examples, or included an inappropriate example. A Basic score of 1 was awarded if the responses included an appropriate example, but it was not explicitly linked to the concept. A score of 2, Proficient, was given if an appropriate example was described and explicitly linked to the concept. A score of 3, Sophisticated, was awarded if multiple, appropriate examples were given and linked to the concept.

Data sources. The rubric design provided the opportunity to use it with three different data sources: written reflections by teachers, teacher interviews, and videotapes of teachers' class sessions. For the videotaped class sessions, however, we were unable to score one criterion: teachers' rationale for using particular strategies. While we assume teacher actions are based on a rationale, the rationale itself is not directly observable. The use of the rubric with each data source is described below.

- Written Reflections: Teachers recorded written responses on a computer template in as close temporal proximity to the professional development or teaching episode as possible. Teachers had access to instructional materials, such as the textbook or lesson plans, for reference. Reflections for five different topics were each collected at four time points during the project: baseline (entry to the program); following professional development on the topic; following teaching of the topic in Year 1; and following teaching of the topic in Year 2. Units and topics included in PRIME were the Cell (transport mechanisms), Evolution (natural selection), Heredity and Genetics (protein synthesis), Matter, Energy, and Organization (photosynthesis), and Interdependence (carrying capacity). Two of the five topics (protein synthesis and carrying capacity) were taught by very few of the teachers and so the reflections for those topics were not included in data analyses.
- Classroom Instruction: Teachers videotaped daily classroom instruction. We selected a lesson (one or more class sessions) on the topic of interest and used the rubric to score the teaching observation for each year of teaching during the project. As noted previously, all rows of the rubric were scored except the one that asked about the rationale linking the teaching strategy to student learning since a "rationale" cannot be viewed in practice.
- Interview Reflections: To determine if interviews might increase the quantity and quality of teachers' responses to the reflection questions, we used the same set of questions in the written reflection in an interview setting. In this setting, interviewers were able to probe teachers' initial responses to obtain more detail and insight into their understanding. Teachers had access to

instructional materials. The interviews were recorded and transcribed. The interview was conducted based on a single topic at the completion of the teachers' two years participation in the project.

Rubric Scoring. Five researchers tested the initial rubric using a sample of teachers' written reflections. Scores were discussed, descriptors on the rubric clarified, and consensus on scoring rules and criteria evolved. Several rounds of refinement resulted in the final rubric. Three raters used the final PCK rubric to score the teachers' written reflections. Before they began scoring reflections for a particular topic, the raters developed a list of common student misconceptions about the topic. They added to this list as necessary during the scoring of the reflections. This is one way that raters maintained consistency of scoring, contributing to the high inter-rater reliability described in the next paragraphs. This scoring process highlights an important element of our rubric: only researchers with high levels of topic specific PCK, or those trained to recognize topic specific PCK, will have success in using this rubric.

Scoring exemplars. The following examples of teacher responses illustrate how the reflections were scored using the PCK Rubric. The first example is one teacher's response, in a reflection about teaching natural selection, to the question about the importance of this natural selection in biology. This response was given a score of 2, Proficient, for the "accuracy" row of the content knowledge section of the rubric.

> *Natural selection is the process by which the variation within a population becomes the basis for new species arising.*

The response is accurate but too brief to be considered advanced. It does include fundamental supporting ideas such as variation, impact on a population (rather than individual), and speciation. To be advanced, this response would have noted that natural selection is the mechanism for evolution, and elaborated on how the environment select the variations, or adaptations, that lead to reproductive success for an individual. Those adaptations are then inherited by their offspring. An advanced response might also have noted that evolution by natural selection explains both the unity (similarities) that exist among organisms as well as the diversity of life.

The paragraph below was written in a reflection on teaching cell transport mechanisms in answer to the question regarding what students struggle with in learning this concept. It received a score of 1, or Basic level, for the "contextual knowledge" row of the rubric.

> *Students know diffusion and osmosis. They struggle with details of the plasma membrane.*

This teacher identifies somewhat specific aspects of the concept that are difficult for students; however, responses are too brief to be considered proficient. In addition, the teacher does not relate the advantages of the teaching strategies s/he uses to address students' struggles with the concept.

The last example is from a reflection about a photosynthesis lesson, in response to the question about the advantages of the instructional strategies used in the lesson. It was scored Advanced for the "rationale for the pedagogical strategies used" row in the pedagogical knowledge section of the rubric.

> *This strategy allows students multiple opportunities to acquire the concept of photosynthesis starting at a low risk level and slowly moving into a higher degree of risk—where answers should be correct. It allows students to work with material visually, auditory, and kinetically.*

Though this is a brief response, the teacher identifies multiple advantages of the teaching strategies in this lesson, and explains how these advantages help students learn, such as decreasing risk and addressing multiple learning modalities.

Inter-rater reliability (IRR). IRR was calculated to test for consistency in scoring between raters. A sample of approximately 10 percent of the reflections was scored by two of the raters (one rater scored all of these, and either one or the other rater also scored these reflections), and inter-rater reliability was calculated using the intraclass correlation coefficient. The results showed no significant differences between raters [$F(1, 48) = 1.604$, $p = 0.212$] with an intraclass correlation coefficient of 0.719 (two-way mixed effects model, single measures, absolute agreement). Interpretation of the intraclass correlation coefficient is similar to that of Cohen's Kappa, i.e., 0.40 to 0.59 is moderate inter-rater reliability, 0.60 to 0.79 substantial, and 0.80 outstanding (Landis & Koch, 1977). Throughout all scoring, researchers were blind to both the teacher and the timing of the reflection.

Two raters used the final PCK rubric to score videotaped class sessions based on the same topics as the reflections. Several videotapes were scored by both raters to establish inter-rater reliability, also calculated using the intraclass correlation coefficient. These results also showed no significant differences between raters [$F(1, 6) = 2.455$, $p = 0.178$] with an intraclass correlation coefficient of 0.776 (two-way mixed effects model, single measures, absolute agreement). All of the interviews were scored by the same rater, so it was not necessary to determine inter-rater reliability for this data source. These data indicate that the PCK rubric can be used with confidence across multiple raters and multiple data sources. As with the written reflections, scorers were blind to the identity of the participant or to the timing of the collection of the data set.

After an early round of analysis of the written reflection, it became clear to the researchers that participants were only providing three to four word answers to many questions. Such limited responses were extremely difficult to score. As researchers, we believed that our participants held more knowledge than they were reporting. As an intervention, we distributed the PCK Scoring Rubric to the project teachers and explained the rationale for each row of the rubric. Then we gave each teacher one of his/her reflections and asked them to score the reflection *as it was written* using the rubric. Our goal was to help participants understand the importance of writing complete, thoughtful responses on the PCK reflections. Although teachers appeared to understand this and indicated that they believed they wrote more subsequently, the scorers did not observe a marked increase in the amount of writing on the reflections following this intervention.

Results and Discussion

Factor analysis. We used the PRIME PCK Rubric with written reflections in a study of 35 teachers over two years of a professional development program that included the implementation of educative curriculum materials. As described previously, teachers provided written reflections on five biology topics four times over the course of the project, including prior to the project start. The data used in this study came from three topics: Cell; Evolution; and Matter, Energy, and Organization. Although the rubric was developed to reflect the hypothesized three components of PCK, a confirmatory factor analysis of the full instrument revealed our PCK instrument was comprised of two rather than three factors. The element of content knowledge remained distinct, with factor loadings for the four items ranging from 0.613 to 0.740 on component 1; all loadings for these items were less than 0.40 for component 2. Factor loadings for the remaining four items (three from our pedagogical knowledge subscale and one from our contextual knowledge subscale) indicated these two subscales should be merged. The factor loadings on component 2 for these items ranged from 0.470 to 0.789, while all loadings on component 1 were less than 0.40. Thus, based on our data, PCK is composed of two distinct factors that we call PCK-CK and PCK-PK. In interviews, the participants agreed that they saw pedagogical knowledge and contextual knowledge as closely related. Researchers who scored the reflections shared this sense of correspondence.

General findings from the written reflections. As shown in Table 1, scores for the PCK-CK factor were generally higher than scores for the PCK-PK factor. In both cases, however, scores were well below the maximum possible score of 12. The range of scores for PCK-CK was 0 to 9.5, and for PCK-PK was 0.5 to 6.5. The range of the total score (PCK-CK + PCK-PK) was 1 to 19.

As noted above, scoring the written reflections was a challenge because many teacher responses were extremely short, frequently consisting of one word to three- or four-word phrases. This was true despite examples and explicit instruction on the expectations for the PRIME PCK Reflection. For this reason we used additional data sources for evaluating teachers' PCK.

Table 1. *Mean PCK-CK and PCK-PK scores for three biology areas based on teacher reflections.*

Content Area (specific topic)	Mean Score (*sd*)					
	PCK-Total		PCK-Content Knowledge		PCK-Pedagogical Knowledge	
	Baseline	Highest Subsequent	Baseline	Highest Subsequent	Baseline	Highest Subsequent
Cell (transport across cell membranes)	6.78 (4.34)	11.24 (3.80)	4.23 (2.64)	7.06 (2.10)	2.81 (1.98)	4.62 (2.00)
Evolution (natural selection)	4.52 (2.26)	10.19 (3.11)	4.18 (1.72)	6.46 (2.21)	0.70 (0.72)	3.72 (1.17)
Matter, energy, & organization (photosynthesis)	7.80 (4.50)	11.41 (4.27)	5.26 (2.62)	8.06 (1.74)	3.16 (1.94)	4.26 (2.35)

Variations in PCK by topic. We entered this study with the assumption that PCK was topic specific. That is, we did not believe that teachers would have one level of "biology PCK," but that their PCK would vary in idiosyncratic ways for each teacher for different biology concepts. To test this hypothesis we divided the possible range of PCK scores (0 to 24) into thirds. Scores in the lower third of the range (0 to 7) were considered a low level of PCK; scores in mid-range (8 to 16) were considered a medium level of PCK, and scores in the upper third of the range (17 to 24) were considered a high level of PCK. Each teacher's scores on the PCK Rubric for the three topics of cell transport mechanisms, natural selection, and photosynthesis were ranked high, medium, or low.

Based on this analysis of teachers' PCK scores on the three topics, our hypothesis was supported. There were *no consistent patterns of PCK scores across all of the teachers.* Upon entry to the program, only five out of 21 teachers exhibited similar levels of PCK for all three topics (not all of the 35 teachers who completed the program had previously taught all three of these topics). Although the levels of PCK for the different topics converged somewhat as teachers completed the program (15 out of 24 teachers exhibited mid-range PCK score levels for all three topics at the end of the program), there were still multiple patterns of higher, mid-range, and lower PCK score levels among the other teachers.

Changes in PCK reflection scores following program participation. To examine the overall changes between baseline reflection scores for the content and pedagogical factors of PCK and those scores for reflections completed at later time points, we averaged the scores for topics in the cell, evolution, and matter-energy-organization areas (the three areas taught by almost all participating teachers). Then we conducted matched pair t-tests between the baseline score and the highest subsequent score. This decision was based on the assumption that participants would not *lose* knowledge, although they may put less effort into completing the reflections at some of the later time points. This assumption was confirmed in the post-program interviews. The *teacher knowledge bases for both factors increased significantly following program participation* (PCK-CK: $t = 6.2137$, $df = 26$, $p < 0.0001$, effect size = 1.691; PCK-PK: $t = 5.8907$, $df = 26$, $p < 0.0001$, effect size = 1.603).

We conducted post-program interviews with the teachers immediately following completion of the program and again one or two years later. In these interviews, teachers indicated that they did not think about their teaching as portrayed by the PRIME PCK Rubric, and that reflection was an uncommon event in their professional lives. Teachers admitted that they put limited effort into the reflections after early administrations because, with no feedback on their performance, they saw no personal value in completing the task.

Comparison of PCK scores from reflections and videotapes. We used the PRIME PCK Rubric to rate videotaped class sessions matched to two of the topics from the reflections (Cell area: transport across cell membranes and Evolution area: natural selection). To facilitate this comparison, the scores on the reflections were adjusted by subtracting the rating for the "rationale for instructional strategy" row of the PCK rubric because the scores for the videotaped sessions did not include this row. Comparison of the videotapes of classroom practice indicated that *teachers had higher scores based on the videotaped sessions than on the reflections*. The overall PCK scores and the PCK-CK scores where the differences were statistically significant (see Table 2). Differences in the scores for the PCK-PK factor from the two data sources were not statistically significant. One explanation for the increase in PCK-CK scores may have to do with the nature of the intervention. The curriculum materials that the teachers used and the professional development in which they participated all focused on the connected nature of the content both across the domain of biology and to the nature of science and inquiry. In other words, the mechanical implementation of the curriculum materials may have been enough to increase teacher scores over baseline. It is also possible that it is simply easier for teachers to *implement* content connections and examples than it is to *express* their thinking about the same ideas in writing.

Table 2. *Results of comparing PCK scores for videotaped class sessions vs. written reflections*

Content area	n	Overall PCK			PCK-CK			PCK-PK		
		t	p	ES	t	p	ES	t	p	ES
Cell	30	2.51	0.005	0.648	3.27	0.009	0.845	0.36	0.669	0.092
Evolution	38	2.75	0.009	0.631	2.94	0.006	0.675	1.82	0.075	0.418

When asked about these findings, teachers admitted that they were more motivated to do well in the classroom than they were on the written reflections. Many expressed surprise that their PCK scores on the reflections did not increase more over time (particularly among the later time points) because they felt that they had made fundamental changes in their teaching. One interpretation of these results suggests that the science teachers in this study were unskilled, unwilling, or lacked practice in reflecting on their teaching. A second may be that PCK exists as tacit knowledge that only fully engages during actual teaching and teachers are unaware or unable to explain this knowledge outside of instruction. Finally, limited change over time may speak to the slow process of change in practice. Our data, however, only supports a change between baseline and later time points in the study; there were no significant differences when comparing later time points to each other.

Comparison of PCK scores from reflections and interviews. We also conducted PCK interviews with the teachers on one biology topic using the questions from the written reflections after they completed the two-year program. The interview topics varied by teacher, depending largely on which topic they had taught most recently. We compared the PCK scores of the teachers who were interviewed about their teaching of natural selection or photosynthesis with their PCK reflections on the same topic. For most participants, *the interviews provided higher scores than the written reflections* when matched to the same time point. As shown in Table 3, the differences were statistically significant for overall PCK score and for the PCK-PK factor for the evolution topic, natural selection. The differences were statistically significant for both PCK components as well as for overall PCK for the matter, energy, & organization topic, photosynthesis. It should be noted that multiple individuals conducted these interviews with varying level of skill, including the ability to probe participant's responses. It is possible that, with more skilled interviewers, we would have seen even more extensive differences in the PCK scores from the two data sources.

Table 3. *Results of comparing PCK scores for interviews vs. written reflections*

Content area	**n**	**Overall PCK**			**PCK-content knowledge**			**PCK-pedagogical knowledge**		
		t	**p**	**ES**	**t**	**p**	**ES**	**t**	**p**	**ES**
Evolution	8	2.18	0.048	1.09	1.21	0.132	0.60	2.79	0.049	1.39
Photo-synthesis	14	5.12	0.0002	1.24	2.71	0.018	2.30	7.21	<.000	2.10

Conclusion

Our results indicate that the PRIME PCK rubric can be used to evaluate teachers' PCK from three diverse data sources: written reflections, videotapes of classroom sessions, and teacher interviews. In addition, the rubric is efficient to use, lends itself to statistical analysis and the establishment of inter-rater reliability, and attends to both the quantity and quality of teacher knowledge, skills, and actions. We further point out that, while our perspective specifies that PCK is topic-specific, the PRIME PCK Rubric itself can be applied to any topic without modification. The PRIME PCK Reflection can be used with only minor modifications; specifically, substituting the appropriate topic into the questions (See Appendix A). Thus, our rubric has an advantage over scenario-based instruments in which researchers must be create new scenarios for evaluating PCK for diverse topics. Furthermore, the use of the same rubric for multiple topics facilitates comparison of teachers' PCK for different topics.

There are distinct advantages and disadvantages for using the PRIME PCK Rubric with different data sources. The written reflections are easy to administer in a short period of time and can be scored more rapidly than either videotapes or interviews. They are limited, however, by the willingness and/or ability of teachers to write thorough responses to the questions that accurately reflect their level of understanding. Similarly, the quality of the interviews conducted may also influence scores. In brief, sketchy responses depress the PCK scores obtained, as we observed in our comparisons of PCK scores for the same topic using different data sources. One solution may be found in the comments of the teachers: motivation. Our participants admitted that they were more motivated to do well when teaching than they were when responding to our probes either in writing or in an interview situation since there was no feedback concerning their work or progress. In future longitudinal studies, it may be valuable to score PCK prompts throughout the project and report scores to participants in order to increase motivation and, perhaps, collect interview data about perceived mismatches between scores and teacher perceptions of growth.

Our results suggest that scoring videotaped class sessions using the PRIME PCK Rubric results in higher scores for the content knowledge component of PCK, while scoring teacher interviews results in higher scores for both components of PCK. The disadvantage of both of these data sources is the time required for collecting and scoring the data. If our results here are found to be consistent in further tests of the

PCK rubric (i.e., using the rubric to score reflections, videotaped class sessions, and interviews on the same topic with the same teachers), it may be that a combination of teacher interviews plus videotaped class sessions will provide the most accurate assessment of PCK. The labor-intensive nature of these data sources could be reduced, however, for researchers who are most interested in specific components of PCK. For example, those who are most interested in studying the pedagogical component of PCK may want to use interviews only, or interviews plus written reflections to collect and score data.

Further, the best use of scores from multiple data sources or points in time warrant careful consideration. For instance, based on the fluctuating scores on PRIME PCK Rubrics from the written responses, we elected to use only the highest score following baseline, working from the assumption that knowledge is not lost within the course of a project while motivation may decrease. Our data, however, show that there are differences in the scores obtained from different data sources. Perhaps the most accurate portrait of PCK is found in the combination of data points rather than the consideration of any one data type alone. For instance, a teacher's PCK scores from multiple data sources might look as follows:

Table 4: *Examples of one teacher's PCK scores across several measures and how they may be combined*

	Written Reflection	Interview	Classroom Observation	Combined Score	Average Score	Highest Score
PCK-content knowledge	8	8.5	10	26.5	8.8	10
PCK-pedagogical knowledge	5	6	3	14	4.7	6
PCK - Overall	13	14.5	13	40.5	13.5	16

The score from a single source will reflect only the strength of that data source while the combined score provides for the most robust measure of PCK. Three options exist. In the first, a researcher may elect to simply add scores from all data sources. While this method may give an accurate score, its use might be compromised if there are missing data sets, making final scores uneven. To attend to the issue of missing data, a researcher may elect to take an average of the scores or take the highest score. In the case of averages, missing data sources can be mitigated, but the assumption exists that all data sources are contributing equally to the final score. Data presented in this paper suggests that this is not the case. The use of the highest score also attends to missing data but includes the acknowledgement that some data sources have a greater sensitivity to some types of PCK knowledge. For instance, if teacher knowledge is tacit, the value of including a highest score that may come from the observation of classroom practice may give a more accurate representation of a teacher's PCK than using an average. The selection of a best model will need further exploration. In addition, researcher effort in collecting and analyzing this data must be balanced against the perceived gains in accuracy.

In summary, the use of the PRIME PCK Rubric and Reflection shows promise for studying PCK. Continued use and refinement of the instrument should provide additional insights into the nature of PCK and how it develops among teachers.

Work made possible with a grant from NSF: ESI-0455846

We acknowledge the contributions of the following members of the PRIME research team to this paper: Janet Carlson, Jane Larson, Molly Stuhlsatz, Joseph Taylor, & Chris Wilson.

To access this paper and power point presentation: www.bscs.org/sessions

References

Ball, D.L., Thames, M.H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59,* 389-407.

Barnett, J., & Hodson, D. (2001). Pedagogical context knowledge: Toward a fuller understanding of what good science teachers know. *Science Education, 85*(4), 426-453.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133-180.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers *American Educational Research Journal, 26*(4), 473-498.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC.: National Academy Press.

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: Macmillan.

Carlsen, W. S. (1991). Effects of new biology teachers' subject-matter knowledge on curricular planning. *Science Education, 75*(6), 631-647.

Carlsen, W. S. (1993). Teacher knowledge and discourse control: Quantitative evidence from novice biology teachers' classrooms. *Journal of Research in Science Teaching, 30*(5), 471-481.

Gess-Newsome, J. (2002). The use and impact of explicit instruction about the nature of science and science inquiry in an elementary science methods course. *Science and Education, 11*, 55-67.

Gess-Newsome, J., Carlson, J., Gardner, A.L., & Taylor, J.A. (2011, April). *Impact of educative materials and transformative professional development on teachers' PCK, practice, and student achievement.* A paper set presented at the annual meeting of the National Association for Research in Science Teaching, Orlando, FL.

Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education* (pp. 3-20). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education.* New York: Teachers College Press.

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.

Kagan, D.M. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks Principle. *Review of Educational Research, 60*(3), 191-469.

Keeley, P. (2005). *Science curriculum topic study: Bridging the gap between standards and practice.* Thousand Oaks, CA: Corwin Press.

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lederman, N.G., & Baxter, J.A. (1999). Assessment and measurement of pedagogical content knowledge. In J. Gess-Newsome & N.G. Lederman (Eds.), *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education* (pp. 147-161). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching, 41*(4), 370-391.

Magnusson, S. J., 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: The Construct and its Implications for Science Education* (pp. 95-132). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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.

Powell, J. C., Taylor, J. A., & Gess-Newsome, J. (2005). Project BEST: Better education for science teachers (Renamed: Project PRIME: Promoting reform through instructional materials that educate). A grant proposal funded by the *National Science Foundation- T*eacher Professional Continuum. Retrieved from BSCS.org\ProjectPRIMEPapers

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

Taylor, J.A., & Gess-Newsome, J. (January, 2007). *Exploring tools and methods for measuring pedagogical content knowledge.* A paper presented at the annual meeting of the *Association of Science Teacher Educators*, Clearwater Beach, FL.

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

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

Wilson, S. M., Shulman, L. S., & Richert, A. E. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring Teacher Thinking*. Sussex, England: Holt, Rinehart, & Winston.

Appendix A: PRIME PCK Reflection

Example: PCK Reflection on Transport Mechanisms in Cells

Think of a lesson you taught about **transport mechanisms in cells**. Use the matrix below to describe the lesson (use as many rows as necessary; if you need more, continue on an attached page).

Step	What I Will Do	What the Students Will Do	Rationale
1			
2			

1. What unit is **transport mechanisms in cells*** a part of in AHA/Insights?
2. What is the overarching concept in the unit named in question 1?
3. What are the connections between the overarching concept of the unit and the sub-concepts within the unit, including **transport mechanisms in cells**?
4. Why is **transport mechanisms in cells** important in the study of biology?
5. What did you find that your students typically knew about **transport mechanisms in cells** when they come to class?
6. What did students typically struggle with when they were studying **transport mechanisms in cells**?
7. What were the advantages of the teaching strategies you used in this lesson?
8. What alternative teaching strategies did you consider in addition to those suggested in the curriculum?
9. If you made changes to the curriculum, what prompted you to do so?

*To use this prompt with other topics, substitute the topic of interest for the text in bold.

Appendix B: PRIME PCK Rubric and Scoring Instructions

PRIME PCK Rubric

Standards and assumptions to use in applying the Rubric:

- Assume that the teacher's use of language reflects the accuracy of his/her understanding. As a general rule, DO NOT "read between the lines" and assume that the teacher understands some point, but did not explain it well in writing: base the score on actual evidence and not on assumptions. A possible exception would be if the response to another question makes it clear the teacher *did* understand the concept. In that case, make a note in the "evidence" column that "teacher's response to Question x indicates s/he does understand this point."
- As a general rule, when assigning scores reviewers should rely more on the teacher's responses to the questions and their rationale for the steps in the classroom procedures than the "what the students do/what the teacher does" lists
- For the "nature of science" sub-row in the CK row, count NOS examples in the procedure and responses to questions, even if the teacher does not explicitly call them examples of NOS or science as inquiry, etc. Consider both the abilities of scientific inquiry and the understandings of scientific inquiry as part of the NOS.
- Note that if a teacher does not describe a lesson that addresses the requested concept (the first sub-row in the CK row) it is likely that most of the following rows and sub-rows will end up scoring 0. That is, DO NOT score subsequent rows based on the inaccurate concept the teacher selected.
- Don't count "examples" unless the teacher explicitly states them, even if you know they're in the materials used. Analogies count as examples (e.g., a baggy or dialysis tubing as examples of cell membranes), but score a 2 or 3 ONLY if the teacher explicitly connects them to the relevant concept (i.e., it may be obvious to us & the teacher that dialysis tubing represents the cell membrane, but is it obvious to students?).
- If the teacher responds "nothing" or "not much" to a question about what students know about a concept, or what they struggle with, score the response as a 0.
- If the teacher reports using a strategy that should elicit prior student knowledge, but his/her rationale does not indicate that this was the purpose of the strategy, score this as a 0.

PRIME PCK Rubric

Teacher ______________________ **Topic** ______________________ **Reviewer** ______________ **PCK Score** __________

Date of completion ______________ **Date of review** ________________ **Level of detail in reflection** **H** **M** **L**

DIMENSION OF PCK	LIMITED 0	BASIC 1	PROFICIENT 2	ADVANCED 3	SCORE/EVIDENCE
CK: Understands concept & role of concepts in the discipline					**Total CK Score:**
--Accuracy	All or mostly inaccurate: Selected lesson does not address requested concept OR Selected lesson addresses requested concept, but there is a glaring inaccuracy or many small inaccuracies	Somewhat inaccurate: Selected lesson tangentially addresses requested concept OR Selected lesson addresses requested concept, but there are a few small inaccuracies	Mostly accurate: Selected lesson accurately addresses requested concept and there are only 1 or 2 small inaccuracies OR Selected lesson accurately addresses concept with no inaccuracies, but responses are too brief to be considered "advanced"	Completely accurate: Selected lesson accurately addresses requested concept with no inaccuracies, large or small	Score
--Interconnections	None of the possible connections between concepts and subconcepts are expressed	Few of the possible connections between concepts and subconcepts are expressed	Some of the possible connections between concepts and subconcepts are expressed	Many of the possible connections between concept and subconcepts are expressed	Score
	None of the possible connections between concepts and the nature of science are expressed	Few of the possible connections between concepts and the nature of science are expressed	Some of the possible connections between concepts and the nature of science are expressed	Many of the possible connections between concepts and the nature of science are expressed	Score
--Examples	No appropriate, accurate examples provided	Potentially appropriate, accurate examples of the concept provided but are not explicitly connected to the concept	One appropriate, accurate example of the concept provided and explicitly connected to the concept	More than one appropriate, accurate example of the concept provided and explicitly linked to the concept	Score

DIMENSION OF PCK	LIMITED 0	BASIC 1	PROFICIENT 2	ADVANCED 3	SCORE/EVIDENCE
CxK:					**Total CxK Score:**
Understands how student variations impact instruction	No evidence of understanding of students' common prior conceptions/difficulties and how they might impact instruction	Narrow understanding of students' common prior conceptions/ difficulties and how they might impact instruction	Adequate understanding of students' common prior conceptions/difficulties and how they might impact instruction	Sophisticated understanding of students' common prior conceptions/difficulties and how they might impact instruction	Score

DIMENSION OF PCK	LIMITED 0	BASIC 1	PROFICIENT 2	ADVANCED 3	SCORE/EVIDENCE
PK: Relationship between rationale to teaching strategies					**Total PK Score:**
--Completeness/ Soundness	No rationale provided OR A rationale that does not establish a relationship between strategies and student learning	Simplistic rationale that establishes a weak or partial relationship between teaching strategies and student learning	Adequate rationale that establishes an accurate, but limited relationship between teaching strategies and student learning	A sophisticated rationale that establishes a comprehensive relationship between teaching strategies and student learning	Score
--Strategies	No strategies are used that elicit prior student understandings	Limited strategies are used that elicit prior student understandings	Appropriately varied strategies are used that elicit prior student understandings	Highly effective and appropriate strategies are used, that elicit prior student understandings	Score
	No strategies are used that allow for student metacognition	Limited strategies are used that allow for student metacognition	Appropriately varied strategies are used that allow for student metacognition	Highly effective and appropriate strategies are used that allow for student metacognition	Score:

Total CK Score ____
Total CxK Score ____
Total PK Score ____
PCK Score